Biochem. J. (1992) 283, 129–136 (Printed in Great Britain)

Identification of the loci of the collagen-associated Ehrlich chromogen in type I collagen confirms its role as a trivalent cross-link

Ronald KUYPERS,* Margaret TYLER,† Lyndon B. KURTH,* Ian D. JENKINS‡ and Douglas J. HORGAN*§
*C.S.I.R.O. Division of Food Processing, Meat Research Laboratory, P.O. Box 12, Cannon Hill, Brisbane, Queensland 4170, Australia, †Deakin Research Limited, C/- C.S.I.R.O. Division of Food Processing, Food Research Laboratory, P.O. Box 52, North Ryde, New South Wales 2113, Australia, and ‡Division of Science and Technology, Griffith University, Nathan, Queensland 4111, Australia

Collagenous peptides containing the Ehrlich chromogen (EC) were selectively isolated from a tryptic digest of bovine tendon by coupling to a diazotized polyacrylamide support. The isolated *p*-phenol-azo-EC peptides were purified and characterized by amino acid and sequence analyses. EC occurred in stoichiometric amounts in trimeric cross-linked chains originating from the known cross-link regions of type-I collagen. The major locus of the EC was α2(I)Hyl-933 × α1(I)Lys(Hyl)-9^N × α2(I)Lys(Hyl)-5^N but it was also shown to occur at the loci α1(I)Hyl-87 × α1(I)Lys(Hyl)-16^C × α1(I)Lys(Hyl)-16^C and α1(I)Hyl-930 × α1(I)Lys(Hyl)-9^N × α2(I)Lys(Hyl)-5^N. After sequence analyses of the *C*-terminal helical cross-link region α2(I)928–963, corrections are presented for residues 927, 930, 932 and 933 of the bovine α2(I) chain. The collagen-associated EC is postulated to be a trisubstituted pyrrole formed by the reaction of the aldehyde form of a telopeptidyl lysine residue with a bifunctional keto amine cross-link. It is also proposed that when the telopeptidyl lysine residue is hydroxylated the above reaction will result in pyridinoline formation.

INTRODUCTION

Type-I collagen cross-linking (see Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988, for recent reviews) is initiated by the enzymic oxidative deamination, by lysyl oxidase, of lysine or hydroxylysine residues in the telopeptidyl regions of the molecule to give aldehyde forms commonly known as allysine and hydroxyallysine respectively. Owing to the quarter-stagger arrangement of type-I collagen molecules, allysine or hydroxyallysine residues in the *C*-terminal and *N*-terminal telopeptides can form cross-links with helical hydroxylysine residues 87 and 930 respectively (Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988). Two possible cross-link pathways can be followed, depending on whether allysine or hydroxyallysine reacts with helical hydroxylysine. On the allysine pathway, aldimine cross-links occur, whereas on the hydroxyallysine pathway rearrangement of hydroxylated aldimines to keto amines occurs (Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988). These keto amine cross-links are highly reactive, and are believed to condense to form multivalent cross-links (Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988). Besides pyridinoline (PYR), which is an established tervalent cross-link on the hydroxyallysine pathway (Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988), several groups have suggested that other keto amine condensation products may exist as collagen cross-links (Eyre *et al.*, 1988; Henkel *et al.*, 1987; Linde & Robins, 1988; Robins & Duncan, 1987). In particular, a collagen-associated Ehrlich chromogen (EC), postulated to be an *N*-substituted pyrrole, has been proposed as another keto amine-derived cross-link (Scott *et al.*, 1981). Its presence was demonstrated in a well-characterized trimeric cross-linked peptide from human type-III collagen (Scott *et al.*, 1983*a*). Recently, by using an affinity method based on diazonium coupling of the EC to a polyacrylamide support, uncharacterized phenol-azo-EC peptides were isolated from a papain digest of bovine skin collagen. From this work the molar absorptivity of the collagen *p*-phenol-azo-EC group was determined, thereby allowing quantification of the EC (Kemp & Scott, 1988). In a previous study, we found that in bovine tendons the site variation in EC content was analogous to that of pyridinoline, thus supporting the contention that EC was a cross-link (Horgan *et al.*, 1990). In the present study we have characterized, by amino acid and sequence analyses, *p*-phenol-azo-EC tryptic peptides isolated from digests of bovine tendon type-I collagen.

EXPERIMENTAL

Preparation of insoluble collagen

Psoas major muscles were obtained from six cattle (approx. 18 months old), slaughtered at the Meat Research Laboratory. After removal of fat and overlying layers of connective tissue, the tendons of insertion were removed from the muscle surfaces, by careful dissection. Any adhering muscle tissue was removed by scraping the tendon with a scalpel. The tendons were washed in saline (0.9% NaCl), and stored frozen in saline until required. Insoluble collagen was prepared by extracting the tendons in 0.05 M-Tris/HCl buffer, pH 7.5, containing 4 M-guanidinium chloride (15 ml/g wet wt. of tendon), with stirring for 24 h at 4 °C. The insoluble residue was collected by filtration, and washed exhaustively with distilled water. The material was then defatted and dehydrated by a series of diethyl ether/ethanol extractions (0:100, 1:3, 1:1, 3:1 and 100:0, v/v), by stirring in 1 litre of each solution for 1 h at room temperature.

Digestion with trypsin

An 11 g portion of the insoluble collagen was suspended in 300 ml of 0.2 M-NH_4HCO_3 containing $CaCl_2$ (1 mM), pH 7.8, and denatured by heating for 15 min at 65 °C. The suspension was cooled to 37 °C, and after the addition of 110 mg of trypsin (Tos-Phe-CH_2Cl-treated; Sigma Chemical Co.), incubated at

Abbreviation used: EC, Ehrlich chromogen.
§ To whom correspondence should be addressed.

37 °C for 4 h. The digest was adjusted to pH 4, centrifuged at 35000 *g* for 30 min and the supernatant collected.

Pyroglutamate aminopeptidase digestion

Pyroglutamate aminopeptidase digestion of azo-EC peptides was carried out by using the procedure of Podell & Abraham (1978) with an enzyme/substrate ratio of 1:100 (w/w). Digested peptides were loaded on a Delta Pak reverse-phase C_{18} column (3.9 mm × 30 cm) that had been equilibrated at 55 °C with 0.1 % (v/v) trifluoroacetic acid (solvent A). Salts were removed by washing with solvent A for 10 min. Peptides were then eluted by using a linear gradient of 0–35 % (v/v) acetonitrile in solvent A over 30 min at a flow rate of 1 ml/min.

Preparation of *p*-phenol-azo-EC peptides

The tryptic digest was adjusted to pH 1.5 with HCl, and 10 g of freshly diazotized polyacrylamide support was added (Kemp & Scott, 1988). Coupling was carried out at 4 °C, in the dark, for 3 h. Unchanged diazonium groups were blocked by reaction with α-naphthol, and the *p*-phenol-azo-EC peptides were isolated by using the procedure of Kemp & Scott (1988). The aqueous extract containing the azo-EC peptides was freeze-dried and stored at −70 °C.

Molecular sieve chromatography

The azo-EC peptides were dissolved in 3 ml of 2 M-guanidinium chloride/50 mM-Tris/HCl buffer, pH 7.5, and loaded on a Sephadex G-50 (superfine grade) column (2.6 cm × 95 cm) that had been equilibrated with the same buffer. The column was eluted at room temperature at a flow rate of 25 ml/h, and 5 ml fractions were collected. The fractions were monitored manually at 250 nm for peptide absorbance and at 400 nm to detect the *p*-phenol-azo-EC group (Kemp & Scott, 1988).

H.p.l.c.

Reverse-phase h.p.l.c. was carried out with a Waters 600E gradient system. Fractions from the Sephadex G-50 column were purified on a Waters Delta Pak C_{18} (30 nm) (15 μm particle size) column (7.8 mm × 30 cm). The column was equilibrated at 55 °C with 0.1 % trifluoroacetic acid (solvent A) at a flow rate of 4 ml/min, and the peptides were eluted using linear gradients of solvent B (70 % acetonitrile) in solvent A. The gradients employed were 0–25 % solvent B in 5 min, then 25–40 % solvent B in 60 min. The eluate was monitored at 230 nm and 400 nm with a Waters 490E multiwavelength detector, and fractions were collected manually. Fractions of interest from the above column were re-run on a smaller Delta Pak C_{18} (30 nm) (15 μm particle size) column (3.9 mm × 30 cm) at 1 ml/min; otherwise the procedure was identical with that described above.

Amino acid analysis

Samples were hydrolysed in the gas phase for 1 h at 150 °C with 6 M-HCl containing 1 % (v/v) phenol. Hydrolysates were dried, then treated with phenyl isothiocyanate to produce phenylthiocarbamoyl-amino acids. These were analysed by the Waters Associates Pico-Tag method.

Amino acid sequence analysis

The amino acid sequences were determined by Edman degradation with an Applied Biosystems 470A automatic gas-phase sequencer coupled to a 120A phenylthiohydantoin analyser with a standard phenylthiohydantoin program.

Nomenclature of cross-linked peptides

The nomenclature used here for cross-linked peptides is as follows: the collagen chain and type are given followed by the residue numbers as assigned to the collagen sequences compiled by Galloway (1982). The residues that participate in the cross-link are shown in brackets.

RESULTS

Molecular sieve chromatography of azo-EC peptides

The azo-EC peptides were initially separated according to molecular mass by Sephadex G-50 chromatography (Fig. 1). Four fractions, I–IV, were collected on the basis of the absorbance of the *p*-phenol-azo-EC group at 400 nm (Kemp & Scott, 1988). As shown in Fig. 1, each 400 nm peak coincided with peptidyl absorbance as determined by u.v. absorption at 250 nm. All fractions displayed the same absorption spectra as that described for the *p*-phenol-azo-EC group (Kemp & Scott, 1988) (results not shown). The fractions were desalted on a Bio-Gel P-2 column (5 cm × 25 cm) with 0.1 M-acetic acid as eluent.

Purification of molecular-mass fractions by reverse-phase h.p.l.c.

Fractions I–IV from the Sephadex G-50 column (Fig. 1) were further purified by reverse-phase h.p.l.c. Fig. 2 shows the separation of fraction I with the fraction collected being designated F1. Fig. 3(*a*) shows that fraction II gave two major peaks

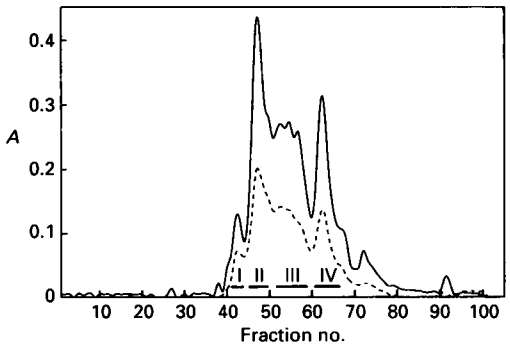


Fig. 1. Molecular-sieve chromatography of *p*-phenol-azo-EC peptides on Sephadex G-50 superfine

The continuous and broken lines are the absorbances at 400 nm and 250 nm respectively. The bars indicate the fractions collected and they are designated by Roman numerals. See the Experimental section for further details.

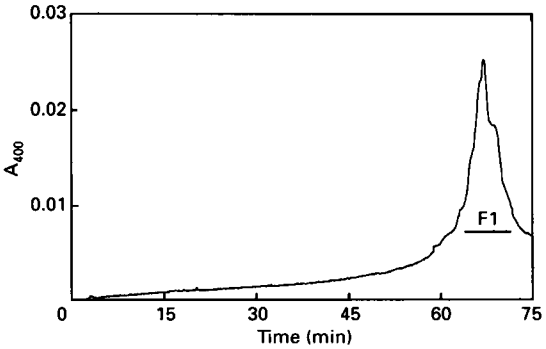


Fig. 2. Reverse-phase h.p.l.c. of fraction I from Fig. 1

Elution was carried out at 55 °C with 0.1 % trifluoroacetic acid for 10 min, followed by linear gradients (0–25 % solvent B/5 min then 25–50 % solvent B/60 min) of solvent B (70 % acetonitrile in 0.1 % trifluoroacetic acid) at a flow rate of 4 ml/min. The absorbance of the eluate was monitored at 400 nm. The bar indicates the collected fraction, designated F1.

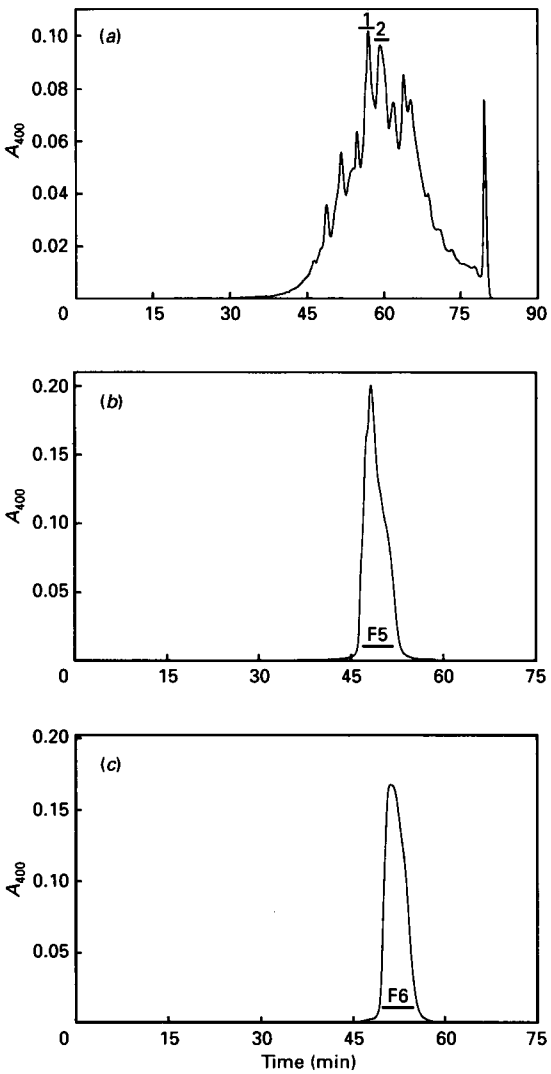


Fig. 3. Reverse-phase h.p.l.c. of fraction II from Fig. 1

(*a*) Fraction II run using the conditions described in Fig. 2. The bars indicate the collected fractions, designated 1 and 2. (*b*) Rechromatography of fraction 1 from (*a*) on a C_{18} Delta Pak column (3.9 mm × 30 cm). Elution conditions were the same as described for Fig. 2, except that gradients were started at zero time and a flow rate of 1 ml/min was used. The bar indicates the collected fraction, designated F5. (*c*) Rechromatography of fraction 2 from (*a*). Conditions as in (*b*). The bar indicates the collected fraction, designated F6.

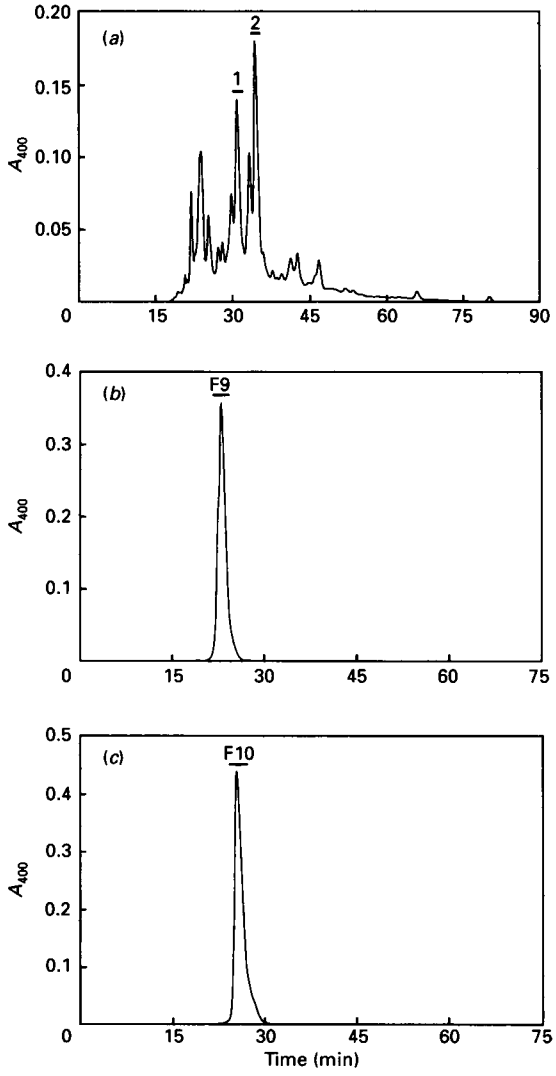


Fig. 4. Reverse-phase h.p.l.c. of fraction IV from Fig. 1

(*a*) Fraction IV run using conditions as described in Fig. 2. The bars indicate the collected fractions, designated 1 and 2. (*b*) Rechromatography of fraction 1; conditions are the same as for Fig. 3(*b*). The bar indicates the collected fraction, designated F9. (*c*) Rechromatography of fraction 2; conditions are the same as for Fig. 3(*b*). The bar indicates the collected fraction, designated F10.

(labelled 1 and 2) which were collected and rechromatographed by h.p.l.c. on the smaller reverse-phase column (3.9 mm × 30 cm). The fractions collected, as indicated on Figs. 3(*b*) and 3(*c*), were designated F5 and F6 respectively. Fig. 4(*a*) shows the separation of fraction IV. Two major peaks (labelled 1 and 2) were collected and rechromatographed using the smaller reversed-phase h.p.l.c. column. The fractions collected (Figs. 4*b* and 4*c*) were designated F9 and F10 respectively. Fraction III was also further fractionated by reverse-phase h.p.l.c. (results not shown).

Amino acid and sequence analysis of purified peptides

Peptide F1. The amino acid composition of peptide F1 is given in Table 1 and agrees closely with the sum of α1(I)993–22^{c} + α1(I)993–22^{c} + α1(I)76–90 (Galloway, 1982). The peptide was therefore estimated to have a total of 102 residues, and contained a stoichiometric amount of EC. The low contents of lysine and hydroxylysine in peptide F1 are consistent with their involvement in the formation of EC as the cross-link. Table 2 gives the sequence data for peptide F1 showing that a total of 44 sequence cycles were completed. The expected phenylthiohydantoin derivatives for the sequences α1(I)993–1007 and α1(I)76–90 were observed together in cycles 1–15 except for the absence of the cross-link residue Hyl-87 in cycle 12. This indicates that this residue participates in the formation of the cross-link. The ratio of two chains of α1(I)993–22^{c} to one of α1(I)76–90 was confirmed by the approx. 2-fold higher yields of threonine to glycine in cycle 1 and glycine to leucine in cycle 2. Table 2 also shows that in sequence cycles 16–44 the phenylthiohydantoin derivatives for the entire sequence α1(I)1008–22^{c} were obtained except for the cross-link residue Lys(Hyl)-16^{c} at cycle 38. Therefore two residues of Lys(Hyl)-16^{c} also participate in the formation of the cross-link in this peptide. Thus the amino acid and sequence data for peptide F1 unequivocally identify it as

Table 1. Amino acid composition (residues/peptide) of peptide F1

Amino acid	Composition	
	F1	Published*
Asp	5.4	6
Glu	7.4	8
Ser	6.2	4
Gly	26.8	27
His	2.6	3
Arg	4.1	3
Thr	3.3	3
Ala	5.7	7
Pro	17.1	20
Tyr	2.7	2
Val	1.3	0
Met	0.2	1
Ile	1.1	0
Leu	6.9	6
Phe	2.0	2
Lys	0.1	0
Hyp	8.5	10
Hyl	0.8	3
Total	102.2	105
EC†	1.2	

* Composition of α1(I)993–22^{C} + α1(I)993–22^{C} + α1(I)76–90 (Galloway, 1982).

† Determined by absorbance at 400 nm of the *p*-phenol-azo-EC group, by using a molar absorption coefficient of 18700 litre·mol^{-1}·cm^{-1} (Kemp & Scott, 1988).

α1(I)993–22^{C} × α1(I)993–22^{C} × α1(I)76–90, cross-linked at residues α1(I)-Hyl-87 and Lys(Hyl)-16^{C}.

Peptides F5 and F6. Sequencing of peptides F5 and F6 yielded the same amino acid sequence (Fig. 5*a*) for both peptides. No residue was found at cycle 6, which indicated a cross-link site (denoted by X in Fig. 5*a*). The closest matching type I collagen sequence (Galloway, 1982) to peptides F5 and F6 was found to be the α2(I)928–963 sequence, which is shown in Fig. 5(*b*) for comparison. The sequence of peptides F5 and F6 differ from the published bovine α2(I)928–963 sequence at residues 930, 932 and 933 where hydroxyproline, leucine and the cross-link occur in place of hydroxylysine, aspartate and arginine. Fig. 5(*c*) shows the sequence of α2(I)928–963 for chick, as determined from nucleotide sequencing (Galloway, 1982). This sequence is also very similar to that of bovine α2(I)928–963, and peptides F5 and F6. The chick sequence displays lysine at position 933, which is a possible cross-link residue (Miller, 1984).

Table 3 shows the amino acid composition of peptides F5 and F6. Both peptides had a total of 76 residues with virtually identical amino acid compositions and each contained stoichiometric amounts of EC. Table 3 also gives the composition of the remaining amino acids, after subtraction of the residues determined by sequencing peptides F5 and F6. The presence of tyrosine in the composition of peptides F5 and F6 (Table 3) suggests that the helical sequences of peptides F5 and F6 are cross-linked to the *N*-terminal telopeptides. The value for the sum of the published (Galloway, 1982) amino acid compositions of the *N*-terminal telopeptides, α1(I)1^{N}–9 + α2(I)1^{N}–9, are also given in Table 3. As can be seen, the sum of the literature value of these *N*-terminal telopeptides is in good agreement with the remainder when the 36 residues of the helical sequence are subtracted from the amino acid composition of peptides F5 and F6.

N-Terminal telopeptides cross-linked to the helical *C*-terminal cross-link region of the α2(I)chain would not be seen during sequencing since both chains are *N*-terminally blocked by pyroglutamate residues (Galloway, 1982). Therefore, in order to reveal the sequence of *N*-terminal telopeptidyl chains, peptides F5 and F6 were digested with pyroglutamate aminopeptidase in order to remove pyroglutamic acid. Table 4 shows the sequence

Table 2. Sequence data for peptide F1

Cycle	Amino acids		Cycle	Amino acids	
	Expected*	Observed†		Expected*	Observed†
1	TG	T(900), G(396)	23	S	S
2	GL	G(757), L(314)	24	G	G
3	DP·	DP·	25	G	G
4	AG	AG	26	Y	Y
5	GT	GT	27	D	D
6	PA	PA	28	L	L
7	AG	AG	29	S	S
8	GL	GL	30	F	F
9	PP·	PP·	31	L	L
10	P·G	P·G	32	P	P
11	GM	GM	33	Q	Q
12	P–	PM	34	P	P
13	P·G	P·G	35	P	P
14	GH	GH	36	Q	Q
15	PR	PRH	37	Q	Q
16	P·	P·R	38	–	–
17	G	G	39	A	A
18	P	PG	40	H	H
19	P·	P·	41	D	H
20	G	G	42	G	H
21	P	P	43	G	None
22	P	P	44	R	DG

* Expected residues for α1(I)993–22^{C} × α1(I)993–22^{C} × α1(I)76–90 (Galloway, 1982).

† Results are for 5 nmol of peptide. Values in parentheses are observed yields in pmol. P· denotes hydroxyproline.

(a)
GLP̣GLXGHDGLQGLP̣GLAGHHGDQGAP̣GAVGPAGPR

(b)
GLḲGDRGHNGLQGLP̣GLAGHHGDQGAP̣GAVGPAGPR

(c)
GLPGLKGHNGLQGLPGLAGQHGDQGPPGNNGPAGPR

Fig. 5. Comparison of amino acid sequences of peptides F5 and F6 with α2(I)928–963 for type I collagen (Galloway, 1982)

(*a*) Sequence of peptides F5 and F6. Amino acids are labelled using the single-letter code, with P̣ and X denoting hydroxyproline and the cross-link site respectively. (*b*) Sequence of bovine α2(I)928–963 for type-I collagen (Galloway, 1982). Ḳ denotes hydroxylysine. (*c*) Sequence of chick α2(I)928–963 as determined from nucleotide sequencing (Galloway, 1982).

Table 3. Amino acid composition (residues/peptide) of peptides F5 and F6

Amino acid	Composition			
	F5	F6	Minus sequence*	Published†
Asp	4.1	4.1	2.1	2
Glu	5.7	5.6	3.7	3
Ser	5.0	4.8	4.9	4
Gly	21.5	21.6	9.6	11
His	2.6	2.6	0	0
Arg	3.3	3.2	2.3	2
Thr	1.8	1.7	1.8	1
Ala	4.6	4.7	0.7	1
Pro	8.5	8.5	6.5	7
Tyr	2.2	2.2	2.2	2
Val	2.2	2.1	2.2	1
Met	1.7	1.8	1.8	3
Ile	1.5	1.4	1.5	1
Leu	6.0	6.0	1.0	2
Phe	0.8	1.1	1.0	1
Lys	0	0	0	2
Hyp	4.2	4.1	1.2	0
Hyl	0.5	0.6	0.6	1
Total	76.2	76.1	43.1	44
EC‡	0.8	0.7		

* Subtracting the sequence found for peptides F5 and F6 (Fig. 5*a*) from their amino acid compositions.

† Amino acid composition for $\alpha 1(I)1^N{-}9 + \alpha 2(I)1^N{-}9$ (Galloway, 1982).

‡ Determined by absorbance at 400 nm of the *p*-phenol-azo-EC group, by using a molar absorption coefficient of 18700 litre·mol^{-1}·cm^{-1} (Kemp & Scott, 1988).

data for the first six cycles of the unblocked peptide F6. Comparison of the unblocked sequence data with the control data in Table 4 reveals the sequence LSYGYD, indicating the *N*-terminal telopeptide $\alpha 1(I)1^N{-}9$ (Galloway, 1982). Similarly, the presence of the *N*-terminal telopeptide $\alpha 2(I)1^N{-}9$ is indicated by phenylalanine and alanine in cycles 1 and 3 respectively, because these residues are unique for this sequence (Galloway, 1982). Similar results to that in Table 4 were also found for peptide F5 (results not shown). Therefore from the amino acid compositions (Table 3) and the results obtained from unblocked peptide F6 (Table 4), the 76-residue azo-EC peptides F5 and F6 are proposed to be the trimeric cross-linked peptide $\alpha 1(I)1^N{-}9 \times \alpha 2(I)1^N{-}9 \times \alpha 2(I)928{-}963$, cross-linked at residues α2(I)Hyl-933, α1(I)Lys(Hyl)-9^N and α2(I)Lys(Hyl)-5^N.

Table 4. Sequence data for peptide F6 after pyroglutamate aminopeptidase digestion

Results are for 5 nmol of peptide. P* denotes hydroxyproline. Values in parentheses are pmol; n.q. indicates not quantified. Underlined residues indicate *N*-terminal telopeptide residues found, and the dash denotes the cross-link residue.

Cycle	Amino acids	
	Control	Digested
1	G(887), L(19), F(19)	G(887), L(177), F(73)
2	L(939), Y(65), D(22), I(43)	L(960), Y(67), D(21), I(25), S(40)
3	P*(n.q.), D(77), P(22)	P*(n.q.), D(58), P(29), Y(159), A(28)
4	G(773), Q(15), E(31), A(31)	G(915), Q(20), A(20)
5	L(662), P(7)	L(656), P(33), T(6), Y(125)
6	–, S(28), A(22), P(48)	–, S(6), P(55), D(77)

Table 5. Amino acid composition (residues/peptide) of peptides F9 and F10

Amino acid	Composition	
	F9	F10
Asp	1.5	3.3
Glu	3.9	2.8
Ser	4.4	5.1
Gly	8.6	9.1
His	1.0	1.0
Arg	2.5	3.7
Thr	1.3	2.2
Ala	0.6	1.5
Pro	4.7	4.7
Tyr	1.7	2.1
Val	1.3	1.4
Met	0.9	0.8
Ile	2.4	2.6
Leu	1.6	2.1
Phe	0.3	0.3
Lys	0	0.6
Hyp	0.9	0.3
Hyl	0.3	0.3
Total	37.9	43.9
EC*	0.9	0.9

* Determined by absorbance at 400 nm of the *p*-phenol-azo-EC group, by using a molar absorption coefficient of 18700 litre·mol^{-1}·cm^{-1} (Kemp & Scott, 1988).

Peptides F9 and F10. The amino acid compositions of peptides F9 and F10 are presented in Table 5. Peptides F9 and F10 had a total of 38 and 44 residues respectively and each contained a stoichiometric amount of EC. On sequencing, both peptides gave the sequence GI–GHR, which corresponds to the helical *C*-terminal cross-link site, α1(I)928–933 (Galloway, 1982). The blank at cycle 3 defined a cross-link site, indicating that Hyl-930 is involved in the cross-link. Attempts to reveal *N*-terminal telopeptides by unblocking with pyroglutamate aminopeptidase proved unsuccessful with both peptides. This may have been due to a combination of steric hindrance of the enzyme by the bulky azo-phenol-EC group and the smaller size of these peptides compared with peptides F5 and F6. However, the presence of two residues of the unique telopeptide amino acid tyrosine in each peptide (Table 5) confirmed the presence of the *N*-terminal telopeptide sequence $\alpha 1(I)1^N{-}9$, which can cross-link to

α1(I)928–933 in a quarter-stagger arrangement (Eyre *et al.*, 1984*a*; Yamauchi & Mechanic, 1988). Similarly, the presence of the α2(I)1^{N}–9 sequence can be inferred from phenylalanine and alanine in the amino acid compositions of peptides F9 and F10 (Table 5), since both these residues are unique to the α2(I)1^{N}–9 chain (Galloway, 1982). Therefore, on the basis of their amino acid composition and the presence of the helical *C*-terminal cross-link region α1(I)928–933, peptides F9 and F10 are proposed to be the trimeric cross-linked peptide, α1(I)1^{N}–9 × α2(I)1^{N}–9 × α1(I)928–933, cross-linked at residues α1(I)Hyl-930, α1(I)Lys(Hyl)-9^{N} and α(I)Lys(Hyl)-5^{N}.

Peptides derived from fraction III. On separation by reverse-phase h.p.l.c., fraction III gave multiple *p*-phenol-azo-EC peptides (results not shown). Amino acid and sequence analysis of the four major peptides indicated that these were mixtures of smaller fragments (approx. 50 residues), obtained from peptides F1, F5 and F6. Another fraction was found to be a mixture of peptides similar to F9 and F10 but with fragments of peptides F5 and F6 also present (results not shown). These fragments probably arose from chymotryptic cleavage of the larger peptides F1, F5 and F6, since it is known that Tos-Phe-CH_2Cl-treated trypsin still contains traces of chymotrypsin activity (Jany *et al.*, 1976). Since no other cross-linked peptides apart from fragments of those already described were found, no results are presented for any of these peptides.

DISCUSSION

EC-containing tryptic peptides were isolated from a digest of insoluble bovine type I tendon collagen by coupling to a diazotized support. Coupling was carried out in acid solution to decrease the reactivity of phenolic and imidazole groups (Scott *et al.*, 1983*b*). Under these conditions, the coupled peptides displayed the absorption spectra of the collagen-associated *p*-phenol-azo-EC group (Kemp & Scott, 1988), indicating that a single reactive compound had coupled to the support. Characterization of these azo-EC peptides has shown conclusively that stoichiometric amounts of EC were present in trimeric cross-linked peptides, originating from the established cross-link regions of type I collagen.

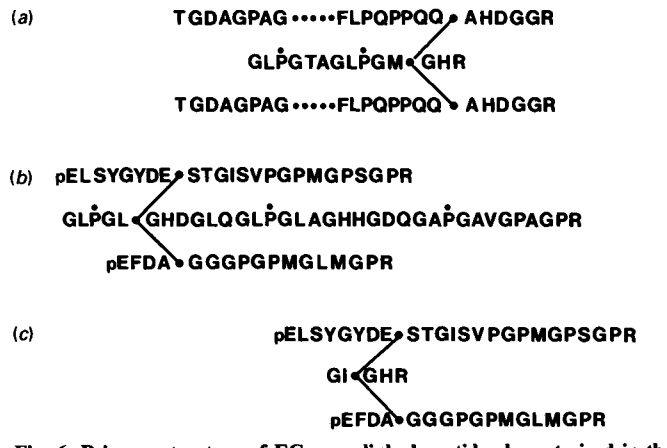


Fig. 6. Primary structure of EC cross-linked peptide characterized in the present study

Amino acids are labelled using the single-letter code, with Ṗ denoting hydroxyproline and pE pyroglutamic acid. The EC cross-link sites are represented as black circles joined by lines. (*a*) Peptide F1 characterized as α1(I)993–22^{C}[Hyl-16^{C}] × α1(I)993–22^{C}[Hyl-16^{C}] × α1(I)76–90[Hyl-87]. The small dots indicate the omission of residues 1001–9^{C} in order to conserve space. (*b*) Peptides F5 and F6 characterized as α1(I)1^{N}–9[Hyl-9^{N}] × α2(I)1^{N}–9[Hyl-5^{N}] × α2(I)928–963[Hyl-933]. (*c*) Peptides F9 and F10 characterized as α1(I)1^{N}–9[Hyl-9^{N}] × α2(I)1^{N}–9[Hyl-5^{N}] × α1(I)928–933[Hyl-930].

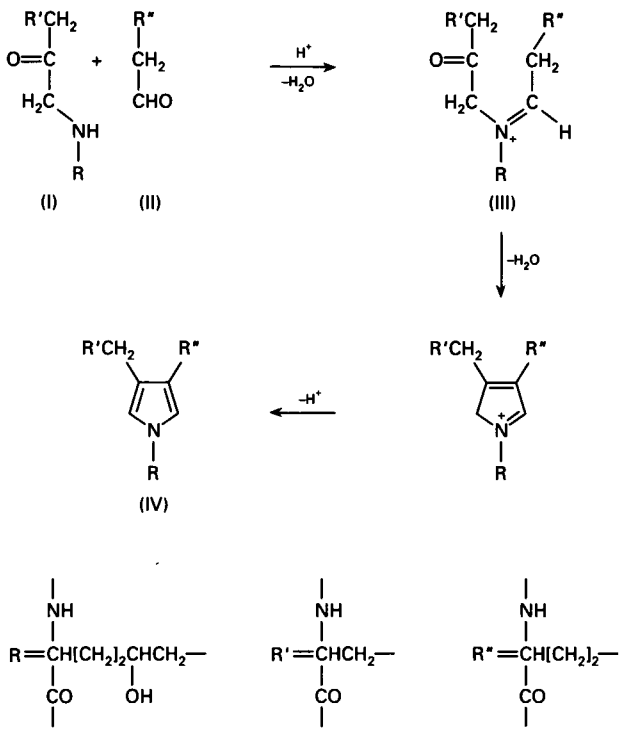
The proposed primary structure of three trimeric cross-linked peptides characterized in this study are shown in Fig. 6. Fig. 6(*a*) shows the structure of F1, identified as the tryptic trimeric peptide α1(I)993–22^{C} × α1(I)993–22^{C} × α1(I)76–90. Since sequence analysis of this peptide indicated a cross-link at positions 16^{C} and 87 and spectrophotometric analysis indicated the presence of 1 mol of EC/mol of peptide, we propose that EC is the cross-link structure in this peptide. We found no evidence for a peptide containing the published (Galloway, 1982) helical *C*-terminal cross-link region of the α2(I) chain in any of the peptides examined. Such a tryptic peptide would display the sequence GDVGL–GDR (the blank indicating the cross-link Hyl-930 residue) and appear as a 50-residue trimer. The only trimeric cross-linked peptides resembling the published α2(I) helical *C*-terminal cross-link region were the 76-residue peptides F5 and F6 identified as α1(I)1^{N}–9 × α2(I)1^{N}–9 × α2(I)928–963 and shown in Fig. 6(*b*). We therefore propose that the published sequence for the bovine α2(I) chain, near the helical *C*-terminal cross-link region, is incorrect, and propose the following substitutions: arginine for valine at residue 927, hydroxyproline for hydroxylysine at residue 930, leucine for aspartic acid at residue 932, and hydroxylysine for arginine at residue 933.

Fig. 6(*c*) shows the proposed primary structure of peptides F9 and F10 as the tryptic cross-linked peptide involving the α1(I) helical *C*-terminal cross-link region, cross-linked to the *N*-terminal telopeptides. Although we were unable to reveal the *N*-terminal telopeptidyl sequences by enzymic unblocking, the results obtained for the peptide's amino acid composition and the sequence identification of the helical *C*-terminal cross-link region, α1(I)928–933, are sufficient to assign the primary structure given in Fig. 6(*c*).

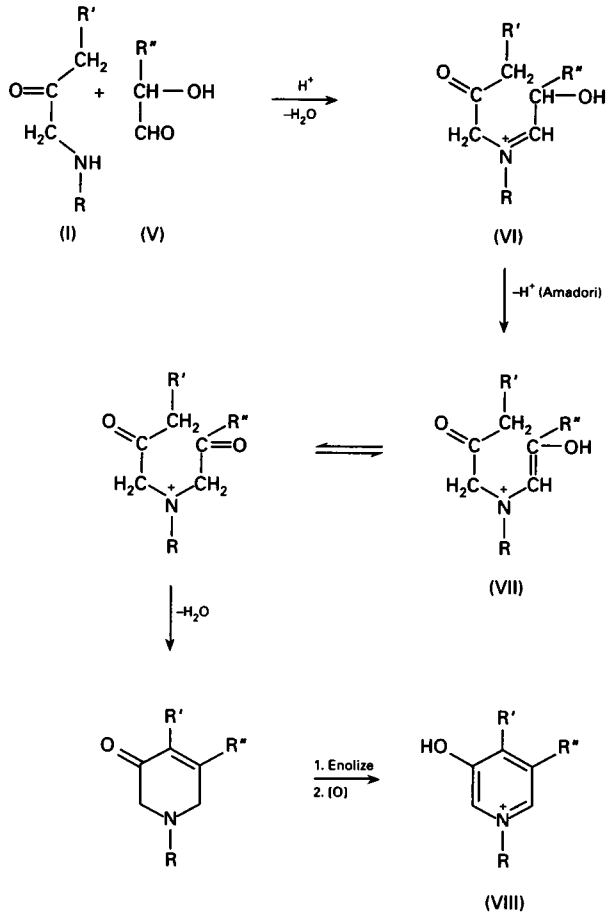
Interestingly, some of the peptides with identical sequences displayed different retention times by h.p.l.c., e.g. peptides F5 and F6 (Fig. 3) and peptides F9 and F10 (Fig. 4). Since the collagen-associated EC has been shown to have either galactose or glycosylgalactose attached to it (Kemp & Scott, 1988), the different chromatographic properties of these peptides could be due to different glycosylation levels of the EC.

Fig. 1 shows that fraction II was the most abundant azo-EC peptide fraction with 34% of the total EC, as quantified by 400 nm absorbance. Since the cross-link site of this molecular mass fraction has been characterized as α2(I)[Hyl-933] × α2(I)[Lys(Hyl)-5^{N}] × α1(I)[Lys(Hyl)-9^{N}], this would appear to be the major locus of the EC in bovine tendon. In bovine dentine collagen, Scott (1980) found that virtually all of the *C*-terminal helical cross-link region of the α2(I) chain was cross-linked to the *N*-terminal telopeptides and that the major cross-link was non-reducible and of unknown structure. This cross-link was probably the EC since the results of this study indicate that the α2(I) chain is the preferred location of the EC.

In each trimeric cross-linked peptide (Figs. 6*a*, *b* and *c*), low levels of hydroxylysine and lysine were revealed by amino acid analyses; furthermore, sequence analyses revealed blanks at the expected residue positions of hydroxylysine and lysine. This together with the stoichiometric amount of EC found in each peptide indicates that it results from the probable condensation of two telopeptidyl aldehydes (derived from either lysine or hydroxylysine), and one helical hydroxylysine residue. Therefore the collagen-associated EC is likely to be a different compound from an EC reported to occur in elastin (Kemp & Scott, 1988), since this latter compound would be formed by the reaction of lysine with lysine-derived aldehydes. Furthermore, the spectra of the reaction products with Ehrlich's reagent have been found to be different in elastin and collagen (Kemp & Scott, 1988).



Scheme 1. Formation of a 1,3,4-trisubstituted pyrrole cross-link by condensation of a keto amine with an allysine



Scheme 2. Formation of the pyridinoline cross-link by condensation of a keto amine with a hydroxyallysine

R groups as in Scheme 1.

The evidence presented for the involvement of lysine (hydroxylysine) residues from known cross-link sites of type-I collagen in the formation of the proposed pyrrolic structure also disproves the suggestion of Hayase *et al.* (1989) that EC arises from non-enzymic glycosylation of collagen.

A trimeric peptide involving the same cross-link sites as shown in Fig. 6(*a*) has previously been shown to contain 0.25 mol of pyridinoline/mol of peptide (Henkel *et al.*, 1987). We failed to detect pyridinoline in purified azo-EC peptides. This was despite using nanomolar loadings in a fluorimetric h.p.l.c. assay with a sensitivity below 2 pmol of PYR (Eyre *et al.*, 1984*b*). This result was not unexpected since pyridinoline does not react with the Ehrlich reagent (Scott *et al.*, 1983*a*,*b*) and would therefore not bind to the diazotized support.

Although there is as yet no firm evidence for the structure of the collagen-associated EC, its chemical reactivity is consistent with that of an *N*-substituted pyrrole postulated to be a keto amine condensation product (Scott *et al.*, 1981). Consistent with this postulate we have recently shown that a high correlation exists between EC, pyridinoline and keto amine concentrations (Horgan *et al.*, 1990). On the basis of these considerations and evidence for the occurrence of EC and pyridinoline at the same cross-link loci (Henkel *et al.*, 1987 and Fig. 6*a*), we propose that EC and pyridinoline formation may occur by the condensation of a keto amine with a lysine or hydroxylysine aldehyde respectively.

In Scheme 1 we propose that the collagen-associated EC is a 1,3,4-trisubstituted pyrrole (IV) formed by the reaction of the keto amine, hydroxylysino-5-oxonorleucine (I) with allysine (II). This reaction order seems most likely since it is established that bifunctional cross-linking occurs initially between adjacent quarter-staggered molecules (Siegel, 1976; Fukae & Mechanic, 1980). Subsequent ring closure of the Schiff base adduct (III) would lead to trisubstituted pyrrole (IV).

In Scheme 2 it is proposed that when hydroxylysino-5-oxonorleucine (I) reacts with hydroxyallysine (V) the hydroxylated Schiff base adduct (VI) so formed would undergo an Amadori rearrangement to compound (VII). Ring closure of compound (VII) would ultimately lead to pyridinoline formation (VIII). Schemes 1 and 2 then have as a common basis the formation of trifunctional cross-links by the reaction of a keto amine with a free telopeptidyl aldehyde. When the aldehyde is hydroxylated, pyridinoline forms and 1,3,4-trisubstituted pyrrole is formed in the unhydroxylated case. Schemes 1 and 2 indicate that in collagenous tissues where keto amine cross-linking initially occurs, the subsequent ratio of pyridinoline to EC (trisubstituted pyrrole) will be determined by the level of telopeptidyl lysine hydroxylation. This proposal is consistent with current evidence. For example, in the highly hydroxylated type II collagen of cartilage, pyridinoline has been shown to be present in stoichiometric amounts (Robins & Duncan, 1983). In a trimeric peptide isolated from calf aorta the *C*-terminal telopeptidyl lysine residues were only partially hydroxylated, resulting in a pyridinoline content of 0.25 mol/mol of peptide (Henkel *et al.*, 1987). Evidence for the presence of a second cross-link, which was not stable to acid hydrolysis, was also found in this peptide. Another trimeric peptide, from human type-III collagen, was also shown to contain 0.25 mol of pyridinoline/mol of peptide and gave a positive colour reaction with the Ehrlich reagent (Scott *et al.*, 1983*a*).

In conclusion, the results of this study provide strong evidence that the EC is a major collagen cross-link involving primarily the *C*-terminal helical cross-link region of the α2(I) chain in bovine

type I collagen. We propose that the EC is a trisubstituted pyrrole, which is formed by the condensation of a bifunctional keto amine cross-link with an adjacent allysine residue.

This work was supported by the Australian Meat and Livestock Research Development Corporation.

REFERENCES

Eyre, D. R., Paz, M. A. & Gallop, P. M. (1984*a*) Annu. Rev. Biochem. **53**, 717–748
Eyre, D. R., Koob, T. J. & Van Ness, K. P. (1984*b*) Anal. Biochem. **137**, 380–388
Eyre, D. R., Dickson, I. R. & Van Ness, K. (1988) Biochem. J. **252**, 495–500
Fukae, M. & Mechanic, G. L. (1980) J. Biol. Chem. **255**, 6511–6518
Galloway, D. (1982) in Collagen in Health and Disease (Weiss, J. B. & Jayson, I. V., eds.), pp. 528–556, Churchill Livingstone, Edinburgh
Hayase, F., Nagaraj, R. H., Miyata, S., Njoroges, F. G. & Monnier, V. M. (1989) J. Biol. Chem. **264**, 3758–3764
Henkel, W., Glanville, R. W. & Greifendorf, D. (1987) Eur. J. Biochem. **165**, 427–436
Horgan, D. J., King, N. L., Kurth, L. B. & Kuypers, R. (1990) Arch. Biochem. Biophys. **281**, 21–26
Jany, K. D., Keil, W., Meyer, H. & Kiltz, H. H. (1976) Biochim. Biophys. Acta **453**, 62–66
Kemp, P. D. & Scott, J. E. (1988) Biochem. J. **252**, 387–393
Linde, A. & Robins, S. P. (1988) Collagen Relat. Res. **8**, 443–450
Miller, E. J. (1984) in Extracellular Matrix Biochemistry (Piez, K. A. & Reddi, A. H., eds.), pp. 41–81, Elsevier, New York
Podell, D. N. & Abraham, G. N. (1978) Biochem. Biophys. Res. Commun. **81**, 176–185
Robins, S. P. & Duncan A. (1983) Biochem. J. **215**, 175–182
Robins, S. P. & Duncan, A. (1987) Biochim. Biophys. Acta **914**, 233–239
Scott, J. E., Hughes, E. W. & Shuttleworth, A. (1981) Biosci. Rep. **1**, 611–618
Scott, J. E., Qian, R., Henkel, W. & Glanville, R. W. (1983*a*) Biochem. J. **209**, 263–264
Scott, J. E., Hughes, E. W. & Shuttleworth, A. (1983*b*) J. Biochem. (Tokyo) **93**, 921–925
Scott, P. G. (1980) Biochemistry **19**, 6118–6124
Siegel, R. C. (1976) J. Biol. Chem. **251**, 5786–5792
Yamauchi, M. & Mechanic, G. L. (1988) in Collagen (Nimmi, M. E., ed.), vol. 1, pp. 157–172, CRC Press, Boca Raton

Received 3 July 1991/23 September 1991; accepted 4 October 1991