Isolation and Characterization of the Cyanogen Bromide Peptides from the α1 Chain of Rat Skin Collagen*

William T. Butler,† Karl A. Piez, and Paul Bornstein‡

ABSTRACT: The α1 chain of rat skin collagen was cleaved at methionyl residues with cyanogen bromide and the resulting fragments were separated by ion-exchange chromatography and gel filtration. Eight unique peptides were obtained in approximately equimolar amounts. The molecular weights ranged from 1500 to 24,000 and totaled 90,000 by sedimentation equilibrium and 93,000 by amino acid analysis, in good agreement with the known weight of the α chain. Furthermore, these peptides account for all of the amino acids in the α1 chain within experimental error. Although additional chromatographic components were observed in small amounts, they could not be distinguished from one of the other peptides by amino acid analysis, acrylamide gel electrophoresis, or an examination of their tryptic maps. It is likely that each of these minor components represents the same amino acid sequence as one of the major peptides. These data give evidence that the eight major peptides represent the whole sequence of the α1 chain of rat skin collagen. Since the α1 chain contains seven methionyl residues, the finding of eight peptides indicates that the two α1 chains of the collagen molecule from rat skin have identical or very similar sequences. These results constitute a suitable first step in the determination of the primary structure of the α1 chain of collagen.

A number of studies have been directed toward derivation of sequence information from digests of collagen (see Gallop and Seifter, 1966; Hannig and Nordwig, 1967). Although these efforts have yielded valuable information concerning the pattern of the primary structure of collagen, sequences have been obtained only for random short regions. It is not possible either to order these sequences or to determine the part of the molecule from which they were obtained. This is a particularly difficult problem since the collagen molecule contains three polypeptide chains, each having a molecular weight of about 95,000. Most collagens appear to contain two similar or identical chains (the α1 chain) and a third chain (the α2 chain) that is readily isolated by chromatography and has a different amino acid composition (Piez *et al.*, 1961, 1963; Schleyer, 1962; Lewis and Piez, 1964; Bornstein and Piez, 1964; Miller *et al.*, 1967; Kao *et al.*, 1967). All three α chains in the collagen from codfish skin, however, have different chromatographic properties and amino acid compositions (Piez, 1964, 1965). Reports that calf skin collagen (Heidrich and Wynston, 1965; Francois and Glimcher, 1966a) and chick bone collagen (Francois and Glimcher, 1966b) also contain three different α chains have been questioned (Piez *et al.*, 1966; Miller *et al.*, 1967).

In order to simplify the starting material in studies of the primary structure of the collagen molecule, we have utilized isolated α1 or α2 chains from rat skin collagen and have employed cleavage at methionyl residues with CNBr (Bornstein and Piez, 1965; Bornstein *et al.*, 1966a; Bornstein and Piez, 1966). Since the α chains of mammalian collagens contain only about six to eight residues of methionine, CNBr cleavage yields a relatively small number of peptides of a size suitable for further study. The amino acid sequences of a pentadecapeptide (α1-CB1) from the cross-linking region at the NH_2-terminal end of α1, a homologous tetradecapeptide (α2-CB1) from α2, and a peptide of 36 amino acids (α1-CB2) adjacent to α1-CB1 in the chain have been reported (Kang *et al.*, 1967; Bornstein, 1967). We report here the isolation and some of the properties of what appear to be the remainder of the CNBr peptides from the α1 chain of rat skin collagen.

Experimental Section

Preparation of α1. Salt- and acid-extracted collagens were prepared as previously described (Bornstein and Piez, 1966) from the skins of 80–100-g Sprague–Dawley rats. Both normal rats and rats made lathyritic by feeding a diet containing 0.1% β-aminopropionitrile fumarate for 3 weeks were utilized. Preparations of α1 from normal and lathyritic rats, whether obtained by salt or acid extraction, were not found to yield different results in any of the experiments reported

* From the Laboratory of Biochemistry, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20014. *Received September 14, 1967.* A portion of this work was presented at the 1967 meeting of the Federation of American Societies for Experimental Biology, Chicago, Ill. (Butler and Piez, 1967).

† Present address: The Institute of Dental Research, The University of Alabama Medical Center, Birmingham, Ala.

‡ Present address: Department of Medicine, University of Washington School of Medicine, Seattle, Wash.

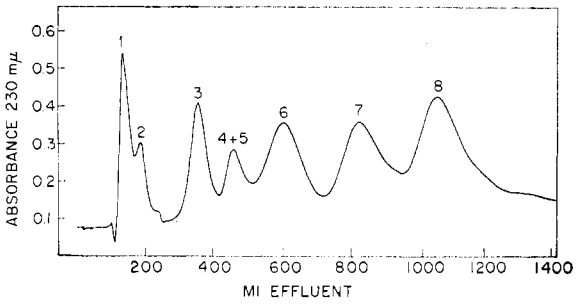

FIGURE 1: Chromatography of CNBr-cleaved α1 of rat skin collagen on CM-cellulose at pH 3.6, 40°. Elution was with a linear gradient from 0.02 M sodium citrate–0.04 M NaCl (pH 3.6) to 0.02 M sodium citrate–0.14 M NaCl (pH 3.6) over a volume of 1600 ml.

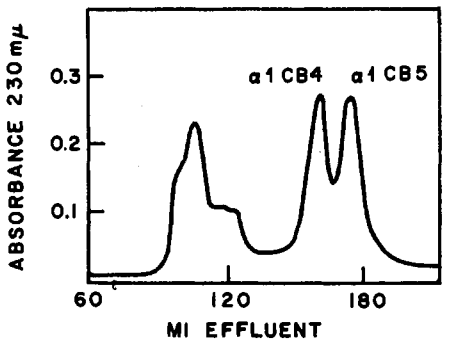

FIGURE 2: Gel filtration of peak 4 + 5 (see Figure 1) on Sephadex G-50. Elution was with 0.03 M ammonium propionate (pH 4.8). The material eluting near the void volume contains peptides α1-CB3 and α1-CB6.

here. The results therefore are given without distinguishing the source of the α1.

α1 was obtained from denatured collagen by chromatography at 40° on 2.5 × 10 cm columns of CM-cellulose (Whatman CM 32) as previously described (Piez *et al.*, 1963; Bornstein and Piez, 1966). The effluent was monitored at 230–236 mμ in a Beckman DB spectrophotometer.

A useful modification of this preparative procedure was found to be elution of the CM-cellulose columns with a gradient prepared from volatile buffers. Stock ammonium propionate buffer was prepared from 131 ml of redistilled propionic acid and 57 ml of concentrated ammonia diluted to 250 ml with water. This stock solution has a nominal ionic strength of 3.5 and solutions prepared by diluting the stock have a pH of 4.8. The CM-cellulose column was eluted at a rate of 300 ml/hr with a linear gradient between 400 (0.07 Γ/2) (1:50 dilution of stock) and 400 ml (0.19 Γ/2) (11:200 dilution of stock). Because of the volatility of the buffer components, it was not possible to deaerate prior to chromatography. Therefore, to prevent the formation of air bubbles in the effluent solution which results in blocking of the light beam through the flow cell, it was found desirable to: (1) wash the column at room temperature after each run with 0.5 M NaCl–0.01 N NaOH that had been deaerated by boiling and cooled just before use; (2) place the fraction collector above the flow cell and column to develop a slight back pressure; and (3) cool the sample compartment of the spectrophotometer. If necessary, place a tee in the line between the column and flow cell to trap bubbles (Piez *et al.*, 1963). The effluent was monitored at 236 mμ. Although a gradual rise in base line accompanied the increase in ionic strength during gradient elution, this did not interfere with the selection of the desired fractions. After lyophilization of appropriate fractions, it was found desirable to redissolve the samples in water by warming them to 40° for a few minutes and to relyophilize them; in this way light fluffy samples which were more easily handled were obtained.

CNBr Cleavage. Approximately 100-mg samples of α1 were cleaved with CNBr as previously described (Bornstein and Piez, 1966). Instead of diluting and lyophilizing to remove excess reagent and solvent after cleavage, some samples were desalted on Bio-Gel P2 (Bio-Rad Laboratories) equilibrated with 0.03 M ammonium propionate (pH 4.5) or with 0.1 M acetic acid, and then lyophilized without diluting.

Chromatography of CNBr Peptides. CM-cellulose columns identical with those used for the preparation of α1 were equilibrated with 0.02 M citrate (pH 3.6) containing 0.04 M NaCl. Samples of CNBr peptides (80–200 mg) were dissolved in 20 ml of this buffer, warmed to 40°, and pumped onto the column. Elution was carried out using a linear gradient between 800 ml each of 0.02 M citrate–0.04 M NaCl (pH 3.6) and 0.02 M citrate–0.14 M NaCl (pH 3.6). The flow rates were 140–200 ml/hr. Effluent fractions were monitored at 228–232 mμ with a Beckman DB spectrophotometer. Fractions of 10 ml were collected.

Further purification of the larger peptides (α1-CB3, α1-CB6, α1-CB7, and α1-CB8) was accomplished by chromatography on CM-cellulose columns at pH 4.8 as described earlier (Bornstein and Piez, 1965). Elution was with a concave gradient between 1000 ml of 0.02 M sodium acetate (pH 4.8) in a 1-l. beaker (mixing chamber) and 740 ml of 0.02 M acetate–0.14 M NaCl (pH 4.8) in a 750-ml erlenmeyer flask.

Resolution of two of the small peptides (α1-CB4 and α1-CB5) that did not separate on CM-cellulose was accomplished by molecular sieve chromatography on a 1.8 × 130 cm column (bed volume 320 ml) of Sephadex G-50, fine beads (Pharmacia). After application of the sample in 1–2 ml of water, the peptides were eluted from the column with 0.03 M ammonium propionate (pH 4.5) at 22 ml/hr. Fractions of 3.0 ml were collected and their absorbances were read at 230 mμ. Each of the peptides was rechromatographed using the same system. Effluent fractions were desalted on columns of Sephadex G-25, coarse beads (Pharmacia), using a pyridine acetate buffer (Piez *et al.*, 1963).

Amino Acid Analysis. Samples were hydrolyzed with

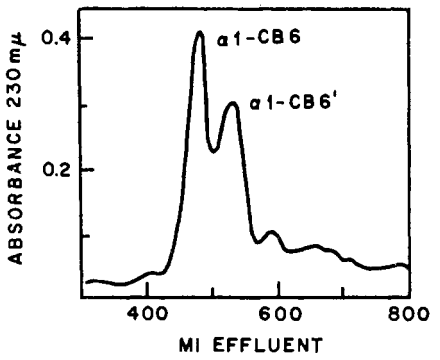

FIGURE 3: Chromatography of peak 6 (see Figure 1) on CM-cellulose at pH 4.8, 40°. Elution was in 0.02 M sodium acetate (pH 4.8) with a concave gradient from 0 to 0.14 M NaCl over a volume of 1740 ml.

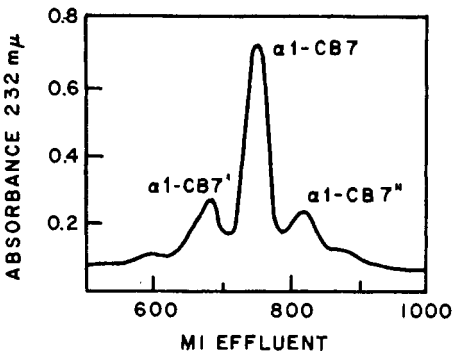

FIGURE 4: Chromatography of peak 7 (Figure 1) on CM-cellulose at pH 4.8, 40°. For further details, see legend to Figure 3.

constant-boiling HCl at 108° for 24 hr in an atmosphere of nitrogen. Analyses were performed on a single-column automatic amino acid analyzer modified for high-speed analysis (Miller and Piez, 1966). Losses of serine, threonine, and tyrosine and the incomplete release of valine were corrected for by previously determined factors (Piez *et al.*, 1960).

Hexose Analysis. Glucose and galactose were measured with the glucostat and galactostat reagents (Worthington) as previously described (Butler and Cunningham, 1966). Prior to the analyses the samples were hydrolyzed with 2 N HCl for 30 min at 108° and then dried.

Peptide Maps. Maps of tryptic hydrolysates of some of the larger peptides were prepared by a modification of the method of Katz *et al.* (1959). Samples (0.8–1.0 mg) were dissolved in a minimal amount of 0.2 M ammonium bicarbonate (pH 8.5) and were digested with 1% by weight of trypsin (Worthington, two-times crystallized). After 6 hr at room temperature an additional 1% by weight of trypsin was added and the digestion was continued overnight. The trypsin hydrolysates were spotted on sheets of Whatman 3MM chromatography paper and subjected to descending chromatography for approximately 6 hr in pyridine–water (8:2). The sheets were dried overnight at room temperature in a hood. High-voltage electrophoresis was then performed in the second dimension at 2000 v for 120 min, using the apparatus and pH 3.7 pyridine acetate buffer described by Katz *et al.* (1959). Peptides were visualized by spraying with 0.2% ninhydrin in acetone and allowing the sheets to stand overnight at room temperature.

Gel Electrophoresis. Acrylamide gel electrophoresis was done by the method of Nagai *et al.* (1964) except that the acrylamide concentration in the running gel was increased to 12.5% and the concentration of *N,N'*-methylenebisacrylamide was reduced to 0.53%.

Molecular Weight Determinations. The molecular weights of the peptides were determined by high-speed sedimentation equilibrium in the Spinco Model E ultracentrifuge as described by Yphantis (1964). Samples of α1-CB4 and α1-CB5 were peak fractions from rechromatography of these peptides on Sephadex G-50 utilizing 0.03 M ammonium propionate (pH 4.5). Buffer from fractions eluted prior to the peptides served as the solvent reference. Centrifugation of these small peptides was at 67,700 rpm and 25° in a titanium rotor employing a cell with a double-sector Kel-F-coated aluminum centerpiece. Sample and reference volumes were 0.2 ml to give approximately a 3-mm liquid column in the cell. Peptides α1-CB6, α1-CB7, and α1-CB8 were peak fractions from chromatography at pH 4.8 on CM-cellulose. They were dialyzed overnight against 100 volumes of 0.15 M potassium acetate buffer (pH 4.8); the buffer outside the dialysis bag was used as the solvent reference. Centrifugation was performed at 35° using several concentrations in a six-compartment Yphantis cell. Column heights were about 1.5 mm. Speeds between 33,450 and 50,740 rpm were used.

Equilibrium was reached in all cases within 400 min. Measurements of Rayleigh interference patterns were made at various times. No significant change in the patterns occurred over a 20-hr period after equilibrium was achieved. The methods for measurement and calculation were those of Yphantis (1964) as previously described (Piez, 1965). The partial specific volumes of the larger peptides (α1-CB3, α1-CB6, α1-CB7, and α1-CB8) were assumed to be 0.700; those for α1-CB4 and α1-CB5 were calculated from their amino acid compositions to be 0.694 and 0.687, respectively.

Results

Chromatography. The chromatography on CM-cellulose at pH 3.6 of the mixture of peptides formed by cleavage of α1 with CNBr is illustrated in Figure 1. A minimum of seven peaks was consistently observed. Samples of CNBr-cleaved β_{11} yielded essentially identical chromatograms confirming earlier indications (Bornstein and Piez, 1966) that β_{11} is derived from α1 by a single cross-link in the region of α1-CB1. β_{11}-CB1, the

TABLE I: Amino Acid Composition of Major and Minor CNBr Peptides.[a]

	α1-CB6	α1-CB6′	α1-CB7	α1-CB7′	α1-CB7″	α1-CB8	α1-CB8′	α1-CB8″
3-Hydroxy-proline	6.3	4.4						
4-Hydroxy-proline	80	79	98	98	93	102	102	97
Aspartic acid	50	49	43	44	45	37	41	42
Threonine	21	21	24	23	23	18	19	18
Serine	57	60	31	33	34	38	45	41
Homoserine[b]			3.4	3.4	4.5	3.6	2.9	4.7
Glutamic acid	63	65	60	65	63	70	78	75
Proline	150	156	143	142	137	118	108	113
Glycine	335	331	342	337	333	328	337	336
Alanine	94	97	105	108	110	134	132	137
Valine	7.1	7.4	29	27	26	18.6	19.3	20.6
Isoleucine	14.5	13.4	10.9	9.7	10.4	4.2	4.0	3.6
Leucine	21	20	15.3	15.2	15.5	15.1	15.9	14.8
Tyrosine	3.6	2.7						
Phenylalanine	5.2	5.5	11.0	11.4	12.7	11.4	12.4	10.6
Hydroxylysine	8.6	9.0	2.5	2.2	3.9	3.7	3.4	3.4
Lysine	16.0	18.1	35	33	36	33	29	30
Histidine	4.6	4.6						
Arginine	55	55	49	47	52	53	49	53

[a] Residues/1000 total residues. Values are averages of three different samples except in the case of α1-CB8″ which is the average of two samples. The absence of a value indicates less than 0.5 residue/1000. [b] Homoserine plus homoserine lactone.

dimer derived from α1-CB1, eluted at the front in the same position as α1-CB1 (peak 1, Figure 1) and did not change the appearance of the chromatogram. Peak 2 contained α1-CB2 which has already been studied in detail (Bornstein and Piez, 1966; Bornstein, 1967). The peak marked 4 + 5 was chromatographed on Sephadex G-50 as illustrated in Figure 2. In this way two small peptides (α1-CB4 and α1-CB5) present in approximately equal amounts could be separated. The material eluting at the void volume (90–100 ml) was shown by amino acid analysis to consist of α1-CB3 and α1-CB6 which overlap α1-CB4 and -5 in the effluent from the CM-cellulose column. The peptides in peaks 3, 6, 7, and 8 were further purified by chromatography of each on CM-cellulose at pH 4.8. By this procedure peak 6 was separated into a major (α1-CB6) and a

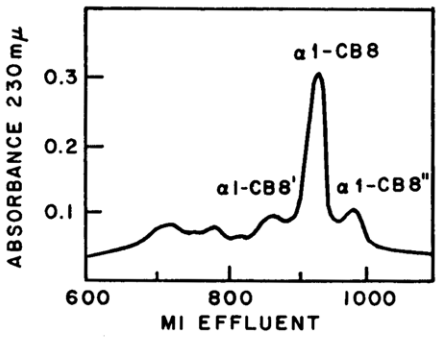

FIGURE 5: Chromatography of peak 8 (Figure 1) on CM-cellulose at pH 4.8, 40°. For further details, see legend to Figure 3.

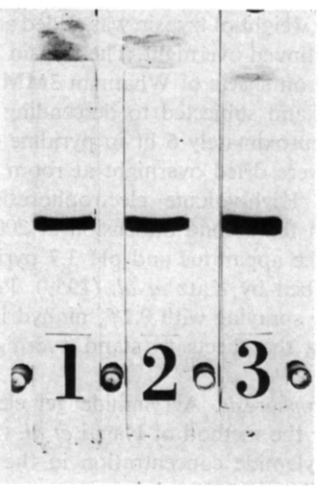

FIGURE 6: Acrylamide gel electrophoresis patterns of α1-CB6 (tube 1), α1-CB6′ (tube 2), and a mixture of α1-CB6 and α1-CB6′ (tube 3). Migration was toward the cathode from top to bottom.

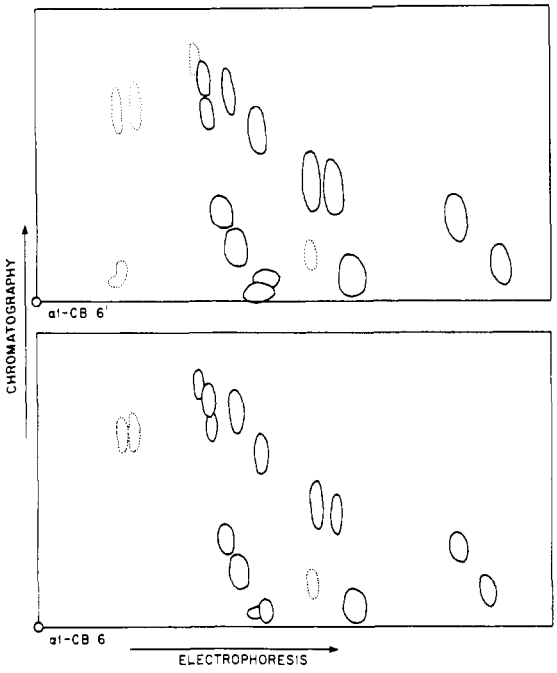

FIGURE 7: Comparison of tryptide maps of α1-CB6 and α1-CB6′. Descending chromatography in the first dimension was in pyridine–water (8:2). Electrophoresis in the second dimension was performed at 2000 v for 2 hr in pyridine acetate (pH 3.5).

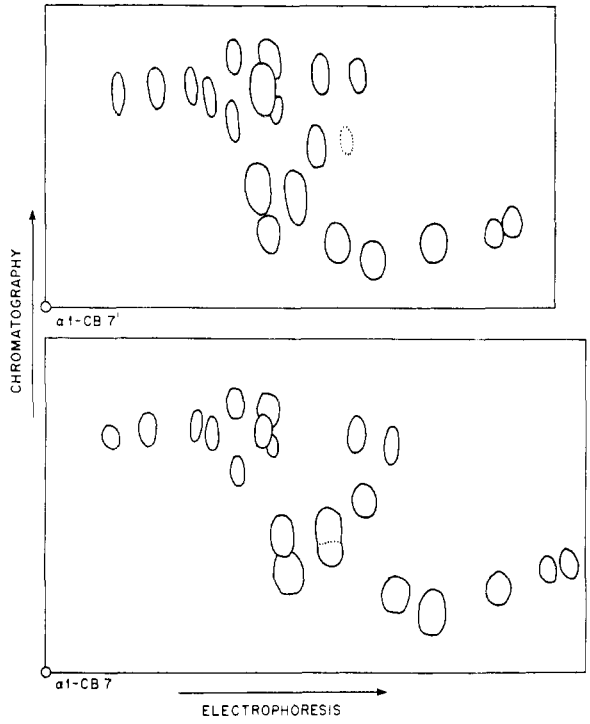

FIGURE 8: Comparison of tryptide maps of α1-CB7 and α1-CB7′. See legend to Figure 7 for further details.

minor (α1-CB6′) component (Figure 3). The material from peak 7 was separated into a major (α1-CB7) and two minor (α1-CB7′ and α1-CB7″) components (Figure 4). Chromatography of the material of peak 8 (Figure 5) also yielded a major (α1-CB8) and two minor components (α1-CB8′ and α1-CB8″). The peaks preceding α1-CB8′ in the effluent were composed of mixtures of peptides containing incompletely cleaved methionyl bonds as evidenced by the presence of methionine in hydrolysates of these fractions; they were not further studied. The material in peak 3 (Figure 1) did not separate into major and minor components but amino acid analysis indicated that contamination was reduced by the rechromatography.

The Major and Minor Chromatographic Components. Table I compares the amino acid compositions of the major and minor components obtained by the chromatography of peaks 6–8 at pH 4.8. No significant differences could be detected between the major and minor components of any set. The small differences in glutamic acid and serine and some other amino acids were not consistent and probably reflect small amounts of impurities. It is apparent from an examination of the amino acid compositions, particularly with regard to those amino acids present in small amounts (*e.g.*, valine, isoleucine, leucine, tyrosine, phenylalanine, hydroxylysine, and histidine), that the major and minor components within each set are either identical or closely related, while there are many differences between the sets.

A comparison of the patterns from gel electrophoresis of α1-CB6 and of α1-CB6′ is shown in Figure 6. No difference could be detected in the mobilities. In a similar series of experiments it was not possible to distinguish between α1-CB7, α1-CB7′, and α1-CB7″ or between α1-CB8, α1-CB8′, and α1-CB8″.

A more critical comparison within each set was obtained by an examination of the tryptides. Figures 7–9 compare diagrammatic representations of peptide maps of the trypsin hydrolysates of some of the major and minor chromatographic components. Weakly staining spots are indicated by dashed lines. There were no obvious differences between the maps for α1-CB6 and α1-CB6′ (Figure 7); those for α1-CB7, α1-CB7′ (Figure 8), and α1-CB7″ (not shown) were also the same; while those for α1-CB8, α1-CB8″ (Figure 9), and α1-CB8′ (not shown) varied only in a single peptide spot (marked with an arrow). This difference could have resulted from the presence of varying amounts of the open and lactone forms of homoserine as discussed below. The number of peptides observed in all cases was consistent with the number predicted from the lysine and arginine contents. On the basis of these observations, it was concluded that each minor component was not fundamentally different from the corresponding major component and therefore that each set represents a single sequence within α1.

α1-CB7′ and α1-CB8′ probably contained the open form of homoserine while the corresponding major components contained homoserine lactone. The charge difference could allow separation of the two forms at

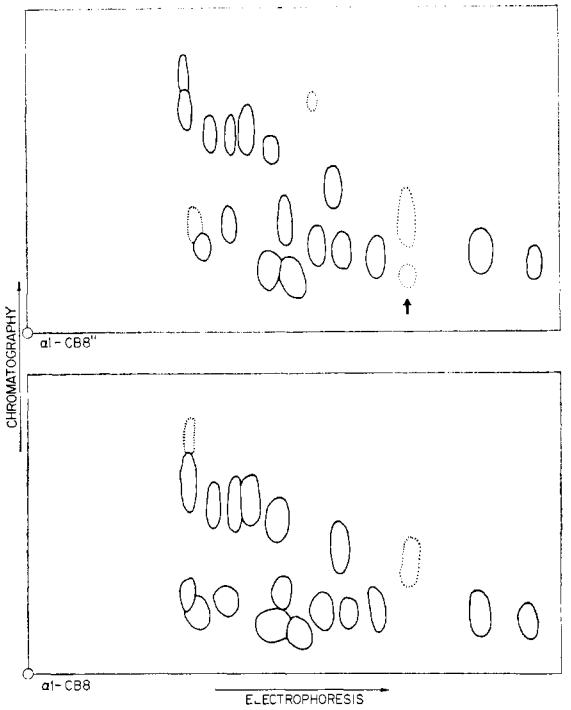

FIGURE 9: Comparison of tryptide maps of α1-CB8 and α1-CB8''. See legend to Figure 7 for further details.

pH 4.8. In support of this explanation, it was found that α1-CB7 could be converted to material that chromatographed in a position identical with α1-CB7' by adjusting the pH of a solution of α1-CB7 to 11, allowing it to stand at room temperature for 5 min, and then readjusting the pH to 4.8 prior to chromatography. These are conditions which open the ring of homoserine lactone. Furthermore, no acidic component analogous to α1-CB7' and α1-CB8' was found by chromatography of peak 6 at pH 4.8. This was to be expected since α1-CB6 contains no homoserine (Table I). No reason was found to explain the presence of α1-CB6', α1-CB7'', and α1-CB8''.

Since the α1 chain might contain repeating sequences and since peptides might arise from nonspecific degradation, it was important to determine the relative yields of the peptides. This estimate was made by measuring the areas under the peaks on several chromatograms like the one seen in Figure 1. Assuming similar extinction coefficients, these areas should have a constant relation to the molecular weights (see below) if the peptides were present in equimolar amounts. This comparison is made in Table II and indicates that approximately 1 equiv of each of the indicated peptides was obtained. A similar conclusion can be reached for α1-CB1, α1-CB2, and α1-CB3 from effluent patterns obtained by chromatography on phosphocellulose (Bornstein *et al.*, 1966). These calculations are only approximate, particularly for the small peptides which may not have the same extinction coefficient; however, to answer the questions concerning repeating sequences and nonspecific degradation it is only necessary to distinguish between a trace, 1 equiv, and 2 or more than 2 equiv.

TABLE II: Relative Recoveries of CNBr Peptides from α1.

α1-CB Peptide	Area[a]	Area/Mol Wt[b]
3	214	0.0157 (0.81)
4 + 5	121	0.0147 (0.75)
6	399	0.0221 (1.13)
7	479	0.0195 (0.10)
8	593	0.0248 (1.27)

[a] Arbitrary units; average of four chromatograms. [b] See Table IV; molecular weights from amino acid analysis. Values normalized to 1.00 for α1-CB7 are given in parentheses.

Amino Acid Composition. The compositions of the CNBr peptides derived from α1 are given in Table III. The data are presented as residues per peptide using those amino acids present in small amounts to calculate a minimal molecular weight and assuming one residue of homoserine per peptide. The compositions of α1-CB1 and α1-CB2 are included for completeness. All of the peptides except α1-CB1 (Bornstein *et al.*, 1966a; Kang *et al.*, 1967) contain one-third glycine. Each peptide has characteristics that distinguish it from the others and from whole α1 chains. For example, α1-CB2 has a relatively high concentration of imino acids. α1-CB3 is the only peptide with a molecular weight near 13,000 and it is rich in glutamic acid. α1-CB4 and α1-CB5 are relatively small like α1-CB2 but are distinctly different in composition; α1-CB5 has one residue of hydroxylysine, one of the two residues of histidine in α1, and approximately one residue each of glucose and galactose (see below). α1-CB6 contains no homoserine and must therefore be derived from the COOH-terminal end of α1; it contains the single residue of 3-hydroxyproline and one of the two tyrosine residues in α1. α1-CB7 and α1-CB8 are both relatively large but the former contains more proline, valine, and isoleucine and less glutamic acid and alanine than the latter.

The number of residues of each amino acid accounted for by the eight CNBr peptides is compared in Table III with an analysis of whole α1 done as part of this study. The differences are within experimental error suggesting that the peptides account for all the amino acid residues in α1. The most important comparison is the methionine content since the number of CNBr peptides can be predicted to be one more than the number of methionine residues. To provide a precise value of the methionine content a number of samples of α1 were hydrolyzed for varying periods of time (24–72 hr) and analyzed. It was found that if the amount of methionine sulfoxide in the sample formed from methionine was small, the total of methionine

TABLE III: Amino Acid Compositions of the CNBr Peptides from $\alpha 1$.[a]

	α1-CB1	α1-CB2	α1-CB3	α1-CB4	α1-CB5	α1-CB6	α1-CB7	α1-CB8	Total CNBr Peptides	α1[b] Found
3-Hydroxyproline	0	0	0	0	0	1 (1.2)	0	0	1	0.8
4-Hydroxyproline	0	5	14	6 (6.0)	3 (2.7)	16	27	27	98	104
Aspartic acid	1	0	7 (7.0)	3 (3.0)	3 (3.0)	10	12	10	46	47
Threonine	0	0	2 (2.0)	1 (1.0)	1 (1.0)	4 (4.2)	7 (6.7)	5 (4.7)	20	20.6
Serine	2	2	3 (3.1)	0	2 (2.0)	12	9 (8.7)	10	40	40
Glutamic acid	1	4	16	3 (3.0)	3 (3.0)	13	16	19	75	76
Proline	2	7	14	6 (5.8)	2 (1.9)	31	39	31	132	126
Glycine	3	12	51	15	12	66	92	87	338	356
Alanine	1	2	20	3 (3.0)	3 (3.1)	19	29	36	113	118
Valine	2	0	4 (4.2)	0	0	2 (1.5)	8 (7.8)	5 (4.9)	21	21.4
Isoleucine	0	0	0	0	0	3 (2.9)	3 (3.0)	1 (1.1)	7	8.0
Leucine	0	1	3 (3.0)	2 (2.0)	1 (1.0)	4 (4.1)	4 (4.2)	4 (4.0)	19	20.5
Tyrosine	1	0	0	0	0	1 (0.6)	0	0	2	1.9
Phenylalanine	0	1	3 (2.9)	0	1 (0.9)	1 (1.1)	3 (3.1)	3 (3.0)	12	12.2
Hydroxylysine	0	0	0 (0.2)	0	1 (1.1)	2 (1.8)	1 (0.7)	1 (1.0)	5	4.3
Lysine	1	0	5 (4.7)	2 (2.0)	2 (1.8)	3 (3.3)	9 (9.0)	9 (8.9)	31	30
Histidine	0	0	0	0	1 (1.0)	1 (0.9)	0	0	2	2.2
Arginine	0	1	6 (6.1)	4 (4.1)	1 (1.0)	11	13	14	50	52
Homoserine	1	1	1 (0.9)	1 (1.0)	1 (0.9)	0	1 (0.9)	1 (1.0)	7	6.6[c]
Total	15	36	149	46	37	200	273	263	1019	1048

[a] Residues per peptide. Values are rounded off to the nearest whole number. Where less than ten residues were found actual values are shown in parentheses except for α1-CB1 and α1-CB2 which have been described previously (Bornstein and Piez, 1966; Bornstein, 1967). [b] Average of three separate analyses of α1. Values are computed for a molecular weight of 95,000 with an average residue weight of 90.6. [c] Methionine.

TABLE IV: Molecular Weights of the CNBr Peptides of $\alpha 1$ as Measured by Amino Acid Analysis and by High-Speed Sedimentation Equilibrium.

Peptide	Amino Acid Anal.	Mol Wt		
		Whole Cell	at $c = 0$	at $r = b$ (M_z)
$\alpha 1$-CB1[a]	1,476	(1,500)		
$\alpha 1$-CB2[a]	3,311	3,200		
$\alpha 1$-CB3	13,627	12,740	11,980	13,920
$\alpha 1$-CB4	4,404	3,770	3,700	3,790
$\alpha 1$-CB5	3,850[b]	4,170	4,180	4,050
$\alpha 1$-CB6	18,048	17,700	17,000	18,100
$\alpha 1$-CB7	24,581	22,260	21,240	23,080
$\alpha 1$-CB8	23,911	25,100	25,600	24,500
Total	93,208	90,440		

[a] From Bornstein and Piez (1966). [b] Amino acids plus one residue each of galactose and glucose.

plus its sulfoxide was constant within experimental error between 24- and 72-hr hydrolysis. If the amount of oxidation was appreciable, as sometimes occurred, the methionine content was low and variable. Using only data where there was less than 10% oxidation, the methionine content was found to be 6.7 ± 0.3 residues (average of seven analysis plus and minus standard deviation) calculated for an $\alpha 1$ chain of 95,000 mol wt. On the basis of a whole number value of seven for methionine content, the eight CNBr peptides isolated provide an exact agreement with the number expected.

Hexose Content. The hexose contents of the CNBr peptides were measured with glucose and galactose oxidases. Hexoses other than glucose and galactose have not been found in soluble collagens (Blumenfeld *et al.*, 1963; Butler and Cunningham, 1966). Equal quantities of glucose and galactose (0.7 mole of each/mole of peptide) were found attached to $\alpha 1$-CB5. Small amounts of hexose (less than 0.2 mole/equiv of peptide) were found in some of the other peptides. Since these latter values are near the lower limit of sensitivity of the method, additional studies will be required to determine their significance.

Molecular Weights. The heterogeneity and molecular weights of the peptides were estimated by high-speed sedimentation equilibrium, except for $\alpha 1$-CB1 which contains only 15 amino acid residues. Good agreement was observed for weight-average molecular weights calculated for the whole cell and extrapolated to zero concentration and z-average molecular weights (Table IV) indicating that the samples were homogeneous. These values compare favorably with the molecular weights calculated from the amino acid compositions in almost every case. The values for $\alpha 1$-CB4 differ by about 15% for an unknown reason. The sum of the molecular weights of the peptides, 90,400 by sedimentation equilibrium and 93,200 by amino acid analysis, agrees well with the reported value of about 95,000 obtained for whole $\alpha 1$ chains (Lewis and Piez, 1964; Piez, 1965; Kang *et al.*, 1966).

Search for Additional Peptides. Although the eight peptides described above account for all the weight and all the amino acids in $\alpha 1$ within experimental error, the data cannot completely exclude the possibility that a peptide was missed. If this were the case, the peptide would have to be small, perhaps a few amino acids, or possibly somewhat larger (molecular weight up to about 3000) if its presence was obscured in the chromatograms by one of the larger peptides. Experiments were undertaken to find such a peptide or to exclude its existence. A CNBr digest of $\alpha 1$ was lyophilized and redissolved, retaining all of the products, and placed directly on a Sephadex G-50 column equilibrated and eluted with 0.2 M acetic acid. Fractions were collected throughout the included volume of the column. Small groups of fractions covering the entire low molecular weight region of the effluent were hydrolyzed and analyzed. A small amount of glutamic acid and serine and traces of other amino acids were found, but the amounts were insufficient to suggest the presence of a previously undiscovered small peptide. Since an equivalent of homoserine would be present in the proposed peptide regardless of its size, the presence of this amino acid in these fractions in only trace amounts gives further evidence of the absence of an additional peptide. A peptide of intermediate size that might chromatograph with one of the large peptides was sought by combining appropriate fractions from the effluent of the CM-cellulose column operated at pH 3.6, concentrating them by lyophilization, and subjecting them to gel filtration on Sephadex G-50. No significant amounts of intermediate-sized peptides were found.

Discussion

After cleavage with CNBr we have been able to

isolate eight unique peptides of varying sizes and properties that account for all of the α1 chain of rat skin collagen on the basis of amino acid composition and molecular weight. Utilization of this technique should provide a suitable initial step in studies of the sequence of the α1 chain, since the molecular weights of the starting materials for such a study are in a reasonable range. The next information essential to this problem is the order of these eight CNBr peptides in the α1 chain. Previous studies (Bornstein *et al.*, 1966a,b) have shown that α1-CB1 and α1-CB2 are peptides 1 and 2, respectively, beginning at the NH_2-terminal end. Since α1-CB6 contains no homoserine, it must be peptide 8 (COOH terminal). Data provided by a previous study (Kang *et al.*, 1966) utilizing tadpole collagenase to cleave native collagen indicate that α1-CB7 is probably peptide 7. The argument is as follows. The tadpole collagenase yields a peptide ($\alpha 1^B$) of about 24,000 mol wt from the COOH-terminal end of α1. $\alpha 1^B$ contains all of the amino acid residues present in the slightly smaller (18,000 mol wt) COOH-terminal CNBr peptide (α1-CB6) including, for example, one residue each of tyrosine and 3-hydroxyproline. Among the amino acids remaining in the segment of about 6000 mol wt not included in α1-CB6 ($\alpha 1^B$ minus α1-CB6) are two residues of isoleucine. Of the remaining CNBr peptides only α1-CB7 has more than one isoleucine. $\alpha 1^B$ must then overlap α1-CB7 indicating that the latter is peptide 7. Experiments are in progress to order the remaining four CNBr peptides.

The finding of only eight CNBr peptides provides strong evidence that the two α1 chains of rat skin collagen are identical. If differences exist in sequence they must be few in number. Previous experiments (Bornstein and Piez, 1965) indicated that there was a sufficient number of CNBr peptides to suggest that the two α1 chains are different. These observations can now be explained by the presence of minor chromatographic components, each of which is probably derived from the same sequence as one of the eight major peptides.

Suggestions have been made that the α chains of collagen are composed of segments (intrachain subunits), perhaps five or six per chain, joined end to end by nonpeptide bonds (Petrushka and Hodge, 1964; Gallop, 1966). The results presented here are clearly incompatible with a model containing identical subunits and appear to be difficult to reconcile with any subunit model in which the subunits consist of only a few kinds. One would predict from these models that more than one equivalent of at least some of the CNBr peptides would have been found. The results presented here are more easily explained by an α1 chain that has an amino acid sequence that is unique throughout its length. The data, of course, do not exclude repeating triplets or longer repeating sequences that may occur by chance since glycine is commonly in every third position and there are high proportions of some of the other amino acids.

The differences that account for the chromatographic separation of the minor components from the corresponding major components are not known. However, there is precedent for heterogeneity of α1 in at least two ways. There is incomplete conversion of a lysyl residue to the δ-semialdehyde of α-aminoadipic acid preliminary to cross-linking (Piez *et al.*, 1967) and there is incomplete hydroxylation of certain proline residues (Bornstein, 1967). Heterogeneity may also arise from differing amounts or kinds of attached carbohydrate. Still another kind of heterogeneity may occur that will explain the minor CNBr peptides observed here.

Acknowledgment

We are indebted to Mrs. Gloria Pakis for expert technical assistance.

References

Blumenfeld, O. O., Paz, M. A., and Gallop, P. M. (1963), *J. Biol. Chem. 238*, 3835.
Bornstein, P. (1967), *J. Biol. Chem. 242*, 2572.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966a), *Proc. Natl. Acad. Sci. U. S. 55*, 417.
Bornstein, P., Kang, A. H., and Piez, K. A. (1966b), *Biochemistry 5*, 3803.
Bornstein, P., and Piez, K. A. (1964), *J. Clin. Invest. 43*, 1813.
Bornstein, P., and Piez, K. A. (1965), *Science 148*, 1353.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Butler, W. T., and Cunningham, L. W. (1966), *J. Biol. Chem. 241*, 3882.
Butler, W. T., and Piez, K. A. (1967), *Federation Proc. 26*, 670.
Francois, C. J., and Glimcher, M. J. (1966a), *Biochem. J. 102*, 148.
Francois, C. J., and Glimcher, M. J. (1966b), *Biochim. Biophys. Acta 133*, 91.
Gallop, P. M. (1966), *Nature 209*, 73.
Gallop, P. M., and Seifter, S. (1966), *Proteins 4*, 153.
Hannig, K., and Nordwig, A. (1967), *in* Treatise on Collagen, Vol. I, Chemistry of Collagen, Ramachandran, G. N., Ed., London, Academic (in press).
Heidrich, H., and Wynston, K. L. (1965), *Z. Physiol. Chem. 342*, 166.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.
Kang, A. H., Nagai, Y., Piez, K. A., and Gross, J. (1966), *Biochemistry 5*, 509.
Kao, K.-Y.T., Hitt, W. E., Dawson, R. L., and McGavack, T. H. (1967), *Arch. Biochem. Biophys. 120*, 518.
Katz, A. M., Dreyer, W. J., and Anfinsen, C. B. (1959), *J. Biol. Chem. 234*, 2897.
Lewis, M. S., and Piez, K. A. (1964), *Biochemistry 3*, 1126.
Miller, E. J., Martin, G. R., Piez, K. A., and Powers, M. J. (1967), *J. Biol. Chem.* (in press).
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.

Nagai, Y., Gross, J., and Piez, K. A. (1964), *Ann. N. Y. Acad. Sci. 121*, 494.
Petruska, J. A., and Hodge, A. J. (1964), *Proc. Natl. Acad. Sci. U. S. 51*, 871.
Piez, K. A. (1964), *J. Biol. Chem. 239*, C4315.
Piez, K. A. (1965), *Biochemistry 4*, 2590.
Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.
Piez, K. A., Lewis, M. S., Martin, G. R., and Gross, J. (1961), *Biochim. Biophys. Acta 53*, 596.
Piez, K. A., Martin, G. R., Kang, A. H., and Bornstein, P. (1966), *Biochemistry 5*, 3813.
Piez, K. A., Weiss, E., and Lewis, M. S. (1960), *J. Biol. Chem. 235*, 1987.
Schleyer, M. (1962), *Z. Physiol. Chem. 329*, 97.
Yphantis, D. A. (1964), *Biochemistry 3*, 297.

Reductive Cleavage of Acylproline Peptide Bonds*

W. F. Benisek,† M. A. Raftery,‡ and R. David Cole

ABSTRACT: To determine the utility of selective, reductive cleavage of proline-containing peptides as a degradative method in structural determinations of polypeptides and proteins, the cleavage of acylproline peptide bonds induced by sodium metal–liquid ammonia was studied in peptides and peptide derivatives. Attention was given to finding reaction conditions sufficient to obtain extensive cleavage, to identifying the nature of the new C-terminal groups produced by the reaction, and to determining the nature and degree of side reactions. The maximal extents of cleavage obtained ranged from 56 to 90%. Variation in the lability of various acylproline peptide bonds was suggested by a comparison of the conditions required for the cleavage of different peptides. Thus, glycylproline and the threonylproline peptide bond of insulin B chain were much more readily reduced than *N*-acetylproline, *N*-acetylglycylproline, or the acylproline peptide bonds in performic acid oxidized apoferredoxin. The extents of reduction of acetylproline in the presence and absence of methanol and sodium amide (a strong base) were essentially the same, indicating that the inclusion of proton donors in the reaction mixture is not necessary for cleavage. The reduction of *N*-acetylglycylproline resulted in the conversion of the glycine residue to both aminoacetaldehyde and ethanolamine residues, but these forms accounted for only about two-thirds of the reduced glycine residues. Using amino acid analysis and amino-terminal analysis of sodium–ammonia–reduced ferredoxin and insulin B-chain derivatives, nonspecific cleavage of internal peptide bonds was found to range from 0 to 4% per peptide bond. The cleavage of amino-terminal peptide bonds was more extensive than that of internal, nonproline peptide bonds. The recoveries of amino acids after acid hydrolysis of reduced polypeptides indicated that side reactions in general were minimal for amino acid residues which did not possess a free α-amino group and which were not amino terminal to prolyl residues.

Recent reports have appeared describing chemical methods for the reductive cleavage of polypeptides at acylproline peptide bonds. Various reducing agents have been employed in this reaction including lithium aluminum hydride (Ruttenberg *et al.*, 1964), lithium in methylamine (Patchornik *et al.*, 1964), and sodium in liquid ammonia (Benisek and Cole, 1965; Wilchek *et al.*, 1965; Ressler and Kashelikar, 1966). All of these techniques result in cleavage of acylproline peptide bonds, to various extents.

In this paper we present the results of a study of the reductive cleavage of proline-containing peptides using sodium metal in liquid ammonia as the reducing agent. This study was undertaken to obtain information which could aid in the design of structural studies of polypeptides. Therefore we have examined the reaction with regard to parameters which affect the extent of cleavage of various acylproline peptide bonds as well as the nature and degree of side reactions, including nonselective peptide-bond cleavage. Some experiments designed to determine the nature of the new C-terminal residues produced by the reductive cleavage are also described.

Materials and Methods

Glycyl-L-proline and acetyl-DL-proline were obtained from Mann Research Laboratories, Inc. Car-

* From the Department of Biochemistry, University of California, Berkeley, California 94720. *Received July 26, 1967.* Supported by U. S. Public Health Service Grants TI GM31, AM 02691, and AM 06482 and by the Agricultural Research Station.

† Present address: Department of Molecular Biophysics, Yale University, New Haven, Conn.

‡ Present address: Department of Chemistry, California Institute of Technology, Pasadena, Calif.