Isolation and Characterization of the Cyanogen Bromide Peptides from the $\alpha 1$ and $\alpha 2$ Chains of Human Skin Collagen*

Eva Marie Click and Paul Bornstein†

ABSTRACT: The $\alpha 1$ and $\alpha 2$ chains of human skin collagen were cleaved with CNBr and the resulting fragments separated by a combination of ion-exchange and molecular sieve chromatography. The eight peptides obtained from the $\alpha 1$ chain varied in molecular weight from 2037 [$\alpha 1$-CB(0,1)] to 23,500 ($\alpha 1$-CB7) and accounted for the amino acid composition and molecular weight of the chain. The data indicate that the two $\alpha 1$ chains in human skin collagen are either identical or very nearly identical in composition and sequence. The six peptides obtained from the $\alpha 2$ chain varied in molecular weight from 270 ($\alpha 2$-CB0) to 29,200 ($\alpha 2$-CB4) and accounted for the amino acid content of $\alpha 2$.

There is a striking conservation of the amino acid composition of collagen during evolution as evidenced by the similarity in the compositions of the human and rat proteins. However, in keeping with studies of other species, the short nonhelical NH_2-terminal sequences of the chains, and particularly that of $\alpha 2$, are more variable in structure. The substitution of leucine in human skin $\alpha 1$ for methionine, present as the second amino acid in the $\alpha 1$ chains of rat tendon and chicken collagen, leads to elimination of the NH_2-terminal dipeptide, $\alpha 1$-CB0, from digests of the human protein. In $\alpha 2$-CB1, at least 3 substitutions occur in the sequence of 14 amino acids. Preliminary evidence indicates that lysyl residues participate in the formation of interchain cross-links, as in the rat and chicken proteins, although the location of the cross-links and extent of the process probably differ from species to species.

Cleavage of polypeptide chains at methionyl residues with CNBr represents an important adjunct to the use of proteolytic enzymes in the elucidation of the covalent structure of proteins. The application of this technique to collagen has proved exceptionally useful because of the limited number of methionyl residues in collagen, the ready solubility of CNBr fragments, and the lack of significant side reactions. Degradation with CNBr first led to the implication of lysyl side chains in the formation of interchain covalent cross-links in rat collagen (Bornstein *et al.*, 1966a; Bornstein and Piez, 1966) and to the demonstration that these bonds are located near the NH_2 terminus of the collagen molecule (Bornstein *et al.*, 1966b). The isolation of the CNBr peptides of the $\alpha 1$ and $\alpha 2$ chains of rat skin collagen (Butler *et al.*, 1967; Fietzek and Piez, 1969) was followed by the determination of the amino acid sequence of some of the smaller fragments (Kang *et al.*, 1967; Bornstein, 1967; Butler, 1970) and by the ordering of these fragments in the polypeptide chains using a combination of chemical, biosynthetic, and electron microscopic techniques (Piez *et al.*, 1969; Vuust *et al.*, 1970).

Useful comparative biochemical information has been obtained by an examination of the CNBr peptides of an avian species, the chicken. The CNBr peptides of both chicken skin collagen (Kang *et al.*, 1969a,b) and chicken bone collagen (Miller *et al.*, 1969; Lane and Miller, 1969) have been isolated and their amino acid compositions determined. The amino acid sequences of the smaller CNBr peptides, located near the NH_2 terminus of the chicken skin collagen molecule, have also been determined (Kang and Gross, 1970) and the order of the peptides in the $\alpha 2$ chain ascertained (Igarashi *et al.*, 1970). At present, information of an incomplete nature is available regarding the CNBr peptides of collagen from three additional species, the cow (Rauterberg and Kuhn, 1968), cod (Laszlo and Olsen, 1969), and rabbit (Bornstein and Nesse, 1970).

The impetus to complete the characterization of the $\alpha 1$ and $\alpha 2$ chains of human collagen derived less from a need to study another mammalian species for comparative purposes than from the importance of the structure of the normal human protein to the evaluation of possible disturbances of collagen structure in pathologic states. This paper describes the separation and analysis of all of the CNBr peptides of the $\alpha 1$ and $\alpha 2$ chains of human skin collagen. The marked similarity in structure between rat and human collagens, suggested by preliminary studies (Bornstein, 1968; Bornstein and Kang, 1970), has been amply confirmed. In the case of every peptide the homology between the human and rat proteins was apparent on the basis of amino acid composition and chromatographic behavior. The nomenclature previously assigned to the rat peptides is therefore used for the human peptides. However, in accordance with current usage (Miller *et al.*, 1969), the NH_2-terminal peptide of the $\alpha 1$ chain is designated $\alpha 1$-CB(0,1) to indicate the substitution of the methionyl residue in the human protein. Evidence is also presented for the similarity of the mechanism of cross-linking in the two proteins.

* From the Department of Biochemistry and Research Training Unit, Department of Medicine, University of Washington, Seattle, Washington 98105. *Received July 27, 1970.* This work was supported by National Institutes of Health Grant No. AM 11248 and a Lederle Medical Faculty award. A portion of this material is taken from a dissertation to be submitted by E. M. C. to the Graduate School of the University of Washington in partial fulfillment of the requirements for the M.S. degree.

† Recipient of Research Career Development Award K4-AM-42582 from the U. S. Public Health Service; to whom to address correspondence.

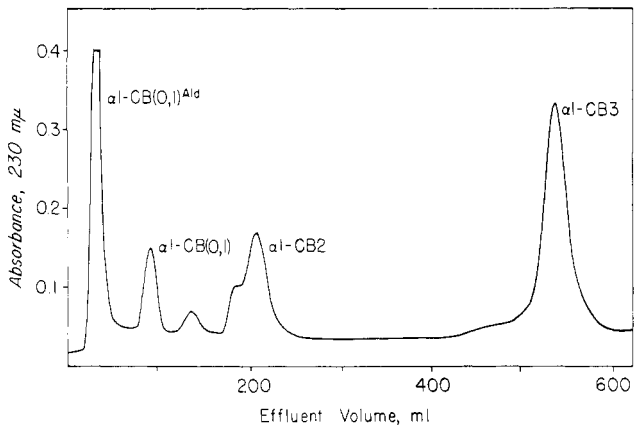

FIGURE 1: Phosphocellulose elution pattern of peptides obtained by CNBr cleavage of the α1 chain of human skin collagen. The sample (65 mg) was dissolved in starting buffer, 1×10^{-3} M sodium acetate (pH 3.8), and eluted with a gradient of NaCl from 0 to 0.3 M over a total volume of 800 ml. The remainder of the α1 peptides were eluted at the conclusion of the gradient with a higher pH and ionic strength buffer and fractionated by other means (see text). The peak eluting between α1-CB(0,1) and α1-CB2 represents the overlapping peptide α1-CB(0,1–2) which results from incomplete cleavage at the methionyl residue linking α1-CB(0,1) and α1-CB2.

Materials and Methods

Preparation of Human Skin Collagen. Samples of normal skin were obtained at postmortem examination from a variety of patients of different ages. Because of the limited extractability of human collagen, pooled skin samples were occasionally used. Care was taken to exclude tissues which might have undergone autolysis. Tissues of patients who had received steroids or anti-metabolic drugs prior to death were also excluded. After removal of the epidermis and subcutaneous tissue with a scalpel the skin was ground to a coarse mince with a mechanical grinder and lyophilized. The dry tissue was then mixed with chips of Dry Ice and milled to a fine powder in a Wiley mill. The skin powder was subjected to sequential extraction with 1 M NaCl containing 0.05 M Tris (pH 7.5), followed by 0.5 M acetic acid, and then by 5 M Gd·HCl[1] (pH 7.5). The resulting collagen or gelatin was purified largely as described previously (Bornstein and Piez, 1964, 1966). The extractions and purifications were performed at 4°. In a typical experiment, 65 g of infant skin (18.9 g dry powder) yielded 605 mg of acid-soluble collagen and 1.3 g of Gd·HCl-soluble material. In another experiment 11.9 g of pooled infant skin powder yielded 185 mg of salt-soluble collagen and 269 mg of acid-soluble collagen.

Preparation of α1, α2, and β_{12}. Collagens were chromatographed on CM-cellulose, after heat denaturation, as described previously (Piez *et al.*, 1963; Bornstein and Piez, 1966). Material, which was insoluble in the starting buffer, 0.06 M sodium acetate (pH 4.8), was removed by centrifugation. The total yield of single- and double-chain components from salt- and acid-extracted collagens, uncorrected for chromatographic and other losses, was 43%. Yields of chromatographically purified components from Gd·HCl-extracted collagen were considerably lower owing to the presence of noncollagenous proteins and higher aggregates of collagen in this material.

CNBr Cleavage. Cleavage of α1, α2, and β_{12} chains with CNBr was performed as described previously (Bornstein and Piez, 1966).

Ion-Exchange Chromatography. Separation of CNBr digests of collagen chains was achieved by a combination of phosphocellulose and CM-cellulose chromatography. In general, initial chromatographic separation of the smaller peptides on phosphocellulose (Bornstein and Piez, 1966) was followed by resolution of the larger fragments on CM-cellulose at pH 3.6 (Butler *et al.*, 1967). Rechromatography on CM-cellulose at pH 4.8 served to purify the larger fragments. A linear rather than the concave NaCl gradient described by Butler *et al.* (1967) was used. Alternatively, unfractionated CNBr digests were subjected to chromatography on CM-cellulose and the smaller unretarded peptides subsequently fractionated by a combination of phosphocellulose and molecular sieve chromatography. The precise steps used in the isolation of individual peptides are described under Results.

Gel Filtration. Collagen chains were separated from buffer salts on Sephadex G-25, coarse beads (Pharmacia), using a pyridine acetate buffer (Piez *et al.*, 1963). CNBr peptides fractionated by ion-exchange chromatography were concentrated by lyophilization and desalted on Bio-Gel P-2 (100–200 mesh, Bio-Rad Laboratories) equilibrated with 0.03 M ammonium propionate (pH 4.5). Fractionation of the smaller peptides was performed on a 1.4×100 cm column of Bio-Gel P-2 (200–400 mesh) equilibrated with ammonium propionate, and on a 2×95 cm column of Sephadex G-50 (fine beads, Pharmacia) equilibrated with 0.2 M NH_4HCO_3 (pH 8.0). Gel filtration on agarose 1.5 m (Bio-Rad Laboratories), equilibrated with 1 M $CaCl_2$–0.05 M Tris (pH 7.5), was used both as a final purification step for some of the larger CNBr peptides and as a means of determining the molecular weights of these fragments (Piez, 1968; Bornstein, 1970).

Enzymatic Hydrolysis. Digestions with trypsin (two-times crystallized, Worthington) and chymotrypsin (three-times crystallized, Worthington) were performed in 0.2 M NH_4HCO_3 (pH 7.8) containing 1×10^{-3} M $CaCl_2$. The enzymes (1% of the substrate by weight) were added as a 0.5% solution in 1×10^{-3} M HCl. Digestion was limited to 4 hr and was terminated by lyophilization.

Amino Acid Sequence Analysis. The partial amino acid sequence of α1-CB(0,1)-T2 was determined by sequential Edman degradation and identification of the original and successively newly formed NH_2-terminal amino acids by dansylation (Gray, 1967). The precise technique and the solvents used in the thin-layer chromatographic identification of dansylamino acids have been described (Bornstein, 1969).

Amino acid analyses were performed on a Beckman 120C analyzer modified for accelerated single-column gradient elution (Miller and Piez, 1966; Bornstein, 1969). Chains and peptides were hydrolyzed *in vacuo* in doubly distilled constant-boiling HCl at 110° for 24 hr. Corrections were made for hydrolytic losses of threonine, serine, methionine, and tyrosine and for incomplete release of valine (Bornstein, 1970).

Aldehyde Determinations. Peptide-bound aldehydes were assayed spectrophotometrically by conversion into azine derivatives with MBTH (Paz *et al.*, 1965). Reactions were

[1] Abbreviations used are: Gd·HCl, guanidine hydrochloride; MBTH, *N*-methylbenzothiazolone hydrazone.

TABLE I: Amino Acid Compositions of the CNBr Peptides from the α1 Chain of Human Skin Collagen.[a]

	α1-CB(0,1)	α1-CB2	α1-CB3	α1-CB4	α1-CB5	α1-CB6	α1-CB7	α1-CB8	Total Peptides	α1 Found[b]
3-Hydroxyproline						0.5			0.5	0.5
4-Hydroxyproline		5.5	14.3	5.3	3.0	15.3	23.0	29.3	96	94
Aspartic acid[c]	1.0		6.3	3.0	3.0	9.2	12	9.1	43	45
Threonine	1.0			0.9		4.0	4.9	5.4	16	16
Serine	2.9	1.8	3.2		1.8	10	8.9	8.2	37	37
Homoserine[d]	0.9	1.0	0.9	1.0	0.9		1.0	0.9	7	6.1[f]
Glutamic acid[e]	2.1	3.9	16	3.2	3.2	13	16	19	76	80
Proline	2.1	6.0	14.4	5.7	2.2	33.0	36.3	31.0	131	134
Glycine	4.2	12	50	16	12	68	88	86	336	340
Alanine		2.1	22	3.4	4.0	21	34	33	119	122
Valine	1.2		4.4			4.3	5.9	6.1	21	22
Isoleucine	1.1					2.1	2.9	1.1	7	7.2
Leucine	1.1	1.0	3.1	2.0	1.1	3.9	4.0	3.9	20	20
Tyrosine	1.8								2	2.2
Phenylalanine		1.0	3.0		0.9	1.9	3.0	2.8	13	13
Hydroxylysine			0.2	0.1	1.4	1.7	0.7	0.7	4.8	4.3
Lysine	0.9		4.7	2.1	1.8	3.8	9.4	8.2	30.8	30.6
Histidine					0.9	0.9			2	2.2
Arginine		1.1	6.3	3.9	1.3	11	13	14	50	50
Total	20	36	147	46	37	204	263	258	1012	1026

[a] Values are expressed as residues per peptide. Actual values are listed for amino acids present as less than 10 residues and for hydroxyproline, proline, hydroxylysine-and lysine (since partial hydroxylation can give rise to noninteger values for these amino acids). A space indicates less than 0.2 residue. [b] Values are averages of three determinations and are calculated for a molecular weight of 93,000 and an average residue weight of 90.6. [c] Includes asparagine. [d] Includes homoserine lactone. [e] Includes glutamine. [f] As methionine.

followed in a Cary Model 15 recording spectrophotometer with the cell compartment jacketed at 40°.

Results

CNBr Peptides Derived from α1

α1-CB(0,1) and α1-CB(0,1)^{Ald}. Chromatography of a CNBr digest of the α1 chain on phosphocellulose (Figure 1) permitted separation of the peptides derived from the NH_2-terminal region. This sequence is present, as in the case of rat collagen, in two forms, α1-CB(0,1) which contains a lysyl residue, and α1-CB(0,1)^{Ald} which contains instead a lysyl-derived aldehyde. Since the starting material included a significant fraction of the α1-α1 dimer, β_{11}, the initial peak in the phosphocellulose elution pattern represents both α1-CB(0,1)^{Ald} and β_{11}-CB1. This identification was supported by the demonstration that the MBTH-produced azine derivatives of the peptide material contained in this peak yielded absorption maxima at both 310 and 330 mμ, indicative of the presence of both saturated and α,β-unsaturated aldehydes (Paz *et al.*, 1965; Bornstein and Piez, 1966).

The amino acid compositions of α1-CB(0,1) (Table I) and α1-CB(0,1)^{Ald} differ only in the presence of a single lysyl residue in the former peptide. α1-CB(0,1)^{Ald} contains instead an aldehydic function, presumably a lysyl-derived δ-semialdehyde of α-aminoadipic acid, as is found in the rat (Bornstein and Piez, 1966).

α1-CB(0,1) was treated with trypsin and the digest chromatographed on Bio-Gel P-2 (200–400 mesh). The two tryptic peptides α1-CB(0,1)-T1 and α1-CB(0,1)-T2 were readily separable, despite similar molecular weights, due to the presence of two tyrosyl residues in α1-CB(0,1)-T1 which led to adsorption of the peptide to the acrylamide gel and its delayed elution. The amino acid compositions of T1 and T2 (Table II) account for that of the starting peptide. The NH_2-terminal sequence of T2 was found to be Ser-Thr-Gly-Gly-. The NH_2 terminus of T1 was blocked, presumably by pyrrolidonecarboxylic acid, preventing sequence determination of this peptide by Edman degradation.

The partial sequence of α1-CB(0,1), arranged to maximize homology with α1-CB0 and α1-CB1 from rat tail tendon collagen, is shown in Figure 2. The substitution of leucine for methionine in position 2 of the α1 chain accounts for the

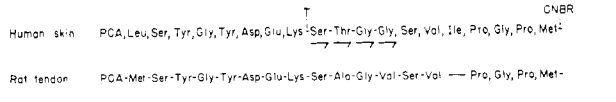

FIGURE 2: Partial amino acid sequence of α1-CB(0,1) arranged to maximize the homology with α1-CBO and α1-CB1 from rattail tendon (Bornstein, 1969). The vertical arrows indicate points of cleavage by CNBr and trypsin (T). The horizontal arrows indicate the sequence determined by Edman degradation. PCA indicates pyrrolidonecarboxylic acid.

TABLE II: Amino Acid Composition of Tryptic Peptides from α1-CB(0,1) and Chymotryptic Peptides from α1-CB2.[a]

	α1-CB(0,1)			α1-CB2		
		T1	T2		C1	C2
4-Hydroxyproline				5.5	2.4	2.3
Aspartic acid	1.0	1.0	(0.2)			
Threonine	1.0		1.0			
Serine	2.9	1.1	1.9	1.8	0.9	1.0
Homoserine[b]	0.9		0.9	1.0		0.9
Glutamic acid	2.1	2.0	(0.2)	3.9	1.0	3.0
Proline	2.1		1.9	6.0	3.6	2.2
Glycine	4.2	1.2	3.3	11.5	6.6	4.9
Alanine	(0.2)			2.1	1.1	1.1
Valine	1.2		1.1			
Isoleucine	1.1		1.0			
Leucine	1.1	1.0		1.0	1.0	
Tyrosine	1.8	1.9				
Phenylalanine				1.0	1.0	
Lysine	0.9	0.9				
Arginine				1.1	1.1	

[a] Values are given as residues per peptide. A space indicates that the amino acid was either entirely absent or present as less than 0.1 residue. Residues in parentheses are fractional residues thought to be impurities. [b] Includes homoserine lactone.

absence of the dipeptide pyrrolidonecarboxylic acid-Hse (α1-CB0) from digests of human skin α1.

α1-CB2 elutes from phosphocellulose after α1-CB(0,1) (Figure 1). The amino acid composition of this peptide (Table I) is identical with that of the rat skin α1-CB2 except for a possible increase in the degree of hydroxylation of the prolyl residues. The compositions of the chymotryptic peptides resulting from cleavage at the phenylalanyl residue (Table II) are also the same as the corresponding peptides from rat collagen (Bornstein, 1967). It is therefore likely that the sequences of rat and human α1-CB2 are identical.

α1-CB3 was obtained by chromatography of a CNBr digest of α1 on phosphocellulose (Figure 1). Alternatively, the peptide can be separated from digests of β_{12} either by chromatography on phosphocellulose (chromatogram not shown) or on CM-cellulose at pH 3.6 (Figure 3). In the latter case the peptide

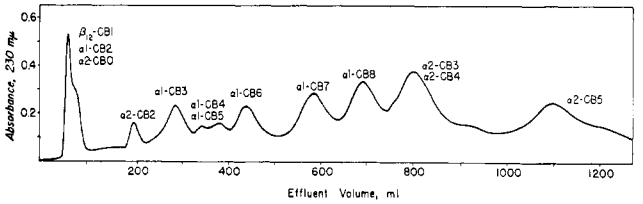

FIGURE 3: CM-cellulose elution pattern of peptides obtained by CNBr cleavage of the β_{12} component of human skin collagen. The sample (75 mg) was dissolved in starting buffer, 0.02 M sodium citrate–0.04 M NaCl (pH 3.6), and eluted with a gradient of NaCl from 0.04 to 0.14 M over a total volume of 1600 ml.

TABLE III: Molecular Weights of the CNBr Peptides from the α1 Chain of Human Skin Collagen.

Peptide	Amino Acid Anal.	Mol Sieve Chromatography[a]
α1-CB(0,1)	2,037	N.D.
α1-CB2	3,294	N.D.
α1-CB3	13,328	12,200
α1-CB4	4,372	N.D.
α1-CB5	3,624	N.D.
α1-CB6	18,388	16,200
α1-CB7	23,506	25,000
α1-CB8	23,271	24,000
Total	91,820	

[a] N.D. = not determined.

must be purified by rechromatography on CM-cellulose at pH 4.8. The amino acid composition of α1-CB3 (Table I) is very similar to that of rat skin α1-CB3 and a high degree of homology is indicated by the virtual identity of fingerprints of tryptic digests of the two peptides (Bornstein, 1968). The molecular weight of α1-CB3, determined by molecular sieve chromatography, agrees well with that calculated by amino acid analysis (Table III).

α1-CB4 and α1-CB5 were obtained from CNBr digests of β_{12} (Figure 3) and α1 (Figure 4) chromatographed on CM-cellulose at pH 3.6. In both instances the peptides were separated as a mixture from adjacent larger peptides by gel filtration on Sephadex G-50. α1-CB4 and α1-CB5 were resolved by rechromatography on CM-cellulose equilibrated with 0.02 M sodium acetate (pH 4.8). A linear gradient of NaCl from 0 to 0.10 M over a total volume of 800 ml was used. The amino acid compositions of the two peptides are tabulated in Table I.

α1-CB6. The COOH-terminal fragment of the α1 chain was isolated from digests of both α1 and β_{12} (Figures 3 and 4). Purification was achieved by rechromatography on CM-cellulose equilibrated with 0.02 M sodium acetate (pH 4.8). A linear gradient of NaCl from 0 to 0.14 M over a total volume of 800 ml was used. The amino acid composition of α1-CB6

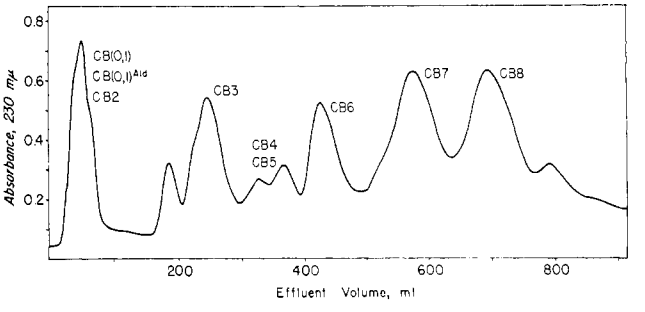

FIGURE 4: CM-cellulose elution pattern of peptides obtained by CNBr cleavage of the α1 chain of human skin collagen. The sample (105 mg) was dissolved in starting buffer, 0.02 M sodium citrate–0.04 M NaCl (pH 3.6), and eluted with a gradient of NaCl from 0.04 to 0.14 M over a total volume of 1600 ml.

TABLE IV: Amino Acid Compositions of the CNBr Peptides from the α2 Chain of Human Skin Collagen.[a]

	α2-CB0	α2-CB1[b]	α2-CB2	α2-CB3	α2-CB4	α2-CB5	Total Peptides	α2 Found[c]
3-Hydroxyproline						0.5	0.5	0.6
4-Hydroxyproline			2.7	22.5	29.2	25.5	80	79
Aspartic acid[d]		1.0	2.0	17	11	15	46	49
Threonine				4.9	5.9	7.0	18	19
Serine			1.9	9.3	11	10	32	33
Homoserine[e]	0.9	0.9	0.9	1.0	0.9		5	5.5[g]
Glutamic acid[f]		1.0	1.2	24	23	20	69	71
Proline		2.0	3.1	39.3	32.0	36.3	113	113
Glycine	1.2	5.0	10	110	108	104	338	343
Alanine			3.2	39	39	30	111	110
Valine		1.0	1.0	10	13	13	38	39
Isoleucine				6.1	3.9	6.0	16	17
Leucine	0.9	1.0	1.0	7.5	12	10	33	33
Tyrosine		0.9				1.5	3	3.6
Phenylalanine				4.0	4.0	3.5	12	11
Hydroxylysine				1.8	4.0	2.0	7.8	9.0
Lysine		0.3		9.0	6.0	5.7	21	20
Histidine				1.7	2.0	6.0	10	11
Arginine			2.8	16	17	16	52	53
Total	3	14	30	324	322	314	1005	1020

[a] Values are expressed as residues per peptide. Actual values are listed for amino acids present as less than 10 residues and for hydroxyproline, proline, hydroxylysine, and lysine (since partial hydroxylation can give rise to noninteger values for these amino acids). A space indicates less than 0.2 residue. [b] Represents a mixture of α2-CB1 and α2-CB1^{Ald}. [c] Values are the averages of five determinations and were calculated for a molecular weight of 93,000 and an average residue weight of 91.4. [d] Includes asparagine. [e] Includes homoserine lactone. [f] Includes glutamine. [g] As methionine.

is listed in Table I and its molecular weight, estimated by both amino acid analysis and gel filtration, in Table III.

α1-CB7 and α1-CB8. These CNBr peptides were obtained from digests of both α1 and β_{12} (Figures 3 and 4). Both peptides were rechromatographed on CM-cellulose at pH 4.8 as described for α1-CB6. The amino acid compositions of α1-CB7 and CB8 are listed in Table I and their molecular weights, determined by agarose gel filtration, are listed in Table III.

The compositions of the eight CNBr peptides account, within experimental error, for the composition of the entire α chain (Table I). Good agreement exists between the molecular weights of the larger CNBr peptides calculated by amino acid analysis and by molecular sieve chromatography (Table III).

CNBr Peptides Derived from α2

Because of the limited availability of α2, considerable reliance was placed on the α1-α2 dimer, β_{12}, as a source of CNBr peptides from the α2 chain. However, fractionation and characterization of digests of α2, together with the homology with the rat α2 chain, made the identification of the α2 peptides unambiguous.

α2-CB0 was isolated primarily from β_{12}. A CNBr digest of β_{12} was placed on a column of Bio-Gel P-2 (200–400 mesh) and the bulk of the peptides eluting at or shortly after the void volume of the column were separated from the remainder of the effluent volume. The latter fractions, including the totally included volume of the column, were concentrated by lyophilization and rechromatographed. The fractions comprising the included volume of the column, although showing little or no absorbance at 230 mμ, were pooled and examined by amino acid analysis (Table IV). A tripeptide (Gly, Leu, Hse) identical in composition with a peptide from the rat α2 chain (Fietzek and Piez, 1969) was demonstrated in this manner.

α2-CB1 and α2-CB1^{Ald}. The NH_2-terminal sequence of the α2 chain was isolated by chromatography of a CNBr digest of α2 on CM-cellulose (Figure 5) or on phosphocellulose (chromatogram not shown). In both chromatograms the lysine- and aldehyde-containing forms of the peptide eluted at the start of the gradient. Separation of a small amount of β_{12}-CB1 (resulting from contamination of α2 with β_{12}) was achieved by filtration on Bio-Gel P-2 (200–400 mesh). The analysis listed in Table IV represents a mixture of α2-CB1^{Ald} and α2-CB1, with the former peptide predominating.

α2-CB2 was isolated by CM-cellulose chromatography of a digest of α2 (Figure 5) as well as by phosphocellulose chromatography of a digest of β_{12} (chromatogram not shown). The amino acid composition of the peptide is listed in Table IV. The peak following α2-CB2 in Figure 5 represents incompletely cleaved material.

α2-CB3, α2-CB4, and α2-CB5. As in the case of the rat α2 chain, three large CNBr peptides accounted for more than 90% of the length of the human α2 chain. Separation of these

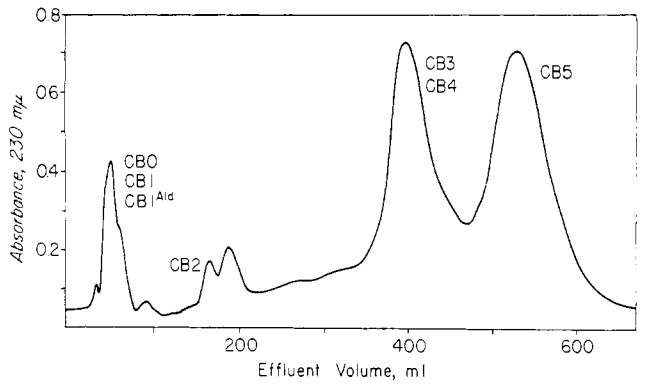

FIGURE 5: CM-cellulose elution pattern of peptides obtained by CNBr cleavage of the $\alpha 2$ chain of human skin collagen. The sample (40 mg) was dissolved in starting buffer, 0.02 M sodium citrate–0.04 M NaCl (pH 3.6), and eluted with a gradient of NaCl from 0.04 to 0.24 M over a total volume of 1200 ml.

peptides proved difficult because of their similar size and charge. However, an initial separation was achieved by chromatography of a digest of β_{12} on CM-cellulose at pH 3.6 (Figure 3). $\alpha 2$-CB3 and $\alpha 2$-CB4 eluted together, well separated from $\alpha 1$-CB8 and $\alpha 2$-CB5. Satisfactory separation of the mixture of $\alpha 2$-CB3 and $\alpha 2$-CB4 was effected by rechromatography on CM-cellulose equilibrated with 0.02 M sodium acetate (pH 4.8). A linear gradient of NaCl from 0 to 0.1 M over a total volume of 800 ml was used. Gel filtration on agarose 1.5 m served as a final purification step and as a means of estimating the molecular weights of the fragments.

$\alpha 2$-CB5 was also purified further by agarose chromatography. This peptide, as obtained by CM-cellulose chromatography, was contaminated with a higher molecular weight fragment which on amino acid analysis had a composition consistent with that of $\alpha 2$-CB5 plus $\alpha 2$-CB3. Similar overlapping peptides designated $\alpha 2$-CB(3-5) were isolated from digests of both chicken and rat $\alpha 2$ chains and assisted in the determination of the order of the CNBr peptides in the $\alpha 2$ chain (Lane and Miller, 1969; Vuust *et al.*, 1970).

Alternatively, $\alpha 2$-CB3, $\alpha 2$-CB4, and $\alpha 2$-CB5 were obtained from a digest of $\alpha 2$. Initial separation on CM-cellulose at pH 3.6 using a higher ionic strength gradient yielded two peaks (Figure 5). The first peak proved to be a mixture of $\alpha 2$-CB3 and $\alpha 2$-CB4 and the second peak a mixture of $\alpha 2$-CB5 and incompletely cleaved material. Separation of all three peptides was again achieved by rechromatography on CM-cellulose at pH 4.8. The amino acid compositions of $\alpha 2$-CB3, $\alpha 2$-CB4, and $\alpha 2$-CB5 are listed in Table IV. Estimates of the molecular weights of these fragments by gel filtration agreed with the values obtained by amino acid analysis.

Discussion

The 14 peptides derived from $\alpha 1$ and $\alpha 2$ adequately account for the amino acid content of the chromatographically purified chains of human skin collagen. Although recoveries of the peptides were not quantitated, gross yields were consistent with their existence in stoichiometric amounts in the chains and indicated the presence of two identical (or nearly identical) $\alpha 1$ chains in the triple-stranded molecule. In calculating the number of amino acids in each chain, the assumption has been made that the molecular weights of human and rat collagen are very similar. Similar molecular weights for rat and human chains are indicated by sedimentation equilibrium measurements (Bornstein and Piez, 1964; Lewis and Piez, 1964), by the comparable position of elution of these chains from agarose gel filtration columns, and by the marked structural homology of all the rat and human CNBr peptides. The value of 93,000 chosen to calculate amino acid content corresponds more closely to recent measurements of the molecular weight of rat chains by sedimentation equilibrium (Kang *et al.*, 1966) and to the results of molecular weight determinations of these chains by molecular sieve chromatography. The amino acid compositions of the $\alpha 1$ and $\alpha 2$ chains listed in Tables I and IV are essentially identical with those previously reported (Bornstein and Piez, 1964).

In view of the obvious homology of the rat, chicken, and human CNBr peptides it can be assumed that the order of the peptides in the chains is the same, *viz.*, 0,1-2-4-5-8-3-7-6 for $\alpha 1$ (Piez *et al.*, 1969); and 1-0-4-2-3-5 for $\alpha 2$ (Vuust *et al.*, 1970; Igarashi *et al.*, 1970). The experiments described in this work also suggest that the mechanism of interchain cross-linking is the same in rat and human collagen. Thus a lysyl-derived aldehyde exists near the NH_2 terminus of the human $\alpha 1$ chain, and cleavage of β_{12} with CNBr yields a peptide, β_{12}-CB1, which apparently results, as in the rat, from an aldol condensation of lysyl-derived aldehydes on the $\alpha 1$ and $\alpha 2$ chains (Bornstein and Piez, 1966; Rojkind *et al.*, 1969). In support of this mechanism the amino acid composition of β_{12}-CB1 (not reported) was found to represent the sum of the compositions of $\alpha 1$-CB(0,1)^{Ald} and $\alpha 2$-CB1^{Ald}, and β_{12}-CB1 contained an α,β-unsaturated aldehyde as indicated by the ultraviolet absorption spectrum of the MBTH derivative (Paz *et al.*, 1965; Bornstein and Piez, 1966). The identification of the cross-linking mechanism in human collagen is an important preliminary step in the evaluation of possible disturbances of cross-linking in various pathologic states.

In contrast to previous structural studies of the rat and chicken proteins, the lack of large amounts of tissue from young healthy individuals and the inability to induce the lathyritic state made it necessary to rely on more highly cross-linked collagen as a source for a large part of the CNBr peptides from human skin. As demonstrated in Figure 3, the β_{12} component can readily be used as a starting material for those CNBr peptides which are not involved in cross-link formation. It may also be possible to identify regions of the chains involved in cross-link formation by the cleavage of β_{12} and higher oligomers obtained from Gd·HCl-extracted collagen or insoluble collagen, followed by fractionation of the resulting peptide mixtures. In view of the marked insolubility of human skin collagen it is likely that a more extensive network of interchain bonds will be found to exist in this protein in comparison with rat collagen.

The striking similarity in structure between rat and human collagens is indicated by the nearly identical compositions of the homologous CNBr peptides. Even when the restriction imposed by the requirement that glycine be in every third position is taken into account, the degree of identity of all the $\alpha 1$ peptides, except $\alpha 1$-CB(0,1), exceeds 90%. The very similar molecular weights indicate that few, if any, deletions or insertions have occurred. In the case of $\alpha 1$-CB2 and $\alpha 1$-CB3 the similarity in amino acid sequence is further demonstrated by the identity or marked similarity of proteolytic digests

of the two fragments. The fewer number and larger size of the $\alpha 2$ peptides make it difficult to determine whether the degree of interspecies homology of the $\alpha 2$ chain is as great as that of the $\alpha 1$ chain, but the similarity is clearly considerable.

To a large extent the structural homology of collagen molecules in the human, rat, and chicken has been made apparent by the constancy of the methionyl residues in these proteins. Only two substitutions involving methionine occur. Leucine replaces methionine in human $\alpha 1$-CB(0,1) and an additional methionine is present in the COOH-terminal region of the chicken $\alpha 1$ chain. The latter substitution leads to the formation of two CNBr peptides from the sequence comprising $\alpha 1$-CB6 (Kang *et al.*, 1969b; Miller *et al.*, 1969). Whether the relative invariance of methionine exceeds that of other similar amino acids is unclear. The constancy of methionyl residues does not extend to the fish proteins since cod collagen contains a much higher content of methionine than do the mammalian and avian proteins (Piez, 1965; Laszlo and Olsen, 1969).

It may be presumed that, in addition to the requirements of the collagen helix, the numerous intermolecular interactions involving collagen severely limit the degree to which functionally acceptable amino acid substitutions can occur. These interactions may include not only the electrostatic and hydrophobic bonds which are thought to play a role in the aggregation of collagen, but also the bonds which play a role in the association of collagen with protein–polysaccharides, glycoproteins, and other components of the intercellular matrix. A substitution which could be compatible with one function of collagen may therefore be rejected by natural selection because the altered protein participates less effectively in its association with other macromolecules. In an analogous fashion, the phylogenetically invariant sequence which exists in cytochrome *c* may have remained constant due to a need both to be compatible with the specific function of the protein and to interact with a portion of the mitochondrial membrane (Margoliash and Schejter, 1966).

Recently, a striking similarity in the segment-long-spacing collagen aggregates from two species of invertebrates, a fish, and a mammal, was demonstrated (Nordwig and Hayduk, 1969). The cross-striations observed in segment-long-spacing aggregates are a reflection of the distribution of charged amino acids in the collagen molecule and similarities in pattern indicate a gross similarity (although not necessarily an identity) in amino acid sequence. This relative constancy despite the wide phylogenetic spread in the source of the proteins attests to the stringent restrictions imposed on the structural variability of collagen.

In contrast to the main body of the collagen molecule, the NH_2-terminal sequences of the chains, particularly that of the $\alpha 2$ chain, demonstrate a somewhat greater interspecies variability (Bornstein, 1968; Bornstein and Kang, 1970). In keeping with this generalization, the composition of human $\alpha 2$-CB1 differs appreciably from that of the homologous rat, chicken, and rabbit peptides whereas $\alpha 1$-CB1 from all four species demonstrates a greater structural uniformity. The NH_2-terminal sequences of the collagen molecule play an important role in the cross-linking process and structural changes in this region could dictate interspecies variation in the ability of collagen to aggregate and polymerize. Since these sequences, by virtue of their amino acid composition and primary structure, cannot assume a triple-helical conformation they are probably less subject to the restrictions imposed upon evolutionary change in the remainder of the molecule.

A consequence of the interspecies variability of the NH_2-terminal sequences in collagen, particularly of the $\alpha 2$ chain, is the finding that when rabbits are immunized with the native rat protein a significant portion of the antigenicity of rat collagen is localized to the sequence encompassed by $\alpha 2$-CB1 (Michaeli *et al.*, 1969). The elucidation of the structure of rabbit $\alpha 1$-CB1 and $\alpha 2$-CB1 *vis a vis* the corresponding sequences in the rat protein has provided a rational explanation for these findings (Bornstein and Nesse, 1970). In view of the structure of human $\alpha 1$-CB(0,1) and $\alpha 2$-CB1 it is likely that a portion of the antigenicity of the native human protein will also be accounted for by $\alpha 2$-CB1. However, antigenic determinants can be identified in other regions of both the rat $\alpha 1$ and $\alpha 2$ chains when individual chains are used as immunizing antigens and when denatured rather than native collagen is used as the detecting antigen (Lindsley *et al.*, 1970). Similar findings will probably be obtained with the use of human collagen chains as antigens.

Acknowledgment

The skillful technical assistance of Felicia Arguelles and Nancy Rittall is gratefully acknowledged.

References

Bornstein, P. (1967), *J. Biol. Chem. 242*, 2572.
Bornstein, P. (1968), *Science 161*, 592.
Bornstein, P. (1969), *Biochemistry 8*, 63.
Bornstein, P. (1970), *Biochemistry 9*, 2408.
Bornstein, P., and Kang, A. H. (1970), *in* Chemistry and Molecular Biology of the Intercellular Matrix, Vol. I, Balazs, E. A., Ed., New York, N. Y., Academic, p 99.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966a), *Proc. Nat. Acad. Sci. U. S. 55*, 417.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966b), *Biochemistry 5*, 3803.
Bornstein, P., and Nesse, R. (1970), *Arch. Biochem. Biophys. 138*, 443.
Bornstein, P., and Piez, K. A. (1964), *J. Clin. Invest. 43*, 1813.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Butler, W. T. (1970), *Biochemistry 9*, 44.
Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.
Fietzek, P. P., and Piez, K. A. (1969), *Biochemistry 8*, 2129.
Gray, W. R. (1967), *Methods Enzymol. 11*, 139, 469.
Igarashi, S., Kang, A. H., and Gross, J. (1970), *Biochem. Biophys. Res. Commun. 38*, 697.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.
Kang, A. H., and Gross, J. (1970), *Biochemistry 9*, 796.
Kang, A. H., Igarashi, S., and Gross, J. (1969a), *Biochemistry 8*, 3200.
Kang, A. H., Nagai, Y., Piez, K. A., and Gross, J. (1966), *Biochemistry 5*, 509.
Kang, A. H., Piez, K. A., and Gross, J. (1969b), *Biochemistry 8*, 1506.
Lane, J. M., and Miller, E. J. (1969), *Biochemistry 8*, 2134.
Laszlo, F., and Olsen, B. R. (1969), *Eur. J. Biochem. 11*, 140.

Lewis, M. S., and Piez, K. A. (1964), *Biochemistry 3*, 1126.
Lindsley, H., Mannik, M., and Bornstein, P. (1970), *J. Clin. Invest. 49*, 59a.
Margoliash, E., and Schejter, A. (1966), *Advan. Protein Chem. 21*, 113.
Michaeli, D., Martin, G. R., Kettman, J., Benjamini, E., Leung, D. Y. K., and Blatt, B. A. (1969), *Science 166*, 1522.
Miller, E. J., Lane, J. M., and Piez, K. A. (1969), *Biochemistry 8*, 30.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 328.
Nordwig, A., and Hayduk, U. (1969), *J. Mol. Biol. 44*, 161.
Paz, M. A., Blumenfeld, O. O., Rojkind, M., Henson, E., Furfine, C., and Gallop, P. M. (1965), *Arch. Biochem. Biophys. 109*, 548.
Piez, K. A. (1965), *Biochemistry 4*, 2590.
Piez, K. A. (1968), *Anal. Biochem. 26*, 305.
Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.
Piez, K. A., Miller, E. J., and Lane, J. M. (1969), *Biochem. Biophys. Res. Commun. 37*, 801.
Rauterberg, J., and Kuhn, K. (1968), *FEBS (Fed. Eur. Biol. Soc.) Lett. 1*, 230.
Rojkind, M., Gutierrez, A. M., Zeichner, M., and Lent, R. W. (1969), *Biochem. Biophys. Res. Commun. 36*, 350.
Vuust, J., Lane, J. M., Fietzek, P. P., Miller, E. J., and Piez, K. A. (1970), *Biochem. Biophys. Res. Commun. 38*, 703.