3500 *Biochemistry* **1987**, *26*, 3500–3509

Weber, P. L., Drobny, G., & Reid, B. R. (1985a) *Biochemistry 24*, 4549–4552.
Weber, P. L., Wemmer, D. E., & Reid, B. R. (1985b) *Biochemistry 24*, 4553–4562.
Wemmer, D. E., Kumar, N. V., Metrione, R. M., Lazdunski, M., Drobny, G., & Kallenbach, N. R. (1986) *Biochemistry 25*, 6842–6849.
Wider, G., Macura, S., Kumar, A., Ernst, R. R., & Wuthrich, K. (1984) *J. Magn. Reson. 56*, 207–234.
Williamson, M. P., Marion, D., & Wuthrich, K. (1984) *J. Mol. Biol. 173*, 341–359.
Wuthrich, K. (1983) *Biopolymers 22*, 131–138.
Wuthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.
Zuiderweg, E. R. P., Kaptein, R., & Wuthrich, K. (1983a) *Eur. J. Biochem. 137*, 279–292.
Zuiderweg, E. R. P., Kaptein, R., & Wuthrich, K. (1983b) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837–5841.

Locus of a Histidine-Based, Stable Trifunctional, Helix to Helix Collagen Cross-Link: Stereospecific Collagen Structure of Type I Skin Fibrils[†]

Gerald L. Mechanic,* Elton P. Katz,[‡] Masayuki Henmi,[§] Claudia Noyes,[∥] and Mitsuo Yamauchi
Dental Research Center and Department of Biochemistry and Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, and University of Connecticut Health Center, Farmington, Connecticut 06032

Received June 30, 1986; Revised Manuscript Received January 29, 1987

ABSTRACT: The loci of the three amino acid residues that contribute their prosthetic groups to form the stable, nonreducible, trifunctional intermolecular cross-link histidinohydroxylysinonorleucine in skin collagen fibrils were identified. Two apparently homogeneous three-chained histidinohydroxylysinonorleucine cross-linked peptides were chromatographically isolated. They were obtained from a tryptic digest of denatured unreduced 6 M guanidine hydrochloride insoluble bovine skin collagen. Amino acid and sequence analyses demonstrated that the prosthetic groups of $\alpha 1$(I)-chain Hyl-87, $\alpha 1$(I)-chain Lys-16^{c}, and $\alpha 2$(I)-chain His-92 formed the cross-link. The latter results served to define the locus of the stable, nonreducible trifunctional moiety. Identical types of analyses were performed on the three-chained peptides isolated after bacterial collagenase digestion of the cross-linked tryptic peptides. This confirmed the initial identification and location of the three peptides linked by the cross-link. In addition, data reported here provide for a correction of the micromolecular structure for the $\alpha 2$(I) chain. Stereochemical considerations concerning this trifunctional cross-link's specific locus indicate that the steric relationships between the α chains of skin and skeletal tissue collagens are fundamentally different and the intermolecular relationships in skin fibrils are specific for skin. The same molecular relationships also indicate that histidinohydroxylysinonorleucine links three molecules of collagen. The stereochemistry of cross-linking for skin collagen is in accordance with and explains the X-ray findings of a 65-nm periodicity found for this tissue [Stinson, R. H., & Sweeny, P. R. (1980) *Biochim. Biophys. Acta 621*, 158; Brodsky, B., Eikenberry, E. F., & Cassidy, K. (1980) *Biochim. Biophys. Acta 621*, 162].

Covalent cross-links between collagen molecules, packed into fibrillar arrays, are principal determinants of the mechanical properties of connective tissue matrices. The cross-linking stereochemistry derives from the reaction of specific peptidyl aldehydes, in the NH$_2$- and COOH-terminal nonhelical peptides, with vicinal ϵ-amino groups of specific peptidyl residues of Lys and Hyl, located in the triple-helical regions of molecules (Yamauchi et al., 1986). His, a sparse amino acid (0.4 residue percent) in collagen, surprisingly also participates in cross-links (Tanzer et al., 1973; Housley et al., 1975; Bernstein & Mechanic, 1980). The cross-linking chemistry in type I collagen fibrils initially involves residues from two collagen chains, but subsequently, residues from other chains participate to form tri- and tetrachain cross-links. The initial cross-link at the COOH-terminal locus covalently ties together two α chains from different collagen molecules by means of an iminium bond (Fukae & Mechanic, 1980; Yamauchi et al., 1986). The reacting moieties are specifically dictated by the intermolecular stagger in the fibril, which juxtaposes the 5-amino-5-carboxypentanal (Lys^{ald})[1] or Hyl^{ald} residue, in the COOH-terminal nonhelical peptide position 16^{c} of an $\alpha 1$(I) chain, with the ϵ-amino of Hyl-87 in the helical region of either an $\alpha 1$(I) or $\alpha 2$(I) chain (Yamauchi et al., 1986). When residue 16^{c} is Hyl^{ald}, rearrangement of the iminium to a ke-

---

[†] This investigation was supported by National Institutes of Health Grants AM 19969, AM 30587, AM 37604, and DE 02668 and National Aeronautics and Space Agency Grant NAG 2-181.

[*] Author to whom correspondence should be addressed at the Dental Research Center.

[‡] Present address: Department of Biostructure and Function, University of Connecticut Health Center, Farmington, CT 06032.

[§] Present address: Tsurumi University, Tsurumi, Japan. Postdoctoral Fellow–Research Associate in the laboratory of G.L.M., 1984–1985.

[∥] Present address: Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514.

[1] Abbreviations: deH-HHMD, dehydrohistidinohydroxymerodesmosine; HHL, histidinohydroxylysinonorleucine; Lys^{ald}, 5-amino-5-carboxypentanal; Hyl^{ald}, 2-hydroxy-5-amino-5-carboxypentanal; Gdn·HCl, guanidine hydrochloride; NEM, N-ethylmaleimide; TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; DEAE, diethylaminoethyl; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/87/0426-3500$01.50/0 © 1987 American Chemical Society

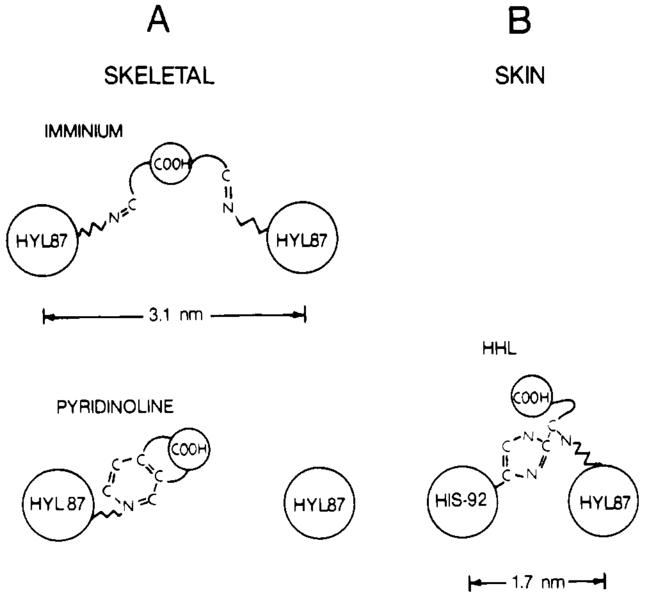

FIGURE 1: Illustration of COOH-terminal cross-links in type 1 skeletal and skin collagen fibrils. Looking down the axis of the collagen molecules, the large circles represent the radial positions (about 0.5 nm) of the β-carbon atoms of residues in the X and Y positions of a Rich–Crick one-bonded collagen triple helix. The small circle represents the possible radial positions (about 0.3 nm) of the α-carbon atoms of the COOH-terminal helical residues. (A) Cross-linking geometry for bone, dentin, tendon, and ligament. An iminium linkage, which is labile, is formed from the reaction of an COOH-terminal Lys-16^c aldehyde in an α1(I) chain and a Hyl-87 in either an α1(I) or α2(I) chain of an adjacent molecule. Two COOH-terminal iminium bonds per molecule involving two different α1 chains on the same molecule are possible (Yamauchi et al., 1986). The circle with the COOH represents this case. The stable, three-chain pyridinoline cross-link is formed with time when one of the iminium bonds dissociates and/or a reaction with a free or iminium-bound COOH-terminal 16^c aldehyde with the remaining iminium bond. The distances between the helical axis of the next nearest neighbors contributing helical residues to the iminium bonds are those estimated for bone or dentin collagen (Katz & Li, 1973b). (B) Postulated cross-linking geometry for HHL formation in skin. This stable, three-chain cross-link involves two helical residues, an α2(I) chain His-92 and an α1(I) chain Hyl-87, and the COOH-terminal Lys^{ald}-16^c, each in a different, nearest-neighbor molecule. The intermolecular distance is for a paracrystalline hexagonal lattice and was calculated from data reported by Brodsky et al. (1980).

toamine can occur (Mechanic, 1974). In the cases of bone, dentin, and tendon fibrils, the geometry of the packing of collagen molecules is such that a trifunctional pyridinoline cross-link can form, from reaction of a ketoamine cross-link with essentially another vicinal COOH-terminal Hyl^{ald}. This three-chain cross-link, which is nonreducible and stable, possibly ties together two molecules in a fibril (Robins & Duncan, 1983) (see Figure 1A).

We report here the localization, through amino acid and sequence analyses, of a three-chain, nonreducible stable cross-link from type I bovine skin collagen fibrils. This cross-link has the imidazole C-2 of His linked to C-6 of Nle, which in turn is linked to the C-6 amino nitrogen of Hyl, and is derived from the reaction of an α1(I) COOH-terminal Lys^{ald}-16^c with the ϵ-amino of Hyl-87 in another α1(I) chain and the imidazole C-2 of His-92 from an α2(I) chain. Its structure was confirmed by fast atom bombardment mass spectrometry, ^{1}H and ^{13}C nuclear magnetic resonance, and we refer to it as HHL (see Figure 2) (Yamauchi et al., 1987). It is a major cross-link at the COOH-terminal locus of an α1(I) chain (Figure 1B). In aging bovine and human skin it approaches a concentration of one cross-link residue per mole of collagen (Yamauchi et al., submitted for publication). If the three chains contributing to the cross-link were from three different molecules in the fibril, then the geometry of packing of molecules in type I fibrils from skin would be fundamentally different than that occurring in fibrils having a more skeletal function. If the chains come from two molecules only, then this would imply that the skin collagen has a different structure from other type I collagen molecules. A unique molecular packing in skin fibrils is indicated by the apparently "anomalous" values for the axial periodicity obtained by X-ray diffraction (Stinson & Sweeney, 1980; Brodsky et al., 1980).

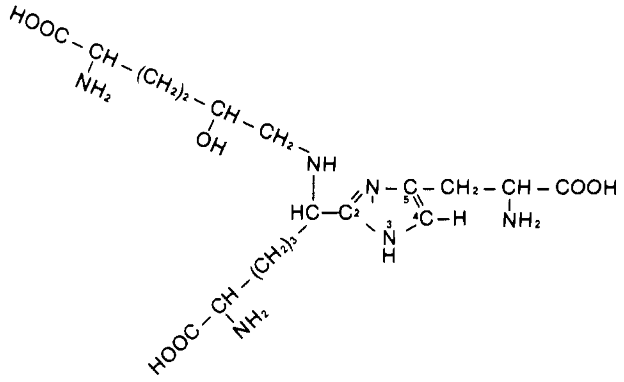

FIGURE 2: Structure of histidinohydroxylysinonorleucine (HHL) confirmed by fast atom bombardment mass spectrometry, ^{1}H and ^{13}C nuclear magnetic resonance (Yamauchi et al., 1987).

The localization of this trifunctional cross-link involved the acquisition of amino acid compositional and sequenching data from apparently homogeneous three-chained peptides. The peptides were isolated from tryptic digests of the residual collagen from 6 M Gdn·HCl extracted bovine skin. Confirmation of its locus was obtained from bacterial collagenase digests of the tryptic three-chain peptides. In the course of this work, deviations from the published sequence for the α2(I) chain were encountered. The details of the corrections to the sequence of the α2(I) chain are described here.

MATERIALS AND METHODS

Apparently homogeneous three-chained HHL cross-linked peptides corresponding to peptides IV and VI were isolated by column chromatography with Sephadex G-50 superfine, phosphocellulose, and DEAE-cellulose as described previously (Yamauchi et al., 1982), with slight modification. The starting material was a tryptic (TPCK-trypsin, Worthington) digest of the denatured collagen residue after 100 g of cleaned lyophilized bovine skin from a 2-year-old female was extracted with 6 M Gdn·HCl.

Additional Chromatographic Procedures. In addition to the chromatographic procedures mentioned above, which were detailed previously (Yamauchi et al., 1982), bacterial collagenase digests of the three-chained HHL peptides were filtered on a Sephadex G-50 superfine column (1.6 × 80 cm) equilibrated with 0.1 M acetic acid and eluted with the same solvent. The flow rate was 12 mL/h, and 2-mL fractions were collected. Peptides IV and VI were chromatographed on the same column to obtain their elution positions. The measured void volume of the column was 59 mL while the total volume (NaCl) was 160 mL (see Results for elution volumes of the peptides).

Bacterial Collagenase Digestion. Approximately 1.4 mg each of peptides IV and VI was treated for 5 h at 37 °C with purified bacterial collagenase (1% w/w, CLSPA, Worthington) in 0.05 M Tris-HCl pH 7.6 buffer containing 0.05 M

$CaCl_2$ and 0.005 M NEM. The digestion mixture was applied directly to the gel filtration column described above.

Amino Acid Analysis. Lyophilyzed aliquots of the isolated peptides IV and VI as well as the fractions from the column after collagenase digestion were hydrolyzed in 6 N HCl in vacuo in a N_2-saturated atmosphere for 24 h at 110 °C. Amino acid compositions were determined in duplicate on an Varian 5560 liquid chromatograph amino acid analyzer as described previously (Yamauchi et al., 1986).

Amino Acid Sequence Analysis. Sequences of peptides IV, VI, C1(IV), and C1(VI) were performed as described previously (Yamauchi et al., 1982) with a Beckman 890 C sequencer.

Results

The three-chained cross-linked peptides IV and VI reported on here were isolated from a tryptic digest of the heat-denatured (65 °C, 10 min, suspended in 0.1 M NH_4HCO_3, pH 8.0, solution) residue of 6 M guanidine hydrochloride extracted bovine dermal layer. Both peptides contained the trifunctional stable nonreducible cross-link, HHL, and each possessed three NH_2-terminal residues.

Essentially similar chromatographic results were obtained as described earlier for the isolation of the three-chained cross-linked peptides corresponding to peptides IV and VI (Yamauchi et al., 1982). Slight variations in the elution position and relative amounts of the three main HHL-containing peptide peaks were obtained, indicating possible cleavage differences. These were attributed to the different batches of TPCK-trypsin that were used in this paper and the previous paper [Yamauchi et al. (1982) and see Discussion]. The third DEAE-cellulose cross-link containing peak (V) separated into two unresolved peaks when refiltered on the Sephadex G-50 column. Since the amino acid composition of peak V was very close to that of peptides IV and VI, it was assumed that the same single peptides were linked together and is not included in this paper because of its heterogeneous nature.

Collagenase Digestion. A typical Sephadex G-50 chromatographic profile of one of the cross-linked three-chained peptides (peptide IV) before and after collagenase digestion is shown in Figure 3. Note that only a single symmetrical peak was obtained with the cross-linked peptide IV, prior to collagenase digestion. This peak had a retention volume of 72 mL while peptide VI had an elution volume of 79 mL. After digestion with collagenase the HHL cross-linked peptide that was derived from peptide IV had a retention volume of 104 mL while the one derived from peptide VI eluted at 108 mL. The other peptide peaks had identical retention volumes in each case (Figure 3). All three peptide chains of peptides IV and VI were cleaved by bacterial collagenase according to the enzyme's classical specificity (see Figure 4). All the Gly triplets containing Hyp NH_2 terminal to residue 84 and COOH terminal to residue 99 in α2CB4 were cleaved from peptide IV to yield C1(IV). No cleavage occurred NH_2 terminal to residue 91 in α2CB4 of peptide VI while cleavage occurred COOH terminal to residue 99 to yield C1(VI) (see Figure 4).

Amino Acid Compositions. The amino acid compositions found for the three-chained HHL peptide IV and the cross-linked peptide, C1(IV), derived after collagenase digestion as well as the compositions obtained from the literature for these peptides are presented in Table I. Table II contains the same information for peptide VI.

Table I: Amino Acid Composition of Peptide IV and Its Collagenase-Derived Product, Peptide C1(IV)

	residues per peptide			
	peptide IV		peptide C1(IV)	
residue	found	lit. value[a]	found	lit. value[b]
Hyp	12 (11.7)[c]	14	0 (0.0)[e]	0
Asp	6 (6.3)	6	4 (4.0)	4
Thr[d]	5 (4.7)	5	0 (0.1)	1
Ser[d]	2 (2.1)	2	2 (1.8)	2
Glu	7 (7.4)	7	3 (3.2)	3
Pro	12 (12.3)	10	5 (5.1)	5
Gly	33 (32.6)	33	11 (11.3)	12
Ala	7 (7.4)	7	1 (1.1)	1
Val	1 (1.2)	1	0 (0.1)	0
Met	1 (0.6)	1	1 (0.8)	1
Ile	1 (0.9)	1	1 (1.0)	1
Leu	7 (7.1)	7	4 (4.1)	4
Tyr	1 (0.9)	1	1 (0.9)	1
Phe	3 (2.8)	3	2 (1.9)	2
His	2 (1.9) + [1][h]	3	2 (1.9) + [1][h]	3
Hyl	3 (2.9) + [1][h]	4	2 (1.8) + [1][h]	2
Lys	0 (0.1) + [1][h]	1	0 (0.1) + [1][h]	1
Arg	4 (4.1)	4	3 (3.1)	3
HHL[e]	1		1	
total[f]	107 [110][g]	110	42 [45][g]	46

[a] Composition of α1CB4-5(76–90) + α1CB6(993–22^{c}) + α2CB4 (76–126) from sequence in Galloway (1982). [b] Composition of α1CB4-5(85–90) + α1CB6(1012–22^{c}) + α2CB4(85–99) from sequence in Galloway (1982). [c] Values in parentheses are actual values found. [d] Not corrected for hydrolysis losses. [e] Histidinohydroxylysinonorleucine; made up of amino acid residues that were originally His, Hyl, and Lys. [f] Totals not in brackets do not include cross-linking residue. [g] Values in brackets are totals found for peptides IV and C1-(IV) with the cross-link as its constituent amino acids. [h] Constituent residue of the cross-link.

Table II: Amino Acid Composition of Peptide VI and Its Collagenase-Derived Product, Peptide C1(VI)

	residues per peptide			
	peptide VI		peptide C1(VI)	
residue	found	lit. value[a]	found	lit. value[b]
Hyp	9 (9.2)[c]	11	0 (0.1)[c]	0
Asp	6 (6.0)	6	4 (4.3)	4
Thr[d]	4 (3.7)	4	0 (0.0)	1
Ser[d]	2 (2.0)	2	2 (1.8)	2
Glu	7 (7.2)	7	3 (3.4)	3
Pro	12 (11.8)	10	5 (5.1)	5
Gly	27 (27.1)	28	9 (9.3)	10
Ala	7 (7.2)	7	1 (1.1)	1
Val	1 (1.2)	1	0 (0.0)	0
Met	1 (0.8)	1	1 (0.7)	1
Ile	0 (0.1)	0	0 (0.1)	0
Leu	6 (6.2)	6	4 (4.3)	4
Tyr	1 (1.1)	1	1 (0.9)	1
Phe	1 (1.1)	1	1 (1.0)	1
His	2 (1.9) + [1][h]	3	2 (1.9) + [1][h]	3
Hyl	2 (2.0) + [1][h]	3	1 (0.9) + [1][h]	1
Lys	0 (0.1) + [1][h]	1	0 (0.0) + [1][h]	1
Arg	3 (3.2)	3	2 (2.1)	2
HHL[e]	1		1	
total[f]	91 [94][g]	95	36 [39][g]	40

[a] Composition of α1CB4-5(76–90) + α1CB6(993–22^{c}) + α2CB4-(91–126) from sequence in Galloway (1982). [b] Composition of α1CB4-5(85–9C) + α1CB6(1012–22^{c}) + α2CB4(91–99) from sequence in Galloway (1982). [c] Values in parentheses are actual values found. [d] Not corrected for hydrolysis losses. [e] Histidinohydroxylysinonorleucine; made up of amino acid residues that were originally His, Hyl, and Lys. [f] Totals not in brackets do not include cross-linking residue. [g] Values in brackets are totals found for peptides VI and C1-(VI) with the cross-link as its constituent amino acids. [h] Constituent residue of the cross-link.

Peptides IV (Table I) had a total of 107 residues and addition (brackets) of the three constituent amino acid residues (His, Hyl and Lys) in the cross-link brought the total amount to 110. The sum of the residues and the composition for

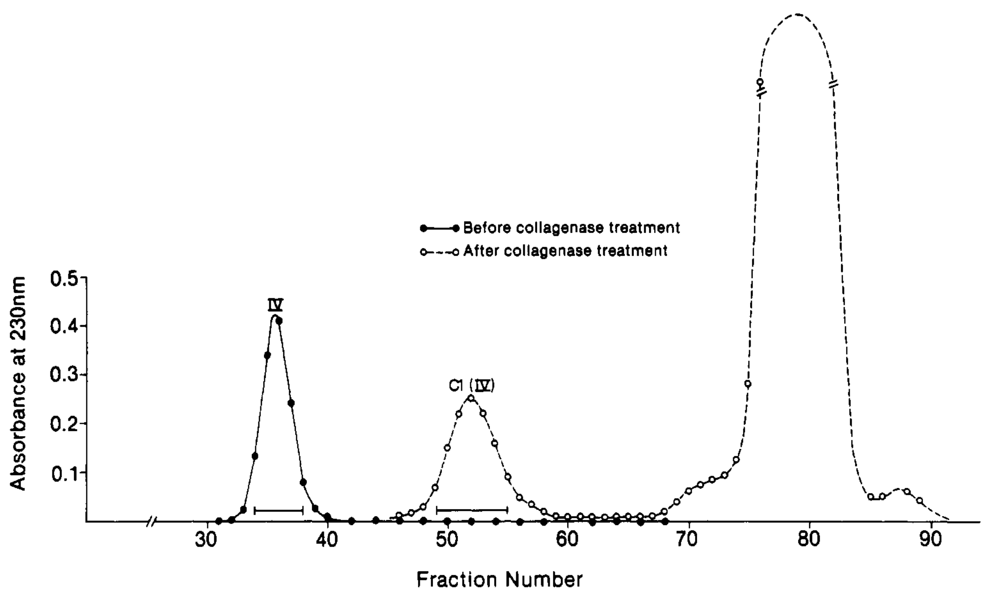

FIGURE 3: Typical gel filtration elution profile of tryptic three-chain peptide. This depicts the gel filtration pattern for peptide IV and its bacterial collagenase digestion products. Original tryptic peptide (●); collagenase digestion products (○). See Materials and Methods for details.

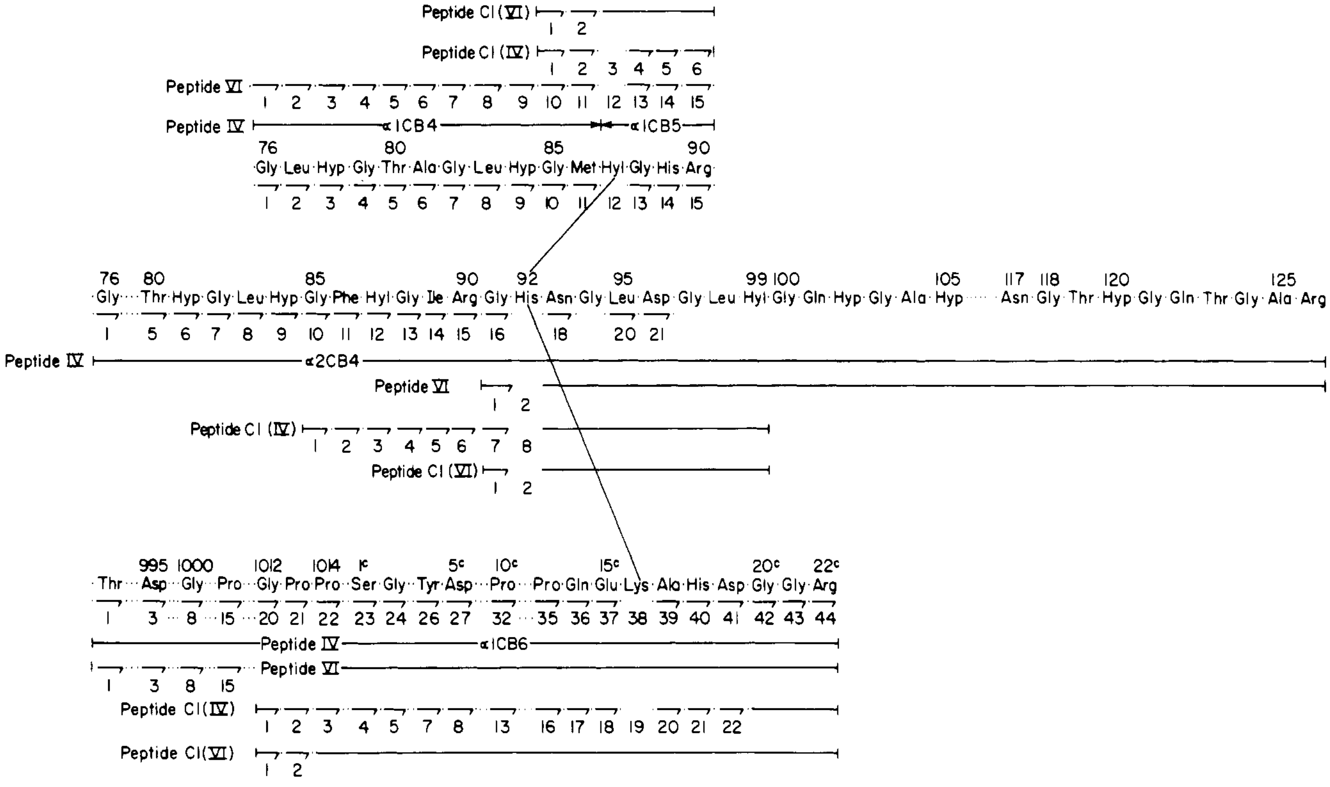

FIGURE 4: Molecular location of HHL. Top sequence of amino acids represents residues 76–90 of the α1(I) chain, middle sequence of amino acids represents residues 76–126 of the α2(I) chain, and bottom sequence of amino acids represents residues 993–22ᶜ of the α1(I) chain. Number above the residue is the residue number in each chain. The half-arrow denotes the actual residue determined in the sequenator with the number just below the arrow representing the sequence cycle. The identification of the peptide sequenced is on the left. Cleavage positions by bacterial collagenase are represented by the vertical bars at the ends of each line representing peptides C1(IV) and C1(VI). The diagonal lines signify the residues that are linked to form the trifunctional cross-link. See Results for other details. Hyl at position 99 in this diagram has been substituted for the Thr-99 reported in the literature (Galloway, 1982) in order to correct the sequence.

peptide IV were equivalent to the sum and the constituent amino acids for the fragments α1CB4-5(76–90) + α1CB6-(993–22ᶜ) + α2CB4(76–126). In the same manner α1CB4-5(76–90) + α1CB6(993–22ᶜ) + α2CB4(91–126) was apparently equal to the amino acid residue composition of peptide VI (Table II). The same was also found for the relationship between α1CB4-5(85–90) + α1CB6(1012–22ᶜ) + α2CB4-(85–99) and peptide C1(IV) (Table I) as well as for α1CB4-5(85–90) + α1CB6(1012–22ᶜ) + α2CB4(91–99) and peptide C1(VI) (Table II) (see Discussion).

The amino acid compositions found for all the peptides are very similar in nature to the constituent amino acids obtained from collagen chain sequences reported in the literature (Galloway, 1982), except for minor differences. The Hyp and Pro content for peptide IV (Table I) was 12 residues for each, while the amounts in the literature are 14 and 10, respectively.

Table III: Typical Sequence Data Obtained for Three-Chained Cross-Linked Peptides

cycle	peptide IV, residues and amounts found[a]
1	Gly (off scale), Thr (nq)
2	Leu (19.5), Phe (18.4), Gly (nq)
3	Hyp (nq), Asp (9.7)
4	Gly (11.5), Ala (12.2)
5	Thr (nq), Gly (nq)
6	Ala (>13), Hyp (nq), Pro (nq)
7	Gly (nq), Ala (>13)
8	Leu (14.8), Gly (nq)
9	Hyp (nq), Pro (nq)
10	Gly (nq), Hyp (nq)
11	Met (11.0), Phe (12.2), Gly (nq)
12	Hyl (nq), Pro (nq)
13	Gly (nq), Hyp (nq)
14	His (2.7), Ile (6.8), Gly (nq)
15	Arg (3.0), Pro (nq)
16	Gly (nq), Hyp (nq)
17	Gly (nq)
18	Asn (0.43 + 0.58 Asp), Pro (nq)
19	Hyp (nq)
20	Leu (0.50), Gly (nq)
21	Asp (nq), Pro (nq)

[a] Values in parentheses are nanomoles, and nq = not quantified.

However, the total amounts of imino acid residues, 24, are the same (Table I) as reported previously.

The total imino acid residue content for peptide VI was found to be 21. However, while the values for Hyp and Pro previously documented are 11 and 10, respectively, we found values of 9 and 12, respectively. It was found that peptide VI contained one less residue of Gly.

The variances from the literature noted above for Hyp and Pro are probably due to differences in hydroxylation of Pro.

The variances from the literature we found were for peptides C1(IV) and C1(VI); both contained one less Gly, one more Hyl, and no Thr. Thr has been reported to be present in the α2(I) chain containing portion of the cross-linked peptide at position 99. The data we obtained and the literature values for the above peptides of interest are presented in Tables IV and V.

Amino Acid Sequences. Typical raw data for sequence cycles obtained for one of the three-chained cross-linked peptides sequenced are presented in Table III. Specifically, Table III contains the data obtained for the first 21 cycles of peptide IV, and along with its amino acid composition (Table I), these data were sufficient to unequivocally identify the three peptides linked by the trifunctional cross-link, HHL. Figure 4 represents a summary of the sequence data and the sequences for peptides IV, VI, C1(IV), and C1(VI).

A total of 44 sequence cycles were successfully completed for peptide IV. Except for the cross-linking residue, Lys-16^{c}, sequence cycle 38, the phenylthiohydantoin derivatives for the entire sequence were obtained for the α1CB6(993–22^{c}) branch of the three-chained peptide. Likewise, the complete sequence for the portion of the peptide represented by α1CB4-5(76–90) was obtained, except for the cross-linking residue, Hyl-87, at sequence cycle 12. Only 21 sequence cycle derivatives were obtained for the third member of the cross-linked peptide, which represented residues 76–96 of the α2 chain. No derivatives were found for the cross-linking residue, His-92, and for Gly-94, which were sequence cycles 17 and 19, respectively. The yields of the derivatives were reduced considerably after sequence cycle 16 and were barely detectable subsequent to cycle 21 at which point sequencing effectively ceased for the α2 chain portion of the three-chained peptide. The blanks at cycles 12, 17, and 38 defined the cross-linking sites. The above data are presented in Figure 4. From the amino acid composition and the sequence data, we propose that the structure for peptide IV is α1CB4-5(76–90)[Hyl-87][2] × α1CB6(993–22^{c})[Lys-16^{c}] × α2CB4(76–126)[His-92].

The phenylthiohydantoin derivatives for the first 15 sequence cycles of a total of 44 were determined for peptide VI. Identical results were obtained as those that were obtained for two of the branches of peptide IV, those being α1CB6(993–22^{c}) and α1CB4-5(76–90). Only the first residue (Gly-91) of the α2 chain portion of the peptide was observed. Sequencing of this chain stopped at the second cycle (Figure 4). However, these data along with the amino acid composition allowed us to assign the structure of peptide VI as α1CB4-5(76–90)[Hyl-87] × α1CB6(993–22^{c})[Lys-16^{c}] × α2CB4-(91–126)[His-92].

A total of 25 sequence cycles were performed on peptide C1(IV), the three-chained collagenase peptide product of peptide IV. However, only derivatives for the first six cycles were examined chromatographically, which were sufficient for the identifications required. The amino acid composition (Table I) along with the sequence results (Figure 4) indicated that α1(I) chain residues 85–90 and 1012–22^{c}, as well as residues 85–99 of an α2 chain, were part of the cross-linked peptide. The structure of this peptide was apparently α1CB4-5(85–90)[Hyl-87] × α1CB6(1012–22^{c}) × α2CB4-(85–99)[His-92].

Only two sequence cycles were examined for peptide C1-(VI), the three-chained collagenase digestion product of peptide VI. The only derivative found in the first cycle was for Gly in an amount that corresponded to three residues for the amount of peptide sequenced. The second cycle yielded Met and Pro as the sole derivatives. These data along with the amino acid composition and the sequencing data for the other peptides were sufficient to identify the portions of the three peptides that constituted the three-chained peptide, peptide C1(VI). The only choice for the peptide's structure was α1CB4-5(85–90)[Hyl-87] × α1CB6(1012–22^{c})[Lys-16^{c}] × α2CB4(91–99)[His-92].

DISCUSSION

On the basis of the structure Housley et al. (1975) proposed for the trifunctional cross-link, Yamauchi et al. (1982) attempted to cleave a three-chained cross-link peptide by specifically degrading the cross-link at its proposed enamine double bond using osmium tetroxide and sodium metaperiodate. An anomalous unknown degradation occurred, and as a result, unexpectedly low yields of cleavage products were obtained. However, the correct identities for the two peptide chains, linked by a new chemically produced unknown bifunctional cross-link, were obtained. The identification and structure of the third chain eluded the latter authors as it did Becker et al. (1975), previously, and this stimulated the current investigation. This paper elucidates the structure of the third chain as well as the locus for the His residue that contributes its prosthetic group to the trifunctional cross-link.

Sequence analysis demonstrated that the same two portions of the α1(I) chains were present in peptides IV and VI. However, sequencing halted at the second cycle for the α2(I) chain portion of the three-chained peptide, peptide VI. Along with the amino acid composition, this indicated that trypsin

---

[2] Designations of the cross-linked peptides in this paper are made in terms of the CNBr fragments obtained from collagen; numbers in parentheses correspond to the residue number assigned in the compiled sequences in Galloway (1982); residues in the brackets are the residues that participate in the cross-link.

Table IV: Amino Acid Composition Differences in α2CB4(85–99)[a] Portion of Three-Chained Peptide C1(IV)

	residues per peptide					
	peptide C1(IV)			4, column 2 minus column 3, α2(85–99)[a]	5, column 1 minus column 3, α2(85–99)[a]	6, column 5 minus column 4
residue	1,[b] found	2,[c] lit. value	3,[d] lit. value			
Hyp	0	0	0	0	0	0
Asp	4	4	2	2	2	0
Thr	0	1	0	1	0	-1
Ser	2	2	2	0	0	0
Glu	3	3	3	0	0	0
Pro	5	5	5	0	0	0
Gly	11	12	7	5	5 (4)[e]	0 (-1)[e]
Ala	1	1	1	0	0	0
Val	0	0	0	0	0	0
Met	1	1	1	0	0	0
Ile	1	1	0	1	1	0
Leu	4	4	2	2	2	0
Tyr	1	1	1	0	0	0
Phe	2	2	1	1	1	0
His	3[f]	3	2	1	1	0
Hyl	3[f]	2	1	1	2	1
Lys	1[f]	1	1	0	0	0
Arg	3	3	2	1	1	0
total	45	46	31	15	15 (14)[e]	0 (-1)[e]

[a] Numbers in parentheses of CB peptides and α chains represent residue numbers in chains of collagen. [b] Composition of collagenase-derived three-chained cross-linked peptide C1(IV) determined in Table I. [c] Literature values for sum of amino acid residues in α1CB4-5(85–90)[a] + α1CB6(1012–22[c])[a] + α2CB4(85–99)[a] (Galloway 1982). [d] Literature values for sum of amino acid residues in α1CB4-5(85–90)[a] + α1CB(1012–22[c])[a]. [e] Values in the parentheses are the residues found for the differences in Gly in Table I. [f] One residue has been added to the actual number found to account for its participation in the trifunctional cross-link histidinohydroxylysinonorleucine (see Table I).

Table V: Amino Acid Composition Differences in α2CB(91–99)[a] Portion of Three-Chained Peptide C1(VI)

	residues per peptide					
	peptide C1(VI)			4, column 2 minus column 3, α2(91–99)[a]	5, column 1 minus column 3, α2(91–99)[a]	6, column 5 minus column 4
residue	1,[b] found	2,[c] lit. value	3,[d] lit. value			
Hyp	0	0	0	0	0	0
Asp	4	4	2	2	2	0
Htr	0	1	0	1	0	-1
Ser	2	2	2	0	0	0
Glu	3	3	3	0	0	0
Pro	5	5	5	0	0	0
Gly	9	10	7	3	3 (2)[e]	0 (-1)[e]
Ala	1	1	1	0	0	0
Val	0	0	0	0	0	0
Met	1	1	1	0	0	0
Ile	0	0	0	0	0	0
Tyr	1	1	1	0	0	0
Phe	1	1	1	0	0	0
His	3[f]	3	2	1	1	0
Hyl	2[f]	1	1	0	1	1
Lys	1[f]	1	1	0	0	0
Arg	2	2	2	0	0	0
total	39	40	31	9	9 (8)[e]	0 (-1)[e]

[a] Numbers in parentheses of CB peptides and α chains represent residues numbers in chains in collagen. [b] Composition of collagenase-derived three-chained cross-linked peptide C1(VI) determined in Table II. [c] Literature values for sum of amino acid residues in α1CB4-5(85–90) + α1CB6(1012–22[c]) + α2CB4(91–99) (Galloway, 1982). [d] Literature values for sum of amino acid residues in α1CB4-5(85–90) + α1CB6(1012–22[c]). [e] Values in parentheses are the residues found for the differences in Gly in Table II. [f] One residue has been added to the actual number found to account for its participation in the trifunctional cross-link histidinohydroxylysinonorleucine (see Table II).

cleavage occurred COOH terminal to Arg-90 in the α2(I) chain producing a peptide 15 residues less than peptide IV, which is consistent with its lower molecular weight (Figure 3). Trypsin digestion was apparently incomplete (yielding peptide IV), possibly due to the proximity of the cross-link, which in this case was the penultimate residue to the NH_2-terminal (Gly-91) cleavage site. The failure to sequence beyond Gly-91 (peptide VI) may be attributed to the derivatized imidazole ring of His-92. Gibson et al. (1980) have reported that blockage of the Edman degradation can occur at imidazole-derivatized His residues. This halt in sequencing did not occur with peptide IV. One might speculate that since peptide VI is smaller than peptide IV, sequencing was at different stages in the α1(I) chains when residue 92 was derivatized. At this stage both α1(I) chains of peptide IV each contained 16 residues less. Therefore, peptide VI was larger when residue 92 was reached. The substantially decreased derivative yields for the α2CB4 portion of peptide IV at cycle 18, with no yield at cycle 19, and then small yields at cycles 20 and 21 are probably due to the Asn–Gly peptide bond between cycles 18 and 19. This sequence corresponds to residues 93 and 94 of the α2(I) chain, which was reported earlier to essentially block the Edman degradation for collagen by formation of significant amounts of β-aspartyl peptide bonds (Bornstein & Balian, 1977; Jörnvall, 1974). In this paper the α2CB4 portion of peptide IV at this stage in the sequencing most likely formed partial β-aspartyl peptide bonds, thereby decreasing the derivative yields.

A sequence inversion in the α1CB6(5^c–22^c) portion of the two-chained cross-linked peptide that was isolated as a degradation product of the three-chained peptide was reported (Yamauchi et al., 1982). It was thought that the peptide residues 13^c–19^c had the apparent sequence -Gly-Gln-Glu-Glu-Ala-Lys(18^c)-His-, while the sequence reported by Rauterberg et al. (1972) was -Gly-Gln-Gln-Glx-Lys(17^c)-Ala-His-. However, the sequence now obtained is -Gly-Gln-Glu-Lys-(16^c)-Ala-His (Figure 4), which is consistent with amino acid compositions (see Tables I and II) and with results obtained by nucleotide sequencing (Fuller & Boedtker, 1981). We attribute the previous discrepancy to carry over errors of the derivatives of Glu-15^c and His-89 in the other α1(I) chain branch of the peptide in the presence of large amounts of the derivative for Ala at residue 17^c in the same cycle.

The complete sequencing of the α1CB4-5(76–90) and the α1CB6(993–22^c) peptide portions of the three-chained cross-linked peptides (Figure 4) are in complete agreement with the literature for these portions of the α1(I) chain of type I bovine skin collagen. The one exception was for residue Glu-15^c; this is discussed above. Considering these data, we have obtained a significant deviation with the literature in the amounts of Hyl and Thr present in the α2CB4(85–99) and α2CB4(91–99) peptide portions of the respective three-chained cross-linked peptides C1(IV) and C1(VI) (Tables IV and V). We unequivocally found by amino acid analyses (Tables I and II) one residue of the cross-link HHL in each peptide, two residues of Hyl for the former peptide, and one residue of Hyl in the latter one, while no Thr was found in either peptide. According to the composition obtained from the sequence compiled by Galloway (1982) and the sequence for α2CB4 published previously (Fietzek & Rexrodt, 1975), there should have been one residue of Thr and one less residue of Hyl than we found in each peptide. We are certain of our findings because we placed sufficient amounts of the respective peptide hydrolysates on the amino acid analyzer to unequivocally identify and quantify Hyl and the other amino acids in multi-nanomolar amounts. Thus, the Hyl content of the three peptides that constitute peptide C1(IV) is three residues (Table IV) and two residues for the three peptides that make up peptide C1(VI) (Table V) with the absence of Thr from both peptides. These data in conjunction with the extreme difficulty in sequencing α2CB4 past the Asn–Gly at residue positions 93 and 94 as found here and in an earlier report (Bornstein & Balian, 1977) allow us to propose the following alterations for the sequence of the α2(I) chain of bovine skin collagen. We suggest that Hyl be substituted for Thr at residue position 99 and Thr be substituted for Hyl at position 123. This sequence change would not alter the compositions of the isolated three-chained peptides (peptides IV and VI, Tables I and II), and they would still conform to the amino acid compositional data found and with that derived from the sequence literature (Galloway, 1982). The fact that we were able to obtain the relatively smaller collagenase-derived peptides C1(IV) and C1(VI) for analysis allowed us to obtain more accurate results than were previously available for this region of the α2(I) chain. The results obtained (see Figure 4) are concordant with the specificity of bacterial collagenase cleavage (Oshima et al., 1979) and supports this suggested sequence correction. In addition, the basic amino acids Lys and Arg are in homologous positions for the first 293 helical residues of both α2(I) and α1(I) chains of bovine collagen except for Thr-99 and Hyl-123 in the α2(I) chain (Galloway, 1982). However, Arg or a hydroxylated Lys can occur instead of a Lys in the α2(I) chain at a position where a Lys is present in the α1(I) chain. The latter is the case for Arg-252 in the α2(I) chain, which does not seem exceptional because both are basic amino acids. The correction to the sequence that is suggested here would place a Hyl at a position in the α2(I) chain that would be homologous with the Lys residue at position 99 in the α1(I) chain. This would produce a closer homology between the α1(I) and α2(I) chains of bovine skin collagen than that which presently exists. Unfortunately, we cannot compare our suggested positional changes with human, rat, or chicken skin collagen since the sequences in this area of the α2(I) have not been reported as yet. It is most probable that the sequences have not been obtained because of the extreme difficulty found in sequencing this portion of the α2(I) chain (see above).

The only types of heterogeneity that have been reported to exist in collagen are due to the now well-known phenomena of partial hydroxylation that may take place for Pro and Lys. It is highly unlikely that a heterogeneity exists involving the substitution of Hyl for Thr at position 99 in the α2(I) chain for insoluble as opposed to soluble collagen. The original sequence (Fietzek & Rexrodt, 1975) was reported for soluble bovine skin type I collagen, and in this paper we have used insoluble skin. This type of heterogeneity would require an additional structural gene to be present for the α2(I) chain, and none has been reported.

The three-chained peptides isolated in this paper showed no indications of the chymotryptic-like cleavages previously encountered (Yamauchi et al., 1982). This was probably due to the different lots of TPCK-trypsin that were used. It has been well established that even TPCK-treated trypsin contains significant chymotryptic activity (Jany et al., 1976). It is not unlikely that different lots of TPCK-trypsin from the same supplier would contain dissimilar amounts of chymotrypsin.

The COOH-terminal cross-linking in type I skeletal collagens involves intermolecular iminium bond formation between the Lys^{ald}-16^c (or Hyl^{ald}-16^c) residue of an α1(I) chain on one molecule to a Hyl-87 of either an α1(I) or an α2(I) chain on another molecule (Yamauchi et al., 1986). Consequently, the cross-linked molecules have to be longitudinally staggered with respect to one another by about 932 (or minus 233) residues. Such a bond can "mature" into a two-molecule, three-chain pyridinoline cross-link by incorporating a COOH-terminal Hyl^{ald}-16^c from another α1(I) chain (Robins & Duncan, 1983). This three-chain cross-link thus contains one residue from a helical region and two residues from nonhelical peptide regions (see Figure 1). It is especially significant that it is not detectable in skin (Eyre et al., 1984). The three-chain cross-link isolated from skin involves Lys^{ald}-16^c on one α1(I) chain, a His-92 on an α2(I) chain, and a Hyl-87 on yet another α1(I) chain. The chemistry of HHL three-chained peptides clearly establishes that it derives only from type I collagen chains and, also, that the aldehyde donor has to be staggered by about 932 residues with respect to the helical components. However, the especially significant features of HHL is that (1) it contains two residues from helical regions and only one residue from a nonhelical peptide region and (2) it is not found in skeletal tissues (Yamauchi, unpublished data). These findings indicate that the steric relationships between the α chains in skin and skeletal collagen fibrils are fundamentally different. There are only two stereochemical possibilities that could explain HHL formation. HHL cross-links could tie either two or three molecules together. As will be discussed below, if the cross-link involved only two molecules, a skin collagen molecule would have to have a unique structure. If it involved three molecules,

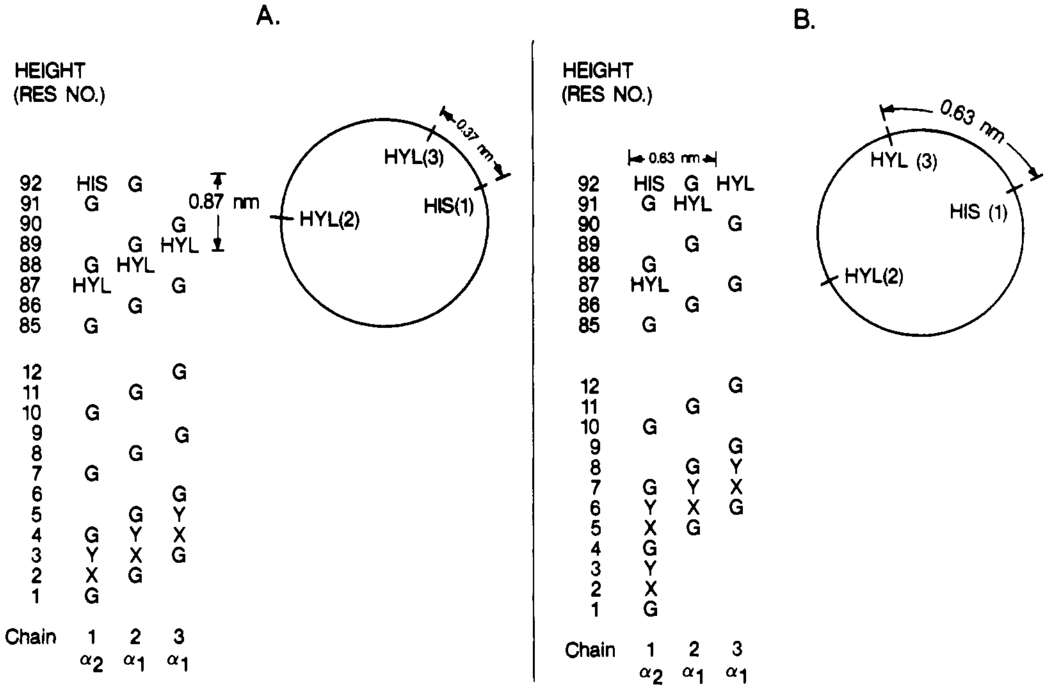

FIGURE 5: Illustration of the effect of chain initiation on the spatial positions of α1(I) Hyl-87 and α2(I) His-92 in an α2(I), α1(I), and α1(I) collagen molecule. (A) The triple helix is initiated by the first triplet of the α2(I) helical sequence. The height along the molecule is measured in units of number of residues, starting with residue 1 of the α2(I) chain. Each number also represents the 0.29-nm rise of the triple helix. The right-hand scheme looks down the axis of the molecule (from COOH to NH_2 terminus). The circle represents the possible radial locations, from the triple-helical axis, of the β-carbon atoms of residues in X and Y positions (approximately 0.5 nm). The angular locations of the Hyl-87 on chains 2 and 3, respectively, and His-92 on chain 1 were calculated from the helical coordinates of Yonath and Traub (1969). An HHL linkage, which would be only 0.68 nm long in this configuration, cannot form as the respective helical residues are 0.95 nm apart. (B) The triple helix is initiated by the second triplet of the α2(I) sequence. The β-carbon atoms for Hyl-87 and His-92 are now only 0.63 nm part, permitting the helical residues on one molecule to participate in HHL formation.

then the intermolecular relationships in skin fibrils would have to be specific for skin. We believe the evidence supports this latter case.

Collagen molecules are conceptualized as having three domains, a major central helix and two small NH_2- and COOH-terminal nonhelical peptide regions. The criterion for helicity is primarily a triplet (Gly-X-Y) repeating sequence and, secondarily, lack of enzymatic susceptibility to the usual proteases. In three dimensions, helically equivalent residues are related to each other by the symmetry operation of a 108° rotation and a rise of 0.29 nm (Yonath & Traub, 1969; Katz, 1970). We do not know which α chain initiates the triple helix nor which of the Gly residues of the sequence are related by the symmetry condition. The conventional view is that the first triplets in the respective α chain sequences are equivalent (Galloway, 1982). This is illustrated in Figure 5 for a molecular structure initiated by an α2(I) chain (i.e., α2, α1, and α1). We have used this structure as an illustration because it has the shortest distance between the residues of interest [α1(I) Hyl-87 and α2(I) His-92]. Even so, this structure would not allow HHL to form. CPK space-filling atomic models indicate that the distance between β-carbon atoms of the helical residues of HHL, when in a fully extended unstrained conformation, is 0.93 nm. However, HHL cannot take up this conformation when the β-carbon atoms are in the same molecule because, then, the cross-link must wind about the cylindrical surface of the collagen molecule. This latter configuration limits the maximal β-carbon atoms' distance between the participating helical residues in a single molecule to only 0.68 nm, which is too small, by about 0.27 nm, for HHL formation (Figure 5). Such a cross-link would be possible within a molecule if it deviated locally from the triple-helical conformation in a major way. We would then have to ask why such a localized conformational perturbation is not present in skeletal type I collagen molecules, allowing HHL formation in these tissues as well. One possible explanation for HHL tissue specificity is that the organization of the chains in the skin collagen molecule is distinctive. A two-molecule HHL link would be possible, for example, if the first triplets of the α chain sequences were not structurally equivalent. By offsetting the registry of the α chains by a triplet, the side chains of Hyl-87 [from the 2nd α1(I) chain] and His-92 [α2(I) chain] are brought into positions to form HHL (see Figure 5B). There is no experimental evidence that rules out such a possibility.

On the other hand, there are X-ray diffraction findings that are most compatible with a three-molecule HHL cross-linking pattern in skin fibrils. The axial period for skeletal collagen fibrils, while in a physiological hydration state, is 67 nm. In skin this drops to 65 nm with a concomitant increase in the "disorientation" of the equatorial scattering (Stinson & Sweeny, 1980; Brodsky et al., 1980). These findings indicate a greater degree of tilt or bend in the skin molecules. This effect does not appear to be intrinsic to the molecule; reconstituted type I skin collagen fibrils have the 67-nm axial periodicity (Brodsky et al., 1980; Li et al., 1981).

A three-molecule HHL cross-link could produce these X-ray diffraction effects. In order for HHL to form, the helical residues would have to be in nearest-neighbor molecules (having an interaxial distance of about 1.5–1.7 nm) that are displaced axially with respect to one another by about 0.4 nm (see Figure 6). This latter distance is the approximate axial intermolecular displacement that brings an Hyl-87 of an α1(I) chain on one molecule to within reaction distance to a His-92 on a nearest-neighbor molecule α2(I) chain. Using a combination of CPK space-filling models and calculations, we have found that this angle is relatively invariant to the estimate of the intermolecular distance in skin fibrils, as it is determined

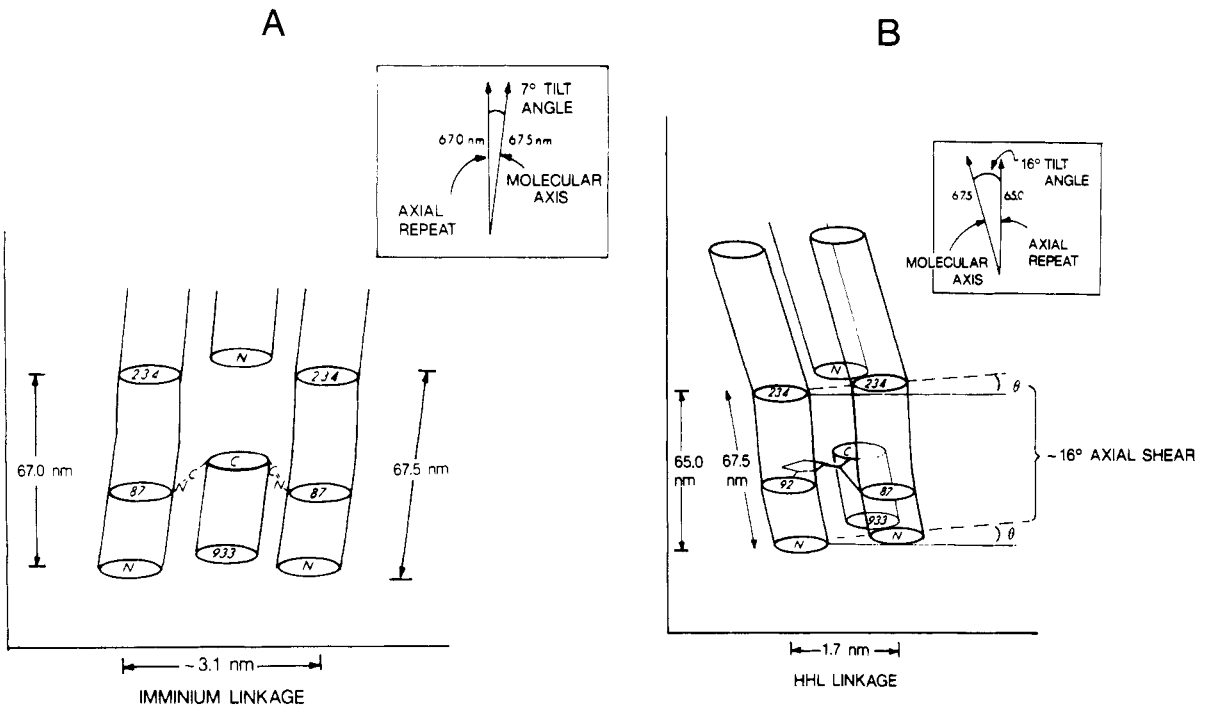

FIGURE 6: Three-dimensional models for the COOH-terminal intermolecular cross-linking geometry for type I skeletal and skin collagens. The collagen molecule is represented as a flexible cylinder. The numbers represent the height from the first amino-terminal helical residue of an $\alpha2$, $\alpha1$, and $\alpha1$ molecule with each increment also corresponding to a helical rise of 0.29 nm. Note that the molecules contributing helical residues have a nonstaggered (OD) relationship. (A) Skeletal collagen fibrils. The inset illustrates that molecular segments are not precisely oriented along the fibril axis. The average flex-tilt of the molecules illustrated in the inset is estimated from the axial repeat distance of 67 nm observed for skeletal collagens and the 67.5-nm stagger distance corresponding to a 233-residue stagger. The intermolecular distance for OD staggered molecules in skeletal fibrils is 2.5–3.1 nm, the latter being the estimate for bone (see Figure 1). (B) Skin collagen fibrils. A three-molecule cross-link of conventional $\alpha2$, $\alpha1$, and $\alpha1$ molecules requires that the molecules containing the helical residues of HHL be nearest neighbors having an intermolecular distance of about 1.7 nm (see Figure 1), longitudinally staggered by about 0.4 nm. The longitudinal displacement raises the Hyl-87 of chain 2 on one molecule to within reaction distance of the His-92 on chain 1 of a neighboring molecule. This kind of displacement is an axial shear, having a shear angle of about 16°. If the molecular segments were oriented normal to the plane passing through the NH_2-terminal residues, as in (A), they would be flex tilted by about 16° with respect to the fibril axis. The inset illustrates that the average axial flex-tilt angle in skin is indeed about 16°. This was calculated with the 65-nm axial period for skin collagen reported by Brodsky et al. (1980) and Stimson and Sweeney (1980).

by the ratio of that minimal axial displacement required for HHL formation which is compatible with a particular intermolecular distance, and these quantities are directly correlated. If the collagen molecules were to align themselves normal to the plane passing through the NH_2-terminal residues, as is the case with "skeletal" collagens of Figure 6A, they would be tilted 16° with respect to the fibril axis. This would produce the 65-nm period and a "sheared" equatorial scattering in the X-ray diffractograms from hydrated skin. It would also account for the 65-nm period and 18° off-axial tilt of the striations observed in electronmicrographs of freeze-fractured skin fibrils (Ruggerri et al., 1979).

Currently we cannot conclude with certainty that Figure 6B describes a unique geometry for HHL formation. However, the HHL content in both adult male and bovine approaches a content of one cross-link per type I collagen molecule (Yamauchi, submitted for publication). This indicates that the essentially unstaggered nearest-neighbor relationship between the molecular segments contributing the His-92 and Hyl-87 residues to the cross-link is the norm rather than the exception and that it is unlikely that HHL formation in type I skin collagen is a result of fluctuations brought about by thermal motion of collagen molecules in a fibril (Torchia, 1982) having the skeletal collagen structure. A disordered packing has been suggested for the packing in the gap regions of fibrils (Piez & Trus, 1981). As the cross-linking occurs adjacent to this region, some configurations like that of Figure 6B might be expected in skeletal fibrils. However, more recently, crystalline packing in the gap region of rat tail tendon fibrils has been demonstrated (Chew & Squire, 1986). This aids to explain why no HHL at all can be detected in those type I collagen fibrils, in which all nearest-neighbor molecules are staggered by distances equal to some multiple of 67 nm (Katz & Li, 1973a; Hulmes & Miller, 1978). In addition, our initial findings on the stereochemistry of a tetrafunctional cross-link, deH-HHMD, occurring at the NH_2-terminal end of the gap regions of the skin fibrils, indicate that, here too, two helical residues (unpublished data), on nonstaggered nearest-neighbor molecules, participate in the cross-link. This is compatible with the configuration of Figure 6B but not with that Figure 6A. Finally, the nearest-neighbor intermolecular stabilization energies associated with fibrillar arrays systematically having some nonstaggered molecules are comparable to those having only staggered molecules (Golub & Katz, 1977). This means that there is no apriori distinction between a skeletal and skin fibril structure on an energy basis.

Registry No. HHL, 108212-74-4.

REFERENCES

Becker, U., Furthmayr, H., & Timpl, R. (1975) *Hoppe-Seyler's Z. Physiol. Chem. 356*, 21.

Bernstein, P. H., & Mechanic, G. L. (1980) *J. Biol. Chem. 255*, 10414.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132.

Brodsky, B., Eikenberry, E. F., & Cassidy, K. (1980) *Biochim. Biophys. Acta 621*, 162.

Chew, M. W. K., & Squire, S. M. (1986) *Int. J. Biol. Macromol. 8*, 27.

Eyre, D. R., Koob, T. J., & Van Ness, K. P. (1984) *Anal. Biochem. 137*, 380.

Fietzek, P. P., & Rexrodt, F. W. (1975) *Eur. J. Biochem. 59*, 113.

Fukae, M., & Mechanic, G. L. (1980) *J. Biol. Chem. 255*, 6511.

Fuller, F., & Boedtker, H. (1981) *Biochemistry 20*, 999.

Galloway, D. (1982) in *Collagen in Health and Disease* (Weiss, J. B., & Jayson, M. I. V., Eds.) p 528, Churchill Livingston, New York.

Gibson, D. R., Gracy, R. W., & Hartman, F. C. (1980) *J. Biol. Chem. 255*, 9369.

Golub, E., & Katz, E. P. (1977) *Biopolymers 16*, 1357.

Housley, T., Tanzer, M. L., Hensen, E., & Gallop, P. M. (1975) *Biochem. Biophys. Res. Commun. 67*, 824.

Hulmes, D. J. S., & Miller, A. (1971) *Nature (London) 293*, 239.

Jany, K. D., Keil, W., Meyer, H., & Kiltz, H. H. (1976) *Biochim. Biophys. Acta 453*, 62.

Jörnvall, H. (1974) *FEBS Lett. 38*, 329.

Katz, E. P. (1970) *Biopolymers 9*, 745.

Katz, E. P., & Li, S. T. (1973a) *J. Mol. Biol. 73*, 351.

Katz, E. P., & Li, S. T. (1973b) *J. Mol. Biol. 80*, 1.

Li, S. T., Sullman, S., & Katz, E. P. (1981) in *The Chemistry and Biology of Mineralized Tissues* (Veis, A., Ed.), pp 69–73, Elsevier/North-Holland, Amsterdam.

Mechanic, G. L. (1974) *Biochem. Biophys. Res. Commun. 56*, 923.

Oshima, G., Shimabukuro, H., & Nagasawa, K. (1979) *Biochim. Biophys. Acta 567*, 392.

Piez, K. A., & Trus, B. L. (1981) *Biosci. Rep. 1*, 801.

Rauterberg, J., Fietzek, P. P., Rexrodt, F. W., Becker, U., Stark, M., & Kuhn, K. (1972) *FEBS Lett. 21*, 7.

Robins, S. P., & Duncan, A. (1983) *Biochem. J. 215*, 175.

Ruggerri, A., Benazzo, F., & Reale, E. (1979) *J. Ultrastruct. Res. 68*, 101.

Stinson, R. H., & Sweeny, P. R. (1980) *Biochim. Biophys. Acta 621*, 158.

Tanzer, M. L., Housley, T., Berube, L., Fairweather, R. Franzblau, C., & Gallop, P. M. (1973) *J. Biol. Chem. 248*, 393.

Torchia, D. A. (1982) *Methods Enzymol. 82A*, 174.

Yamauchi, M., Noyes, C., Kuboki, Y., & Mechanic, G. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7684.

Yamauchi, M., Katz, E. P., & Mechanic, G. L. (1986) *Biochemistry 25*, 4907.

Yamauchi, M., London, R. E., Guenat, C., Hashimoto, F., & Mechanic, G. L. (1987) *J. Biol. Chem.* (in press).

Yonath, A., & Traub, W. (1969) *J. Mol. Biol. 43*, 465.

Octopus dofleini Hemocyanin: Structure of the Seven-Domain Polypeptide Chain[†]

Josette Lamy,[‡] Michèle Leclerc,[‡] Pierre-Yves Sizaret,[‡] Jean Lamy,[‡] Karen I. Miller,[§] Reginald McParland,[§] and K. E. van Holde*[,§]

Laboratoire de Biochimie, Faculté de Pharmacie, 37032 Tours Cédex, France, and Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331

Received October 9, 1986; Revised Manuscript Received December 30, 1986

ABSTRACT: *Octopus dofleini* hemocyanin exists as a native molecule composed of 10 identical 350-kDa polypeptide chains. It is homogeneous by sedimentation methods, by N-terminal sequence homology, and by the presence of a single immunoprecipitation peak, corresponding to the subunit. Electron microscopy shows seven globular domains, and there are seven immunologically distinct products (*Od* 1 to *Od* 7) following limited proteolysis. By combining the results of proteolysis and immunological methods with amino acid sequencing, the order of the domains within the subunit has been established with one remaining uncertainty. Trypsin and serine protease from *Staphylococcus aureus* produce fragments corresponding to domain *Od* 1 and the six-domain fragment (2,3,4,5,6,7). The amino acid sequence and the digestion conditions for domain *Od* 1 are the same as those that Takagi [Takagi, T. (1986) in *Invertebrate Oxygen Carriers* (Linzen, B., Ed.) pp 259–262, Springer-Verlag, Berlin and Heidelberg] found for the C-terminal domain of *Paroctopus* hemocyanin. Thus, the sequence becomes N-(2,3,4,5,6,7)-1-C. Trypsin, subtilisin, and other enzymes produce fragments (1,2,5) and (3,4,7). This information permits ordering the domains additionally as follows: N-[(3,4,7),6]-(2,5)-1-C. α-Chymotrypsin produces fragments (1,2,5,6) and (2,5,6), allowing as the next step this sequence: N-(3,4,7)-6-(2,5)-1-C. Digestion with subtilisin produces the unique fragment (4,7), and a secondary digestion of fragment (3,4,7) by trypsin yields the fragment (3,4). The sequence now is N-(3-4-7)-6-(2,5)-1-C. Amino acid sequencing demonstrates that fragment (4,7) has the same N-terminal sequence as the whole subunit. Thus, the sequence now becomes N-7-4-3-6-(2,5)-1-C. The uncertainty of the order of the (2,5) fragment remains.

Molluscan hemocyanins are composed of 10 or 20 subunits of 350–450 kDa arranged to form hollow cylinders with molecular weights of approximately 3.5×10^6 to 9.0×10^6. Most of the information on the structure of molluscan hemocyanin has been obtained from studies of *Helix pomatia* and other gastropods (Ghiretti-Magaldi et al., 1981; van Bruggen et al., 1981a; van Holde & Miller, 1982; Lontie, 1983). These studies have revealed that gastropod hemocyanins are composed of a mixture of several different subunits, a fact that has made them difficult to study, because their association reactions are complex (Brouwer et al., 1979). One must sort out individual subunit types and determine their role in the

---

[†] This work was supported by RCP 080816 of Centre National de la Recherche Scientifique (CNRS) (J.L.) and Grant DMB 85 17310 from the National Science Foundation (K.E.v.H. and K.I.M.).

[‡] Faculté de Pharmacie.

[§] Oregon State University.

0006-2960/87/0426-3509$01.50/0