Lai, T. F., & Marsh, R. E. (1972) *Acta Crystallogr., Sect. B B28*, 1982–1989.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Madsen, N. B., Kasvinsky, P. J., & Fletterick, R. J. (1978) *J. Biol. Chem. 253*, 9097–9101.
Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.
Pullman, B., Saenger, W., Sasisekharan, V., Sundaralingam, M., & Wilson, H. R. (1973) *Jerusalem Symp. Quantum Chem. Biochem. 5*, 815–820.
Ross, P. D., & Subramanian, S. (1981) *Biochemistry 20*, 3096–3102.
Rossmann, M. G., Liljas, A., Branden, C. I., & Banaszak, L. J. (1975) *Enzymes, 3rd Ed. 2*, 61–102.
Saran, A., Perahia, D., & Pullamn, B. (1973) *Theor. Chim. Acta 30*, 31–44.
Sarma, R. H. (1980) in *Nucleic Acid Geometry and Dynamics* (Sarma, R. H., Ed.) p 145, Pergamon Press, Oxford.
Segel, I. H. (1975) *Enzyme Kinetics*, pp 481–492, Wiley-Interscience, New York.
Sprang, S., & Fletterick, R. J. (1979) *J. Mol. Biol. 131*, 523–551.
Sturtevant, J. (1964) in *Rapid Mixing and Sampling Techniques in Biochemistry*, pp 89–103, Academic Press, New York.
Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236–2240.
Sturtevant, J., & Lyons, M. (1969) *J. Chem. Thermodyn. 1*, 201–212.
Sundaralingam, M. (1973) *Jerusalem Symp. Quantum Chem. Biochem. 5*, 417–456.
Sundaralingam, M. (1975) *Ann. N.Y. Acad. Sci. 255*, 3–42.
Sutor, J. (1959) *Acta Crystallogr. 11*, 453.
Tanford, C. (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd ed., pp 18–21, Wiley, New York.
Thewalt, U., Bugg, C. E., & Marsh, R. E. (1970) *Acta Crystallogr., Sect. B B26*, 1089–1101.
Withers, S. G., Sykes, B. D., Madsen, N. B., & Kasvinsky, P. J. (1979) *Biochemistry 18*, 5342–5348.

Amino Acid Sequence of Chick Skin Collagen α1(I)-CB8 and the Complete Primary Structure of the Helical Portion of the Chick Skin Collagen α1(I) Chain†

John H. Highberger, Clare Corbett, S. N. Dixit, Wing Yu,‡ Jerome M. Seyer, Andrew H. Kang, and Jerome Gross*

ABSTRACT: The primary structure of chick skin collagen α1-CB8, the 279-residue CNBr peptide from the helical portion of the α1(I) chain, has been determined by automated amino acid sequence analysis of tryptic peptides of the maleylated and of the cyclohexanedione-treated material, of thermolytic peptides, and of a single 40-residue chymotryptic fragment. The sequence thus obtained showed 95% identity with that of the corresponding peptide from rat collagen. Completion of this work permits the assembly of the complete helical amino acid sequence of the chick skin α1(I) chain.

At the present time at least five genetically distinct types of collagen are recognized, although there are indications that additional forms may exist. Type I is by far the most common, making up the framework of most of the tissues of the animal body, but is especially characteristic of skin, tendon, and bone. The molecule consists of three α chains, each containing over 1000 amino acid residues in peptide linkage with no sequence repetition beyond 6 residues except for one short 9-residue stretch near each end (Hulmes et al., 1973). Two of these, the α1(I) chains, are identical, and one, the α2 chain, is different in amino acid sequence. The molecule is thus usually represented as $[\alpha 1(I)]_2\alpha 2$. Type II, or $[\alpha 1(II)]_3$, is the predominant protein of cartilage. Type III, $[\alpha 1(III)]_3$, occurs primarily in cardiovascular tissue, although it is also found in most loose connective tissues. The chain structure of type IV collagen found in basement membrane of several different tissues is still controversial, as is that of the so-called type V or A–B collagen.

Primary structural analyses of type I collagen available in the literature are still incomplete in that the published sequence of the complete α1(I) chain is a hybrid consisting of part rat and part calf skin (Hulmes et al., 1973; Bornstein & Traub, 1979). Portions of the sequence of skin collagen from human and guinea pig sources are also available (Fietzek et al., 1974; Clark & Bornstein, 1972). The amino acid sequence analysis of the α2 chain of calf skin is also complete, but as yet unpublished (P. P. Fietzek, personal communication). The complete amino acid sequence of calf skin type III (Fietzek et al., 1979; Dewes et al., 1979a,b; Bentz et al., 1979; Lang et al., 1979; Allmann et al., 1979) and human liver type III (Seyer & Kang, 1977, 1978, 1981; Seyer et al., 1980) has recently appeared in the literature, and a large portion of type II collagen is also published (Butler & Ponds, 1971; Butler et al., 1976, 1977).

With this publication of the cyanogen bromide peptide, α1-CB8, the amino acid sequence determination of the α1(I) chain of chick skin collagen, including the nonhelical amino-

† From the Developmental Biology Laboratory, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114 (J.H.H., C.C., W.Y., and J.G.), and the Veteran's Administration Medical Center and Departments of Biochemistry and Medicine, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104 (S.N.D., J.M.S., and A.H.K.). *Received June 17, 1981; revised manuscript received November 24, 1981.* This is Publication No. 895 of the Robert W. Lovett Memorial Group for the Study of Diseases Causing Deformities. This work was supported by U.S. Public Health Service Grants AM 3564 and AM 16506 and Training Grant AM 07258 from the National Institutes of Health, The Janie Fund, The Gebbie Foundation, and the Medical Research Service of the Veterans Administration.

‡ Present address: Thermo Electron Corp., Woburn, MA.

0006-2960/82/0421-2048$01.25/0

Table I: Amino Acid Composition of Chick Skin Collagen α1-CB8, Some of Its Tryptic Peptides after Maleylation, and Chymotryptic Peptide C1[a]

amino acid	α1-CB8	C1	TM1	TM2	TM2T1	TM2T2	TM3	TM4	TM5	TM6
4-hydroxyproline	30 (30)	5.3 (5)	3.8 (4)	6.4 (7)	3.7 (4)	2.4 (3)	4.9 (5)	1.8 (2)	1.9 (2)	8.0 (8)
aspartic acid	10 (10)	2.1 (2)	2.0 (2)	1.9 (2)	1.9 (2)	0.3	2.7 (3)	0.2	1.0 (1)	2.0 (2)
threonine	5.7[b] (6)	1.9 (2)	1.9 (2)	1.0 (1)		0.6 (1)	0.9 (1)			1.8 (2)
serine	11 (11)	1.9 (2)	0.4				2.7 (3)	0.1	0.1	3.1 (4)
glutamic acid	22 (22)	2.2 (2)	2.3 (2)	2.2 (2)	2.0 (2)	0.4	5.8 (6)	1.7 (2)	1.2 (1)	2.9 (3)
proline	30 (30)	5.2 (5)	4.8 (5)	3.3 (3)	3.1 (3)	0.3	6.6 (7)	2.0 (2)	2.1 (2)	7 (7)
glycine	93 (93)	13 (13)	13 (13)	15 (15)	9.3 (9)	5.6 (6)	18 (18)	5.2 (5)	6.0 (6)	21 (21)
alanine	41 (41)	3.5 (4)	7.9 (8)	11 (11)	6.4 (6)	4.5 (5)	5.1 (5)	2.3 (2)	2.2 (2)	6.6 (7)
valine	2.9[c] (3)	0.9 (1)					1.1 (1)			1.1 (1)
isoleucine	1.9 (2)			0.9 (1)		0.7 (1)		0.1	0.9 (1)	
leucine	4.4 (4)		0.2					0.7 (1)		1.7 (2)
phenylalanine	2.9 (3)		0.8 (1)	1.0 (1)		1.2 (1)		0.2	0.9 (1)	
hydroxylysine	1.2 (1)	0.4			0.4					0.4
lysine	8.6 (9)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	0.2	3.8 (4)	0.2	0.8 (1)	2.8 (3)
arginine	15 (15)	2.3 (2)	1.1 (1)	1.0 (1)		0.7 (1)	1.0 (1)	0.9 (1)	1.1 (1)	2.9 (3)
homoserine	0.8[d] (1)	0.9[d] (1)								
total	281	40	39	45	27	18	54	15	18	63

[a] Results are expressed as residues per peptide. The number is rounded off to the nearest whole number where more than 10 residues occur. The numbers in parentheses indicate the residues calculated from the elucidated sequence. [b] Obtained by extrapolating results of timed hydrolyses to zero time. [c] Obtained from a 48-h hydrolysate. [d] Includes homoserine lactone.

terminal telopeptide and the total helical region, is now complete.

Materials and Methods

Preparation of α1-CB8. Chick skin collagen was isolated from 3-week-old White Leghorn lathyritic chicks by neutral salt and acid extraction, and the CNBr peptides of the collagen were prepared and purified as previously described (Highberger et al., 1975). The amino acid composition of chick skin collagen α1-CB8 as used in the present work is given in Table I.

Sequencing Strategy. Restricted tryptic cleavage of the whole α1-CB8 was used as the basic subfragmentation method. The action of trypsin was confined to the arginyl residues by maleylating the substrate and to the lysyl residues by the action of cyclohexanedione. Sequence data of peptides isolated from these digests in combination with those from several analyses on intact α1-CB8, and several thermolytic peptides plus one 40-residue COOH-terminal peptide from a chymotryptic digestion of α1-CB8, provided sufficient data for reconstruction of the original sequence.

Maleylation of α1-CB8. α1-CB8 was maleylated by the method of Butler et al. (1969) with minor modification. After tryptic digestion, the maleylated material was dissolved in 0.1 M pyridine–acetate, pH 3.1, and heated in a closed vial for 6 h at 60 °C to remove the maleyl groups. The material was then frozen and lyophilized.

Cyclohexanedione Treatment of α1-CB8. The arginyl groups of α1-CB8 and a thermolytic peptide, Th1 (residues 1–103), were blocked with cyclohexanedione as described by Patthy & Smith (1975). Excess unreacted reagent was removed by extensive dialysis against large volumes of 0.1 M triethylamine–acetate, pH 9. The bag contents were then lyophilized.

Enzymatic Hydrolysis. Enzymatic digestions with trypsin (Worthington; 3× crystallized, TPCK treated), chymotrypsin (Worthington), and thermolysin (3× crystallized and lyophilized, grade A; Calbiochem) were performed at 37 °C for 2 h, 40 min, and 30 min, respectively. The peptide (4 mg/mL) was dissolved in 0.2 M ammonium bicarbonate/1 mM $CaCl_2$, pH 7.9, and was incubated with the appropriate enzyme. A 1:100 molar ratio of enzymes to substrate was used. At the end of the incubations, the digestion mixture was acidified to pH 3 and lyophilized.

Chromatographic Peptide Separation. Tryptic, thermolytic, and chymotryptic digestion mixtures were initially fractionated on a large (2.5 × 110 cm) column of Sephadex G-50S as previously described (Highberger et al., 1975). The column was calibrated for molecular weight estimation by using known collagenous peptides and tritiated water.

Where analyses indicated further purification to be necessary, this was first attempted by additional molecular sieve chromatography in order to minimize sample loss, using either Sephadex G-50S or one of the Bio-Gel P resins of suitable grade for the particular peptide. Where ion-exchange chromatography was deemed necessary, this was carried out on phosphocellulose (Whatman P11 cellulose phosphate) columns of appropriate size by using the method previously described (Highberger et al., 1975).

Amino Acid Analyses. Samples for amino acid analysis were hydrolyzed for 24 h in tubes sealed under nitrogen, in glass-distilled, constant boiling HCl at 110 °C. The analyses were run on a Beckman 121 amino acid analyzer with an Autolab System AA integrator, using the buffer system of Trelstad & Lawley (1976). Corrections were not usually made for destruction of amino acids during hydrolysis, but in certain cases the hydrolysis time was extended to 72 h for the determination of valine. Serine and threonine were determined by extrapolating the results of various timed hydrolysates to zero time.

In the case of peptides obtained by the tryptic hydrolysis of α1-CB8 treated with cyclohexanedione (CHD-peptides), Patthy & Smith (1975) have shown that the arginine of such peptides is largely destroyed in the acid hydrolysis preceding analysis. They suggested addition to the hydrolyzing acid of an excess of thioglycolic acid to prevent the action. This expedient, however, was not applicable in the present case, because thioglycolic acid eluted from the analyzer column in a position which interfered with the hydroxyproline peak. Therefore, the amino acid analyses for arginine for these peptides were used only to indicate the presence or absence of arginine, and the number of residues was determined from the sequencing results, or in some cases, from overlapping peptides.

Sequencing Methods. Sequencing was carried out in an updated Beckman 890B sequencer by using the methods previously described (Highberger et al., 1975). The fast protein–Quadrol program of Beckman Instruments was used

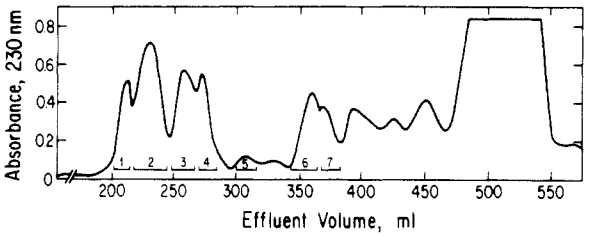

FIGURE 1: Chromatography of the tryptic digest of the maleylation product of 50 mg of α1-CB8 on a 2.5 × 110 cm column of Sephadex G-50S. The column was eluted with 0.03 M sodium acetate/4 M urea, pH 4.8, at 42 °C at a flow rate of 20 mL/h. Numbered horizontal brackets show how fractions were pooled.

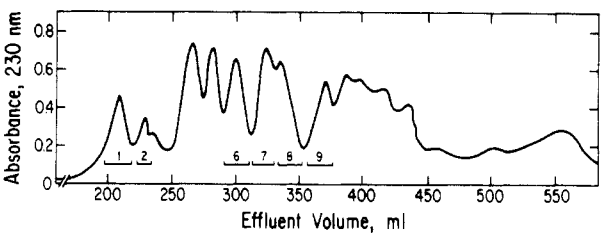

FIGURE 2: Chromatography of the tryptic digest of the product of the treatment of 50 mg of α1-CB8 with cyclohexanedione, on the same column and with the same eluant used in Figure 1. Temperature 44 °C; flow rate 22 mL/h.

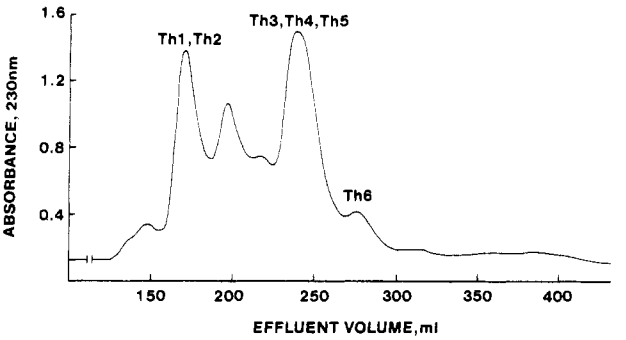

FIGURE 3: Chromatography of the thermolytic peptides on a G-50S column (2.5 × 110 cm). The column was eluted with 0.04 M acetate, pH 4.8, at room temperature at a flow of 16 mL/h.

for some of the peptides whereas the slow peptide–DMAA program was employed for other peptides. Peptide modifications were made with 4-sulfophenyl isothiocyanate (S-PITC) (Braunitzer et al., 1970) and/or with 2-amino-1,5-naphthalenedisulfonic acid (ANS) as described earlier (Foster et al., 1973) and used by us successfully (Dixit et al., 1975a,b). Identification of PTH-amino acids was done by gas chromatography as previously described (Highberger et al., 1975) or by amino acid analysis following back hydrolysis with hydroiodic acid (Smithies et al., 1971). In the later stages of the work, PTH-amino acids were identified by high-pressure liquid chromatography by using an Altex Model 420 system with Altex Model 110 high pressure pumps. The reverse-phase column was 4.6 mm i.d. × 25.0 cm, packed with Altex Li-Chrosorb C_{18}, 10 μm. Buffer A was 0.01 N sodium acetate, pH 5.0–acetonitrile (95:5). Elution buffer B was 100% acetonitrile, increasing from 20% at the start to a final concentration of 60%. The flow rate was 2.0 mL/min. This is a modification of Zimmerman et al. (1976). All solvents used were purchased from Burdick & Jackson.

Results

Nomenclature Used for Peptides. In the following, peptides isolated from tryptic digests of maleylated or cyclohexanedione-treated α1-CB8 and from thermolytic digests of α1-CB8 are designated TM, TCHD, and Th, respectively. In each case these initial letters are followed by a number indicating the position of the particular peptide, sequentially from the NH_2 terminus in the α1-CB8 sequence as finally determined, relative to other peptides isolated from similar digests. One additional peptide, isolated from chymotryptic digests of α1-CB8, was termed C1.

Tryptic Peptides of Maleylated α1-CB8. The tryptic digests of maleylated α1-CB8 were initially fractionated on a Sephadex G-50S column as indicated. In the early experiments it was found that the tryptic peptides of α1-CB8 appeared to show a great tendency to associate, making purification difficult under ordinary conditions. It was later found that the presence of 4 M urea in the eluting buffers helped greatly. The columns were also operated at a slightly higher temperature, 42 °C.

Figure 1 shows a typical G-50S chromatogram of the tryptic digest of maleylated α1-CB8 made under these conditions. The eluted fractions were pooled separately as shown, the urea and acetate buffer was removed by desalting on Bio-Gel P2 columns of appropriate size, and then the fractions were lyophilized and analyzed. In most cases it was possible to purify the fractions sufficiently for sequencing by repeating the molecular sieve chromatography on Sephadex G-50S. In this way the following peptides were obtained: a 63-residue fragment, TM6 from fraction 1 (Figure 1), TM3 of 54 residues from fraction 2, TM2 of 45 residues of fraction 3, and a 15-residue TM4 peptide from fraction 7. Amino acid composition of these peptides is shown in Table I; the details of sequence analyses are given in Table III and the results in Figure 5.

Tryptic Peptides of Cyclohexanedione-Treated α1-CB8. Cyclohexanedione-blocked α1-CB8 was digested with trypsin and initially fractionated on Sephadex G-50S in a similar manner used on the tryptic digest of the maleylated protein. A typical chromatogram is shown in Figure 2.

Fraction 6 yielded the 37-residue fragment, TCHD3, while the 33-residue TCHD1 was obtained from fraction 7. TCHD4, a 28-residue peptide which contained 1 residue of homoserine and must, therefore, be the COOH terminal in α1-CB8, came from fraction 8 and TCHD2, a 20-residue peptide, from fraction 9. The latter was further purified on phosphocellulose (not shown). The amino acid compositions of these peptides are shown in Table II.

The three initial smaller peaks represent larger overlap fragments arising because of slower cleavages at one or more cleavable bonds. This interpretation was supported by the fact that their amount decreased with increasing time of digestion. As they were not needed in the final sequence reconstruction, however, they were not further investigated.

Thermolytic Peptides of α1-CB8. The initial fractionation of thermolysin-digested α1-CB8 on Sephadex G-50S is shown in Figure 3. The first fraction containing Th1 (residues 1–103) and Th2 (residues 104–199) was rechromatographed on a phosphocellulose column to resolve these peptides (Figure 4A). The fraction containing Th3, -4, and -5 (Figure 3) was similarly resolved by phosphocellulose chromatography (Figure 4B). The amino acid composition of the thermolytic peptides is presented in Table II.

Thermolytic peptide Th1 was treated with cyclohexanedione as above to block arginyl residues. After tryptic digestion, the material was chromatographed on a phosphocellulose column when two major peptides Th1A and Th1B were obtained (figure not shown). Only Th1A (residues 52–96) was necessary for the determination of the sequence.

Table II: Amino Acid Composition of Purified Tryptic Peptides of Cyclohexanedione-Treated α1-CB8 and Th1 and Thermolytic Peptides of α1-CB8[a]

amino acid	TCHD1	TCHD2	TCHD3	TCHD4	Th1A	Th1	Th2	Th3	Th4	Th5
4-hydroxyproline	4.1 (4)	1.7 (2)	3.5 (4)	3.0 (3)	3.8 (4)	11 (11)	11 (11)	4.6 (5)	5.2 (5)	5.2 (5)
aspartic acid	0.4		1.2 (1)	1.0 (1)	1.9 (2)	4.0 (4)	4.0 (4)	1.2 (1)		2.0 (2)
threonine	0.8 (1)		0.2	0.6 (1)	0.8 (1)	2.7 (3)	1.1 (1)			1.9 (2)
serine	1.5 (2)		1.0 (1)	0.2	1.8 (2)	2.7 (3)	3.8 (4)	0.9 (1)	1.8 (2)	1.8 (2)
glutamic acid	1.4 (1)	3.6 (4)	2.1 (2)	1.8 (2)	6.2 (6)	7.3 (7)	8.3 (8)	4.8 (5)	3.2 (3)	2.3 (2)
proline	4.3 (4)	3.0 (3)	3.1 (3)	3.6 (4)	5.2 (5)	12 (12)	9.3 (9)	4.3 (4)	3.2 (3)	5.1 (5)
glycine	11 (11)	7.1 (7)	12 (12)	9.4 (9)	15 (15)	35 (35)	32 (32)	15 (15)	13 (13)	13 (13)
alanine	5.9 (6)	2.3 (2)	6.2 (6)	4.0 (4)	7.0 (7)	18 (18)	13 (13)	5.7 (6)	5.0 (5)	3.8 (4)
valine	0.3	0.9 (1)	0.1	0.7 (1)			0.8 (1)	0.7 (1)	0.9 (1)	0.8 (1)
isoleucine	0.8 (1)		0.8 (1)				0.8 (1)		0.8 (1)	
leucine	0.1	0.1	0.9 (1)			1.1 (1)	1.0 (1)	1.0 (1)	1.1 (1)	1.0 (1)
phenylalanine	0.8 (1)		0.9 (1)			1.1 (1)	1.9 (2)	1.0 (1)		
lysine	0.8 (1)	0.9 (1)	1.0 (1)	0.3	0.8 (1)	2.0 (2)	3.8 (4)	0.8 (1)	3.0 (3)	1.0 (1)
arginine[b]	1 (1)		4 (4)	2 (2)	2.0 (2)	6.0 (6)	5.0 (5)	3.8 (4)	2.0 (2)	2.0 (2)
homoserine[c]				0.7 (1)						0.8 (1)
total	33	20	37	28	45	103	96	45	39	41

[a] Results are expressed as residues per peptide. The number is rounded off to the nearest whole number where more than 10 residues occur. The numbers in parentheses indicate the residues calculated from the elucidated sequence. [b] Determined from sequence results. [c] Includes homoserine lactone.

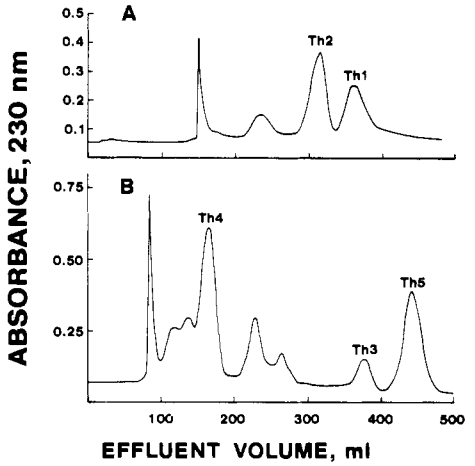

FIGURE 4: Phosphocellulose chromatography of thermolytic peptide mixtures obtained from Figure 3. The column (1 × 6 cm) was equilibrated with 0.001 M acetate, pH 3.8, and was eluted with a linear gradient consisting of 250 mL of starting buffer and 250 mL of 0.001 M acetate, pH 3.8, containing 0.2 M NaCl at a flow rate of 45 mL/h.

COOH-Terminal Chymotryptic Peptide of α1-CB8. Chymotryptic digests were also initially fractionated on a large column of Sephadex G-50S, as described earlier, but in these cases with no urea present in the eluting buffer.

Experience with this method appeared to indicate that there probably are no bonds present in α1-CB8 that are specifically cleavable by chymotrypsin. Although there are three phenylalanyl residues present, later results showed that these are all immediately followed by Hyp, effectively blocking them from cleavage. On the other hand, all the chymotryptic digests gave rise to an easily resolvable peak (chromatograph not shown), which proved to be a 40-residue peptide arising from cleavage of a Leu–Thr bond at 239–240 of α1-CB8 and which must represent the COOH-terminal peptide of α1-CB8 since it contained a residue of homoserine. This was called peptide C1. The amino acid composition is shown in Table I.

Reconstruction of the α1-CB8 Sequence. The sequence of α1-CB8 was established as follows from the results obtained on the component peptides described above (Table III). Each peptide was analyzed at least twice. The two analyses on intact α1-CB8 gave the NH_2-terminal sequence up to residue 80 with the 39-residue TM1 fitting residues 22–60 (Figure 5). Peptide Th1 was sequenced through the first 40 residues, confirming the results obtained from analyses of α1-CB8 and TM1.

Table III: Automatic Sequence Analyses of Enzymatically Produced Fragments of Chick Skin Collagen α1-CB8

peptide	residue no.	μmol	program	modification	cycles of degradation
α1-CB8	1–279	0.72	Quadrol	none	80
TM1	22–60	0.50	Quadrol	ANS	39
TM2	70–114	0.54	Quadrol	ANS	45
TM2T1	70–96	0.38	Quadrol	ANS	27
TM2T2	97–114	0.24	Quadrol	ANS	18
TCHD1	97–129	0.48	Quadrol	ANS	33
TM3	115–168	0.50	Quadrol	S-PITC, ANS	34
TCHD2	148–167	0.80	Quadrol	S-PITC	20
TCHD3	168–204	0.54	Quadrol	S-PITC	37
TM4	172–186	0.31	Quadrol	ANS	15
TM5	193–210	0.41	Quadrol	ANS	18
TM6	211–273	0.48	Quadrol	ANS	47
CH1	240–279	0.54	Quadrol	ANS	40
TCHD4	252–279	0.44	Quadrol	ANS	28
Th1	1–103	0.50	DMAA	ANS	40
Th2	104–199	0.50	DMAA	ANS	47
Th3	155–199	0.20	DMAA	ANS	20
Th4	200–238	0.20	DMAA	ANS	24
Th5	239–279	0.30	DMAA	ANS	24
Th1A	52–96	0.20	DMAA	ANS	29

Sequence analysis of Th1A (52–96) from residues 52–80 aligned TM1 and TM2 (70–114), bringing the known sequence through 114 residues. Certain gaps were then filled by the sequence of the component tryptic peptides TM2T1 (70–96) and TM2T2 (97–114). This permitted the alignment of the 33-residue TCHD1 (97–129), which brought the sequence to 129 residues. The alignment of 54-residue TM3 (115–168) was then deduced from the sequence data of TCHD1 and of TM3 and Th2 (104–199) sequenced through the first 34 and 47 residues, respectively, bringing the known sequence through 150 residues. The sequence of TCHD2 (148–167) plus the sequence of Th3 (155–199) through the first 20 residues overlaps peptides TM3 and TCHD3 (168–204), confirming their alignment and bringing the sequence through 204 residues. Further alignment was obtained from the sequence of Th4 (200–238) through the first 24 residues which overlaps TM5 (193–210) and TM6 (211–273). TM6 was sequenced through the first 36 residues, extending the sequence through 246 residues.

Chymotryptic peptide C1 (240–279) must represent the COOH-terminal 40-residue peptide since it contains a residue

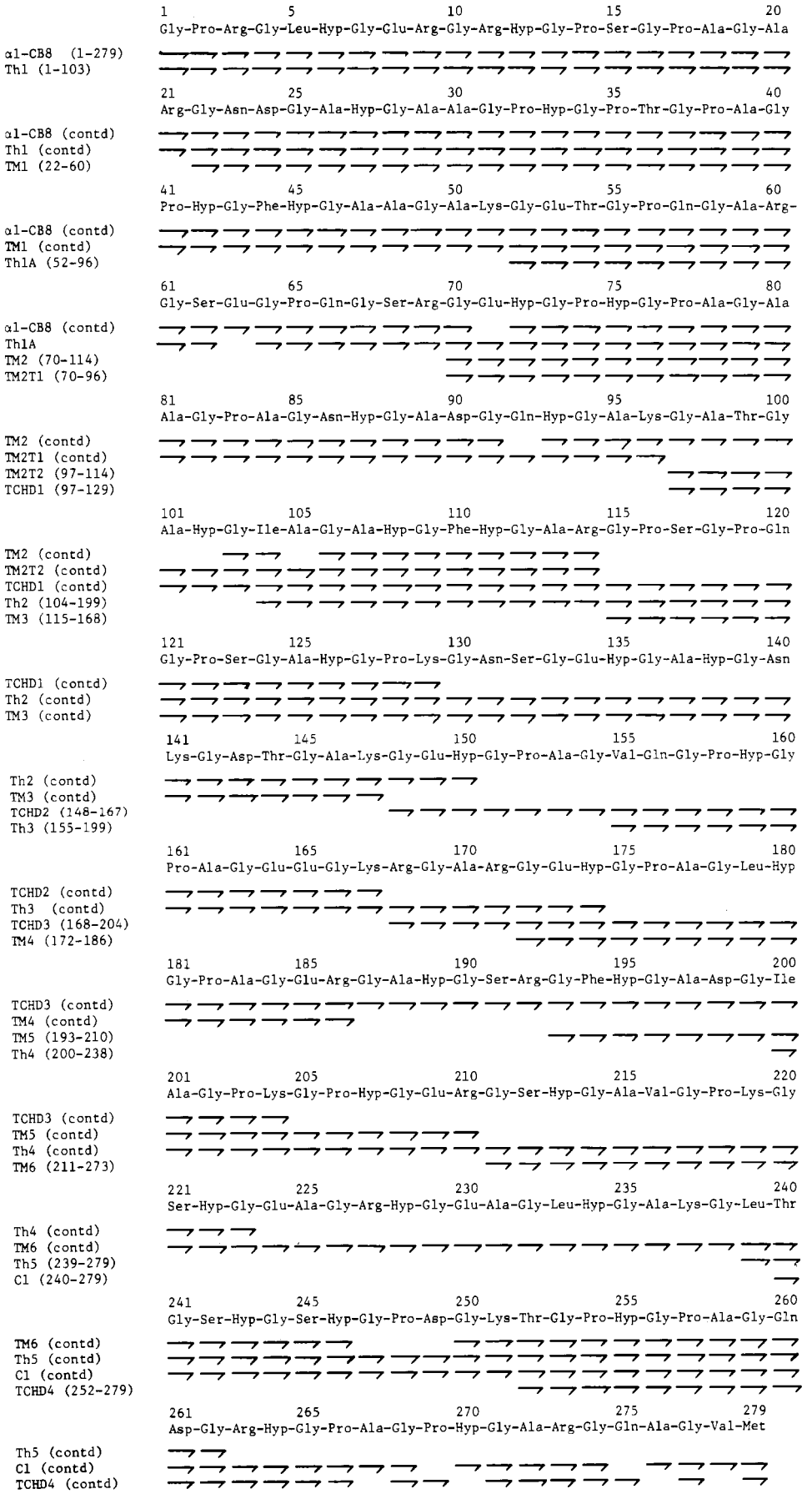

FIGURE 5: Reconstruction of the sequence of chick α1-CB8 from results obtained on the component peptides. Each horizontal half-arrow indicates one cycle of Edman degradation with identification of the residue by one or more of the methods described in the text.

Table IV: Comparison of Chick Skin α1(I) Sequence with Those of Rat-Calf α1(I) and Chick Skin α2[a]

No. in helix	5	6	11	14	15	17	21	26	35	36	39	45	51	53	56	57	101	102	138	140	150	162	177	191	222	300
Chick α1(I)	Pro	Ala	Leu	Pro	Hyp	Ala	Gln	Pro	Ala	Ser	Met	Ala	Asn	Asp	Glu	Ala	Gln	Hyp	Ser	Pro	Hyp	Ala	Thr	Ser	Thr	Ala
Rat/Calf α1(I)		Ser										Hyp					Asn	Thr	Hyp	Ser	Val	Thr	Ala	Val	Asn	Ser
Chick α2	Leu	Met	Pro	Ala	Ser	Pro	Hyp	Val	Gln	Thr	Gln	Hyp	Ala	Glu	Ala	Leu	___	___	___	___	___	___	___	___	___	___

No. in helix	306	311	323	330	332	335	336	338	339	341	344	345	347	354	357	359	360	363	365	372	375	378	380	381	383	387
Chick α1(I)	Ala	Ala	Ile	Hyp	Glu	Ser	Hyp	Ala	Val	Pro	Ser	Hyp	Glu	Ala	Hyp	Ala	Lys	Thr	Ser	Asp	thr	Hyp	Pro	Ala	Glu	Hyp
Rat/Calf α1(I)	Hyp	Gly	Val	Ala					Ala																	
Chick α2	___	___	___	Ala	Asn	Ala	Ser	Pro	Ala	Val	Pro	Asn	Asp	Hyp	Met	Pro	Arg	Hyp	Gln	Ala	Glu	Val	Phe	Hyp	Ala	Val

No. in helix	390	393	395	398	399	401	402	410	411	417	420	422	423	425	426	428	429	431	432	434	435	437	438	446	447	449
Chick α1(I)	Ala	Hyp	Ala	Gln	Ala	Val	Met	Ala	Ala	Hyp	Arg	Ala	Hyp	Pro	Hyp	Ala	Val	Ala	Ala	Lys	Asp	Glu	Ala	Pro	Thr	Pro
Rat/Calf α1(I)										Ala		Val						Pro							Ala	
Chick α2	Ile	Ala	Asn	Glu	Hhp	Asn	Ile	Pro	Thr		Lys	Asn	Val	Leu	Ala	Pro	Arg		Hyp	Pro	Glu	Asn	Asn	Val		Asn

No. in helix	450	452	453	456	461	471	474	483	485	486	488	489	491	495	497	503	507	509	510	513	516	518	531	522	524	527
Chick α1(I)	Ala	Glu	Arg	Gln	Ala	Ala	Hyp	Gln	Val	Hyp	Asn	Ala	Ala	Ala	Ala	Phe	Arg	Val	Gln	Hyp	Gln	Pro	Ala	Asn	Ala	Asn
Rat/Calf α1(I)					Ser							Leu		Ser					Glu		Ala					
Chick α2	Gln	Ala	Lys	Thr	Pro	Ser	Ala	Arg	Leu	His	Glu	Phe	Val		Pro	Leu	Ser	Ala	Val	Ala	Ile	Ser	Pro	Ser	Pro	Pro

No. in helix	530	533	534	536	537	539	540	542	543	545	546	548	549	551	557	560	563	566	567	569	570	572	573	575	576	579
Chick α1(I)	Ala	Asp	Ala	Ala	Hyp	Ala	Hyp	Asn	Glu	Pro	Hyp	Leu	Glu	Met	Ala	Leu	Ala	Asp	Arg	Asp	Hyp	Pro	Lys	Ala	Asp	Pro
Rat/Calf α1(I)								Ser	Gln	Ala			Gln				Pro				Ala				
Chick α2	Asn	Glu	Hyp	Asn	Val	Pro	Ala	Ala	Hyp		Ala	Pro	Hyp	Ile	Val	Val	Gly	Glu	Lys	Ala		Leu	Arg	Asp	Thr	Thr

No. in helix	581	584	588	590	593	599	600	603	606	609	615	617	620	623	629	630	633	642	650	651	653	654	656	659	660	663
Chick α1(I)	Lys	Leu	Thr	Pro	Pro	Ala	Hyp	Lys	Ala	Hyp	Thr	Ala	Ala	Asp	Pro	Hyp	Ala	Asp	Glu	Thr	Asp	Ala	Ala	Asp	Ala	Hyp
Rat/Calf α1(I)		Val								Ser									Asp	Hyp						
Chick α2	Arg	Ala	Hyp	Ala	Ala	Gly	Ala	Arg	Gly	Ala	Ala	Phe	Ile	Glu	Arg	Val	Ser	Ala	Ala	Arg	Pro	Lys	Pro	Glu	Thr	Thr

No. in helix	665	666	669	671	672	675	677	681	684	686	689	695	707	711	713	714	716	726	728	729	731	734	735	738	741	744
Chick α1(I)	Pro	Ala	Thr	Ala	Hyp	Ala	Glx	Hyp	Hyl	Ala	Ser	Ala	Pro	Ser	Asn	Ile	Leu	Glx	Ser	Lys	Glu	Thr	Thr	Ala	Hyp	Hyp
Rat/Calf α1(I)			Ala	Pro		Thr	Asn				Pro	Phe				Ala	Pro		Pro		Pro					Val
Chick α2	Ala	Ile	Ile		Ser	Hyp	Pro	Ala	Ala	Pro	Asp	Met	Thr	Ala	Ile	Thr	Pro	Asp	Pro	Arg	Leu	Asp	Val	Val	Thr	Gln

No. in helix	746	747	752	753	758	759	761	762	764	765	767	777	779	780	782	783	788	794	797	798	800	801	806	807	834	845
Chick α1(I)	Pro	Ala	Pro	Hyp	Ser	Hyp	Ala	Asp	Pro	Ile	Ala	Ala	Gln	Arg	Val	Val	Gln	Phe	Leu	Hyp	Pro	Ser	Lys	Gln	Ala	Ala
Rat/Calf α1(I)				Ala	Ala					Ala															Ser	Ser
Chick α2	Ile	Ala	Phe	Ala	Pro	Ser	Glu	Ala	Ala	Ala	Pro	Leu	Ala	Hyp	Ile	Leu	Ser	Leu	Ile	Ala	Ala	Thr	Pro	Leu	___	___

No. in helix	851	852	854	866	869	885	897	900	929	947	971	972	995	996	999
Chick α1(I)	Ala	Ala	Pro	Pro	Ala	Asn	Ala	Hyp	Met	Ala	Ala	Ala	Glu	Val	Val
Rat/Calf α1(I)	Ser	Hyp	Ala	Ala	Pro	Ser	Ile	Val	Ile	Ser	Ser	Hyp	Asp	Ala	Ala
Chick α2	___	___	___	___	___	___	___	___	___	___	___	___	___	___	___

[a] Only positions where substitutions are found in chick skin α1(I) are shown, and glycines in the first triplet position are omitted. Dashes indicate that that sequence has not been determined. Rat-calf α1(I) sequence, Hulmes et al. (1973) and Bornstein & Traub (1979). Chick α2 sequence: residues 5–57, Kang & Gross (1970) and S. N. Dixit et al. (unpublished results); residues 330–357, Highberger et al. (1971); residues 359–489, Dixit et al. (1977a); residues 491–695, Dixit et al. (1977b); residues 707–807, Dixit et al. (1979).

of homoserine. The overlap of TM6 with TCHD4 (252–279) was established by sequence determination of C1 and Th5 (239–279) through the first 24 residues. Peptides C1 and TCHD4 (252–279) extend the sequence through 279 residues of α1-CB8. The number of residues in α1-CB8 established by the sequence data, 279, is in agreement, within the experimental error of the analytical methods, with the analytical value of 281 (Table I). The order of peptide alignment determined in this study is in agreement with the reported alignment of homologous peptides of α1-CB8 of rat skin collagen (Balian et al., 1971, 1972).

Discussion

The amino acid sequence of chick skin collagen α1-CB8, as determined in the present work, shows 95% identity with the α1-CB8 of rat skin collagen (Balian et al., 1971, 1972). In the rat sequence it was found that a certain amount of underhydroxylation of proline existed, to wit, residues 87 and 93. Our data for the chick peptide also show significant proline underhydroxylation near this region. The analytical data for the peptide TM2 (residues 70–114 of α1-CB8) show this clearly as seen in Table I. Most of this seems to occur in the component tryptic peptide TM2T2 (residues 97–114), which contains three hydroxyprolines at positions 102, 108, and 111. Our data, however, do not permit determination of the distribution of the total underhydroxylation among these.

The determination of the sequence of chick skin α1-CB8 permits the assembly of a complete amino acid sequence for the chick skin collagen α1(I) chain. This is shown in Figure 6. As indicated there, this includes only 6 residues of the 26 presumed to exist in the COOH-terminal nonhelical telopeptide by homology with the calf sequence. For reasons as yet unknown, it has thus far not been possible to isolate the complete COOH-terminal telopeptide in spite of exercising all of the obvious precautions.

The 1014 residues in the helical portion of the chick sequence show 93.4% sequence identity with the corresponding part of the rat–calf composite assembled by Hulmes et al. (1973) and Bornstein & Traub (1979). A complete comparison of the chick α1 with chick α2 cannot yet be made, but the 539 residues of the α2 chain that are presently published (Dixit et al., 1977a,b, 1979) show 84.1% sequence identity with the corresponding part of the α1 sequence. Thus, the interspecies homology of the α1(I) chains is considerably greater than the intraspecies homology of the α1 and α2 chains. This comparison was previously noted by Dixit et al. (1979). Table IV shows a complete comparison of the substitutions in the chick skin α1(I) chain with the rat–calf sequence and with the chick skin α2 chain as far as published.

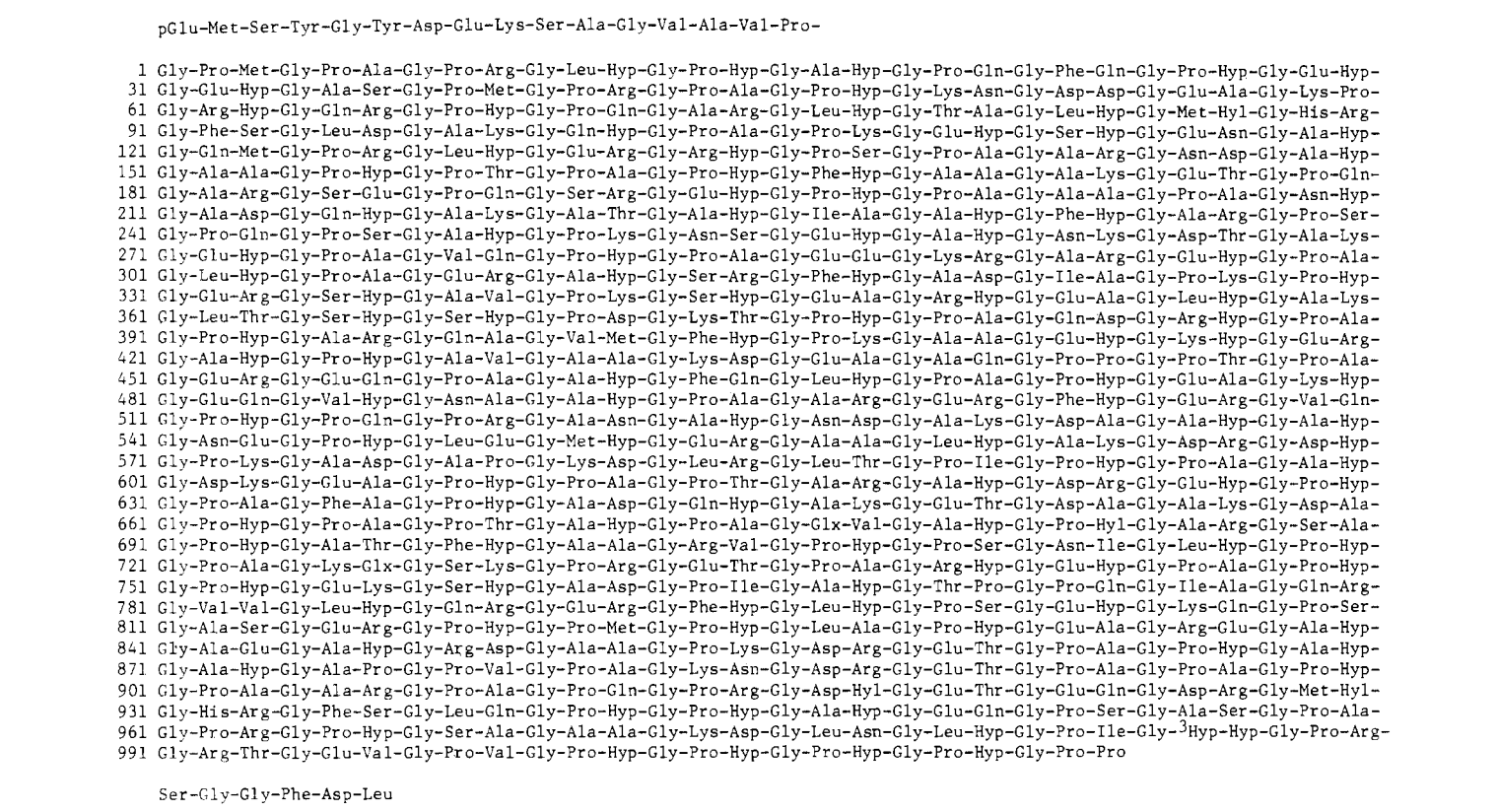

FIGURE 6: Amino acid sequence of chick skin collagen $\alpha 1$(I). Residue numbering starts with the helical portion. NH_2 terminus through residue 39, Kang & Gross (1970); residues 40–123, Kang et al. (1975); residues 124–403, J. H. Highberger et al. (this paper); residues 404–551, Dixit et al. (1975a); residues 552–822, Highberger et al. (1975, 1978); residues 823–929, Dixit et al. (1975b); residues 930–1020, Dixit et al. (1978).

The largest CNBr peptide of the $\alpha 1$(I) chain is CB8. As such, it is not surprising that it is one of the chief loci of the chromatographic heterogeneity discovered in the earliest investigation (Butler et al., 1967) of the CNBr peptides of the $\alpha 1$(I) chain of rat collagen. The standard procedures of purifying $\alpha 1$-CB8 by removing the minor components (the more acidic single prime and the more basic double prime) were carried out in the present case in preparing the material used for sequencing. The chromatographic separation of the components, however, is far from perfect, and it is to be expected that the major components will be contaminated with small amounts of the minor components.

The explanation proposed by Butler et al. (1967) for these results, namely, that they are the result of the equilibrium between the free acid form of the COOH-terminal homoserine and its lactone form, no longer holds in the light of a more recent experience with CNBr peptides. Offord (1972) has shown that in peptides the homoserine equilibrium occurs much less readily than with free homoserine and apparently even less readily the larger the peptide. This is confirmed by our own experience with the 30-residue $\alpha 2$-CB2 (Highberger et al., 1971), where chromatographic heterogeneity due to homoserine equilibrium was only observed after it had been cleaved by trypsin into two 15-residue components. But the heterogeneity we are concerned with occurs only in the two largest CNBr peptides of $\alpha 1$(I) and is not seen at all in the remaining smaller peptides, all of which, except the COOH-terminal one, also contain homoserine. A much more probable explanation is that minor sequence differences occur, giving rise to the major and single prime variant forms. This leaves the double prime minor variant unexplained. It seems possible that this may come from type III collagen since two of the CNBr peptides of this collagen chromatograph very close to the positions of $\alpha 1$(I)-CB7 and $\alpha 1$(I)-CB8. Butler et al. (1977a,b) have recently found a similar situation to that outlined above, in the case of type II collagen of bovine nasal cartilage. These authors have presented evidence that at least two types of $\alpha 1$(II) chains exhibiting minor sequence differences exist in this tissue. They have designated these $\alpha 1$(II) major and $\alpha 1$(II) minor.

This situation in the case of chick collagen $\alpha 1$(I) is now under investigation in one of our laboratories.

References

Allmann, H., Fietzek, P. P., Glanville, R. W., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 861–868.

Balian, G., Click, E. M., & Bornstein, P. (1971) *Biochemistry 10*, 4470–4478.

Balian, G., Click, E. M., Hermodson, M. A., & Bornstein, P. (1972) *Biochemistry 11*, 3798–3806.

Bentz, H., Fietzek, P. P., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 833–840.

Bornstein, P., & Traub, W. (1979) *Proteins (3rd Ed.) 4*, 412–632.

Braunitzer, G., Schrank, B., & Ruhfus, A. (1970) *Hoppe-Seyler's Z. Physiol. Chem. 351*, 1589–1590.

Butler, P. J. G., Harris, J. I., Hartley, B. X., & Leberman, R. (1969) *Biochem. J. 112*, 679–689.

Butler, W. T., & Ponds, S. L. (1971) *Biochemistry 10*, 2076–2081.

Butler, W. T., Piez, K. A., & Bornstein, P. (1967) *Biochemistry 6*, 3771–3780.

Butler, W. T., Miller, E. J., & Finch, J. E., Jr. (1976) *Biochemistry 15*, 3000–3006.

Butler, W. T., Finch, J. E., Jr., & Miller, E. J. (1977a) *J. Biol. Chem. 252*, 639–643.

Butler, W. T., Finch, J. E., Jr., & Miller, E. J. (1977b) *Biochemistry 16*, 4981.

Clark, C. C., & Bornstein, P. (1972) *Biochemistry 11*, 1468–1474.

Dewes, H., Fietzek, P. P., & Kühn, K. (1979a) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 821–832.

Dewes, H., Fietzek, P. P., & Kühn, K. (1979b) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 851–860.

Dixit, S. N., Kang, A. H., & Gross, J. (1975a) *Biochemistry 14*, 1929–1933.

Dixit, S. N., Seyer, J. M., Oronsky, A. O., Corbett, C., Kang, A. H., & Gross, J. (1975b) *Biochemistry 14*, 1933–1938.

Dixit, S. N., Seyer, J. M., & Kang, A. H. (1977a) *Eur. J. Biochem. 73*, 213–221.

Dixit, S. N., Seyer, J. M., & Kang, A. H. (1977b) *Eur. J. Biochem. 81*, 599–607.

Dixit, S. N., Seyer, J. M., Kang, A. H., & Gross, J. (1978) *Biochemistry 17*, 5719–5722.

Dixit, S. N., Mainardi, C. L., Seyer, J. M., & Kang, A. H. (1979) *Biochemistry 18*, 5416–5422.

Fietzek, P. P., Furthmayr, H., & Kühn, K. (1974) *Eur. J. Biochem. 47*, 257–261.

Fietzek, P. P., Allmann, H., Rauterberg, J., Henkel, W., Wachter, E., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 809–820.

Foster, J. A., Bruenger, E., Hu, C. L., Albertson, K., & Franzblau, C. (1973) *Biochem. Biophys. Res. Commun. 56*, 70–74.

Highberger, J. H., Kang, A. H., & Gross, J. (1971) *Biochemistry 10*, 610–616.

Highberger, J. H., Corbett, C., Kang, A. H., & Gross, J. (1975) *Biochemistry 14*, 2872–2881.

Highberger, J. H., Corbett, C., Kang, A. H., & Gross, J. (1978) *Biochem. Biophys. Res. Commun. 83*, 43–49.

Hulmes, D. J. S., Muller, A., Parry, D. A. D., Piez, K. A., & Woodhead-Galloway, J. (1973) *J. Mol. Biol. 79*, 137–148.

Kang, A. H., & Gross, J. (1970) *Biochemistry 9*, 796–804.

Kang, A. H., Dixit, S. N., Corbett, C., & Gross, J. (1975) *J. Biol. Chem. 250*, 7428–7435.

Lang, H., Glanville, R. W., Fietzek, P. P., & Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 841–850.

Offord, R. E. (1972) *Biochem. J. 139*, 499–501.

Patthy, L., & Smith, E. L. (1975) *J. Biol. Chem. 250*, 557–564.

Seyer, J. M., & Kang, A. H. (1977) *Biochemistry 16*, 1158–1164.

Seyer, J. M., & Kang, A. H. (1978) *Biochemistry 17*, 3404–3411.

Seyer, J. M., & Kang, A. H. (1981) *Biochemistry 20*, 2621–2627.

Seyer, J. M., Mainardi, C. L., & Kang, A. H. (1980) *Biochemistry 19*, 1583–1589.

Smithies, O., Gibson, D., Fanning, E. M., Goodfliesch, R. M., & Gilman, J. G. (1971) *Biochemistry 10*, 4912–4921.

Trelstad, R. L., & Lawley, K. R. (1976) *Anal. Biochem. 70*, 287–289.

Zimmerman, C. L., Apalla, E., & Pisano, J. J. (1976) *Anal. Biochem. 75*, 77–85.