Grotto, L., Moroz, C., de Vries, A., and Goldblum, N. (1967), *Biochim. Biophys. Acta 133,* 356.
Goucher, C. R., and Flower, H. H. (1964), *Toxicon 2,* 139.
Hendon, R. A., and Fraenkel-Conrat, H. (1971), *Proc. Nalt. Acad. Sci. U.S.A. 68,* 1560.
Houssay, B. A. (1930), *C. R. Seances Soc. Biol. Ses Fil. 105,* 308.
Keyser, J. W. (1964), *Anal. Biochem. 9,* 249.
Kondo, H., Kondo, S., Ikezawa, H., Murrata, R., and Ohsaka, A. (1960), *Jpn. J. Med. Sci. Biol. 13,* 43.
Laure, C. J. (1975), *Hoppe-Seyler's Z. Physiol. Chem. 356,* 213.
Lin, Y., Means, G. E., and Feeney, R. E. (1969), *J. Biol. Chem. 244,* 789.
Mandelbaum, F. R., Reichl, A. P., and Asakura, M. T. (1975), *Toxicon 13,* 109.
Markland, F. S., and Damus, P. S. (1971), *J. Biol. Chem. 246,* 6460.
Nikai, T., Sugihara, H., and Tanaka, T. (1977), *Yakugaku Zasshi, 97,* 507.
Ohsaka, A., Ikezawa, H., Kondo, H., and Kondo, S. (1960), *Jpn. J. Med. Sci. Biol. 13,* 73.
Ohsaka, A., Just, M., and Habermann, E. (1973), *Biochim. Biophys. Acta 323,* 415.
Omori-Satoh, T., and Ohsaka, A. (1970), *Biochim. Biophys. Acta 207,* 432.
Oshima, G., Iwanaga, S., and Suzuki, T. (1968), *J. Biochem. (Tokyo) 64,* 215.
Ownby, C. L., Kainer, R. A., and Tu, A. T. (1974), *Am. J. Pathol. 76,* 401.
Penke, B., Ferenczi, R., and Kovacs, K. (1974), *Anal. Biochem. 60,* 45.
Seifter, S., and Harper, E. (1970), *Methods Enzymol. 19,* 613.
Spiekerman, A. M., Fredericks, K. K., Wagner, F. W., and Prescott, J. M. (1973), *Biochim. Biophys. Acta 293,* 464.
Strickland, E. H. (1974), *CRC Crit. Rev. Biochem. 2,* 113.
Takahashi, T., and Ohsaka, A. (1970), *Biochim. Biophys. Acta 207,* 65.
Toom, P. M., Squire, P. G., and Tu, A. T. (1969), *Biochim. Biophys. Acta 181,* 339.
Tsao, F. H. C., Keim, P. S., and Heinrikson, R. L. (1975), *Arch. Biochem. 167,* 706.
Tu, A. T. (1973), *Annu. Rev. Biochem. 42,* 235.
Tu, A. T. (1977), Venoms: Chemistry and Molecular Biology, New York, N.Y., Wiley.
Tu, A. T., Bjarnason, J. B., and Hruby, V. J. (1978), *Biochim. Biophys. Acta 533,* 530.
Wagner, F. W., Spiekermann, A. M., and Prescott, J. M. (1968), *J. Biol. Chem. 243,* 4486.
Yu, N., and Jo, B. H. (1975), *Arch. Biochem. Biophys. 156,* 469.

Covalent Structure of Collagen: Amino Acid Sequence of Five Consecutive CNBr Peptides from Type III Collagen of Human Liver†

Jerome M. Seyer* and Andrew H. Kang

ABSTRACT: Type III collagen was solubilized from human liver by limited pepsin digestion and purified by differential salt precipitation and carboxymethylcellulose chromatography. Digestion with cyanogen bromide yielded the nine distinct peptides previously described and an additional tripeptide not recognized in earlier studies. Five of these peptides, α1(III)-CB1, 2, 4, 8, and 10, were further purified by molecular sieve and/or ion exchange chromatography. They contained 12, 40, 149, 125 and 3 amino acid residues, respectively. The amino acid sequence of these peptides was determined by automated Edman degradation of tryptic (before and after maleylation), chymotryptic, thermolytic or hydroxylamine-derived peptide fragments as well as the intact peptides. The alignment of these five peptides within the collagen chain is deduced to be 1-8-10-2-4 by homology with known α1(I) sequences. The known CNBr peptide alignment of the NH_2-terminal portion of type III collagen so far would, therefore, be α1(III)-CB3-7-6-1-8-10-2-4 and correspond to the homologous region of α1(I)-CB0-1-2-4-5-8-3 or residues 11–567 of the α1(III) collagen chain.

Collagen represents nearly one-third of the total body protein of most vertebrates. At least four genetic types are known. The most abundant and extensively studied is type I collagen consisting of two α1(I) chains and one α2 chain as a triple-stranded helix of over 1000 amino acids per chain (Traub & Piez, 1971; Gallop et al., 1972). A composite amino acid sequence of the α1(I) chain can now be constructed from CNBr peptides of calf, chick, and rat skin collagen (Hulmes et al., 1973; Gallop & Paz, 1975; Piez, 1976; Fietzek & Kuhn, 1976) and the sequence of α2 is also nearly complete (Fietzek & Rexrodt, 1975; Dixit et al., 1977a,b).

The genetically distinct type III collagen has recently been identified and found to be nearly as widespread in tissue distribution as type I collagen. Both collagen types I and III occur simultaneously in most connective tissues with the exception of bone which contains type I only, and cartilage, which contains only type II collagen (Miller & Lunde, 1973; Chung & Miller, 1974; Epstein, 1974). Type III collagen, unlike type I, contains intramolecular cysteine-derived cross-linkages and, therefore, can be separated chromatographically as a 280 000 molecular weight species containing three identical α1(III)

† From the Connective Tissue Research Laboratory, Veterans Administration Hospital and Departments of Medicine and Biochemistry, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104. *Received February 9, 1978.* This study was conducted under VA Research Projects No. 4826-01, 4826-02, and 7820-01 and was supported in part by U.S. Public Health Service Grant No. AM-16506.

This article not subject to U.S. Copyright. Published 1978 by the American Chemical Society

chains from 95 000 molecular weight chains of type I collagen (Chung & Miller, 1974). Nine CNBr peptides from type III collagen of human skin and liver have been isolated and characterized (Chung et al., 1974; Seyer et al., 1976). The amino acid sequences of three of these CNBr peptides α1(III)-CB3-7-6 from the amino terminus of the helical region have been determined (Seyer & Kang, 1977), as have partial sequences from α1(III)-CB4, 5, and 6 of calf aorta (Fietzek & Rautherberg, 1975; Fietzek et al., 1977).

This report presents the amino acid sequence of five additional CNBr peptides, α1(III)-CB1, 2, 4, 8, and 10, which together contain 329 amino acid residues.

Materials and Methods

Preparation of CNBr Peptides. Human cirrhotic liver was obtained after autopsy and type III collagen prepared as previously described (Seyer & Kang, 1977; Seyer et al., 1976). The CNBr peptides were obtained by digestion in 70% formic acid and ion-exchange chromatography on CM-cellulose[1] (Seyer & Kang, 1977).

The larger peptides, α1(III)-CB4 and α1(III)-CB8,[2] were further purified by a rechromatography on CM-cellulose. Forty-milligram samples of each peptide, denatured at 45 °C for 10 min, were applied to a 1 × 7 cm column of the resin previously equilibrated with 0.02 M sodium acetate, pH 4.8, at 43 °C (the starting buffer). A concave gradient of NaCl was formed utilizing 830 mL of starting buffer and 500 mL of 0.02 M sodium acetate/0.12 M NaCl, pH 4.8, as the limiting solvent (Chung et al., 1974). A flow rate of 250 mL per h was used and the effluent was continuously monitored at 230 nm. Fractions representing each peak were pooled, lyophilized, desalted on Bio-Gel P-2 in 0.1 M acetic acid, and relyophilized.

The peptides α1(III)-CB1 and α1(III)-CB2 were purified after initial CM-cellulose by gel filtration on Sephadex G-50 SF (Pharmacia). The samples (20 mg) were dissolved in 0.1 M acetic acid, heated to 45 °C for 10 min to ensure denaturation, and chromatographed on 4 × 110 cm columns of Sephadex G-50 SF equilibrated with 0.1 M acetic acid. The effluent was continuously monitored at 230 nm and a drop of tritiated water was used to mark the column volume (Piez, 1968). The smaller peptide, α1(III)-CB1, was further purified on a 0.9 × 25 cm column of PA-35 ion-exchange resin (Beckman Instruments, Palo Alto, Calif.) (Kang & Gross, 1970). An automatic peptide analyzer (Technicon Instruments, Inc., Ardsley, N.Y.) equipped with a split-stream device enabled continuous monitoring of a portion of the effluent for ninhydrin reactivity with the remainder (75%) being collected in a fraction collector. Separation was achieved at 60 °C utilizing a nine-chamber gradient device beginning with 0.02 M sodium citrate buffer, pH 3.1 (Technicon peptide methodology). Fractions were lyophilized and desalted as previously described (Seyer & Kang, 1977).

The peptide α1(III)-CB10 was isolated in a somewhat different manner. Whole CNBr digests of α1(III) chains were lyophilized and chromatographed on a Bio-Gel P-2 column (2 × 110 cm) in 0.1 M acetic acid (Kang & Gross, 1970). The effluent was monitored continuously at 228 nm and the material which eluted immediately prior to tritiated H_2O was collected, lyophilized, and rechromatographed using the automatic peptide analyzer with PA-35 resin as described above. The peptide α1(III)-CB10 failed to elute under these conditions but was removed by a stepwise buffer change to 0.5 M NaCl, in 0.35 M sodium citrate, pH 8.0.

Amino Acid Analysis. Samples were hydrolyzed in constantly boiling HCl at 108 °C for 24 h under an atmosphere of N_2. Analyses were performed on an automatic analyzer (Beckman Instruments, Palo Alto, Calif.) using a single column method previously described (Kang, 1972). No correction factors were used for losses of the labile amino acids or for the incomplete release of valine.

Enzymatic Hydrolysis. Digestions with trypsin (TPCK-treated, 3× crystallized, Worthington) were performed in 0.2 M NH_4HCO_3 (pH 8.3). An enzyme to substrate ratio of 1:50 (w/w) was used as previously described (Seyer & Kang, 1977). In some cases maleylation prior to trypsin digestion was used to limit the cleavage to Arg residues (Butler et al., 1969). Recrystallized maleic anhydride was used (20-fold molar excess over the total number of lysyl residues) and demaleylation was achieved with pyridine acetate, pH 3.0 (1:10) (Dixit et al., 1975).

Chymotrypsin digestion (α-chymotrypsin, 3× crystallized, Worthington) was performed using a 300:1 (w/w) ratio of peptide:enzyme at 37 °C for 30 min. The enzyme together with soybean trypsin inhibitor was dissolved in 0.1 mM HCl (1 mg/mL) and added to the peptide which was previously dissolved in 0.2 M NH_4HCO_4, pH 8.0 (5 mg/mL). After incubation, the digestion mixture was diluted five-fold with cold H_2O and lyophilized (Dixit et al., 1975). Thermolysin digestions were also performed on α1(III)-CB4. Thirty-six milligrams of the peptide dissolved in 7 mL of 0.2 NH_4HCO_3, pH 8.0, was treated with 0.3 mL [1 mL/mL H_2O thermolysin (Worthington)] for 30 min at 37 °C. The reaction was terminated by adjusting the pH to 3.5 with acetic acid and the products chromatographed directly on Sephadex G-50 SF in 0.1 M acetic acid.

Hydroxylamine Cleavage. Cleavage of α1(III)-CB8 was performed with hydroxylamine. The peptide (10 mg) was dissolved in 2 mL of H_2O and mixed with an equal volume of cold, freshly prepared 2 M NH_2OH in 1 M K_2CO_3 (pH 9.0) (Balian et al., 1971). The reaction was allowed to proceed 30 min at 37 °C and terminated by adjusting the pH to 3.5 with HCl. The preparation was applied directly to a Sephadex G-50 SF column (2.0 × 110 cm) previously equilibrated with 0.1 M acetic acid. The separated peptides were monitored at 230 nm.

Edman Degradation. Automatic Edman degradations were performed with a Beckman Sequencer, Model 890C, according to the principles described by Edman & Begg (1967). Both the SLOW PROTEIN-QUADROL (072172C) and the SLOW PEPTIDE-DMAA (071472) programs of Beckman Instruments were employed. Small peptides were treated with 2-amino-1,5-napthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide to minimize extraction of the peptides from the reaction cup (Foster et al., 1973). Slight modifications included the use of smaller quantities of the ANS and CDI allowing identification of free Pth-glutamic acid and aspartic acid, and the degradation of the peptides to their penultimate COOH terminus (Dixit et al., 1975). The homoserine peptides were treated initially with 0.1% NH_4OH for 10 min to convert the homoserine lactone to free homoserine. Ammonia was removed by lyophilization. The Pth-amino acids were identified by either gas chromatography before and after trimethylsilylation (Pisano & Bronzert, 1969),

[1] Abbreviations used: CM-cellulose, carboxymethylcellulose; Pth, phenylthiohydantoin; ANS, 2-amino-1,5-napthalenedisulfonic acid; CDI, *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide.

[2] Nomenclature of CNBr peptides used was according to Chung et al. (1974). Tryptic, chymotryptic, thermolytic, and hydroxylamine-derived peptides were numbered in the order of their elution from the specific columns.

```
                    3
α1(III)-CB10   R G M

                          12
α1(III)-CB1    G P A G I P*G A P*G L M

                                                  30
α1(III)-CB2    P*G S P*G G P*G S D G K P*G P P*G S Q G E S G R P*G P P*G P

                        40
α1(III)-CB2    S G P R G Q P*G V M
```

FIGURE 1: The complete amino acid sequence of α1(III)-CB1, 2, and 10 of human liver type III collagen.

TABLE I: Amino Acid Composition of α1(III)-CB1, 2, and 10 of Type III Collagen from Human Cirrhotic Liver.[a]

amino acid	α1(III)-CB1	α1(III)-CB2	α1(III)-CB10
Hyp	2.0 (2)	7.6 (8)	
Asp	0.3	1.3 (1)	0.3
Thr		0.2	
Ser		4.6 (5)	
Glu		3.2 (3)	0.2
Pro	1.2 (1)	3.9 (4)	
Gly	4.1 (4)	13.7 (14)	1.4 (1)
Ala	2.3 (2)	0.3	0.2
Val		0.8 (1)	
Ile	0.8 (1)	0.2	
Leu	0.9 (1)		
Phe			
Hyl			
Lys		1.1 (1)	
Arg		2.0 (2)	0.9 (1)
Hse[b]	0.8 (1)	0.7 (1)	1.2 (1)
total	12	40	3

[a] Values expressed as residues per peptide. No entry indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values. [b] Includes homoserine lactone.

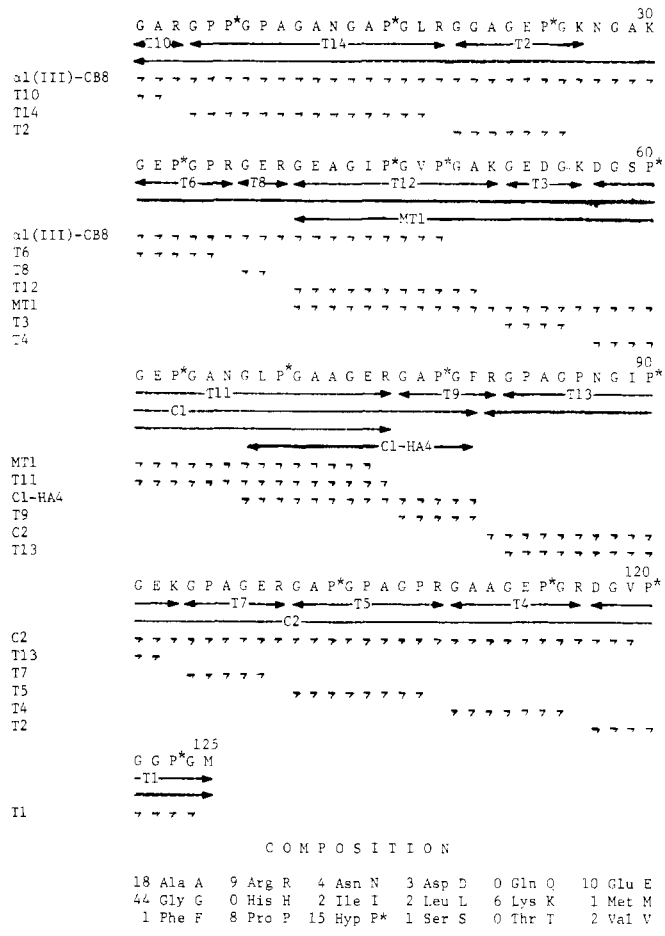

FIGURE 2: The complete amino acid sequence of α1(III)-CB8 of human liver type III collagen. The trypsin, chymotrypsin, and hydroxylamine-derived peptides isolated are indicated by long arrows (↔). Short horizontal arrows (→) indicate the extent of Edman degradation of each peptide.

or by high pressure liquid chromatography (Zimmerman et al., 1973). The COOH-terminal residues were, in most cases, not identified but inferred from the amino acid composition of the peptides and specificities of either trypsin, hydroxylamine, or CNBr. Subsequent overlapping sequence analysis confirmed their identity. In many cases COOH-terminal lysine or phenylalanine was identified as the Pth derivative even though the COOH terminus had been previously modified with ANS.

Results

Amino Acid Sequence of α1(III)-CB1, CB2, and CB10. Three small peptides, α1(III)-CB1, CB2, and CB10, were isolated by a combination of CM-cellulose, gel filtration and, in the case of α1(III)-CB1 and CB10, ion-exchange chromatography on PA-35 resin using an automated peptide analyzer (see paragraph concerning supplementary material at the end of this paper). The complete amino acid sequence of each peptide was determined directly by automated Edman degradation. Their amino acid composition and sequences are presented in Table I and Figure 1, respectively. Derivatization of the COOH-terminal Hse in each case allowed sufficient retention for complete amino acid sequence analysis to the final residue.

Amino Acid Sequence of α1(III)-CB8. The peptide α1(III)-CB8 was isolated by CM-cellulose chromatography at pH 3.8 (data presented in supplementary material) followed by CM-cellulose chromatography at pH 4.8 (data presented in supplementary material). Its amino acid composition is presented in Table II. Amino acid sequence analysis of intact α1(III)-CB8 enabled the identification of the 48 NH_2-terminal residues (Figure 2).

A tryptic digest of α1(III)-CB8 was fractionated on Bio-Gel P-2 which enabled the separation into excluded (>1200 mol wt) and included (<1200 mol wt) peptides. Those small enough to be retained by the gel were further fractionated by ion exchange chromatography (PA-35) into 10 peptides, supplementary material (T1-10), possessing the amino acid compositions presented in Table II. The four larger peptides (T11-14) excluded by Bio-Gel P-2 were separated by phosphocellulose chromatography (supplementary material) into homogeneous peptides. Their amino acid composition is also shown in Table II. All of the lysine and arginine residues were accounted for except an Asn-Gly-Ala-Lys tryptic peptide (residues 27-30) which was not isolated. Its position and sequence was inferred from the Edman degradation of intact α1(III)-CB8, and it is henceforth referred to as T15.

In order to obtain information on the alignment of the tryptic peptides, α1(III)-CB8 was digested with chymotrypsin and fractionated on CM-cellulose (supplementary material). Two chymotryptic peptides, C1 and C2, were further purified by gel filtration on Sephadex G-50 SF. They eluted from the Sephadex column at positions corresponding to estimated molecular weight of 8000 and 4000, respectively. Their amino acid composition is presented in Table II. The peptide C2 contains 45 residues including a residue of homoserine and, therefore, must be the COOH terminus, and C1 containing 80 residues including the phenylalanyl residue must be the NH_2

TABLE II: Amino Acid Composition of α1(III)-CB8, Tryptic, and Chymotryptic Peptides of Type III Collagen from Human Cirrhotic Liver.[a]

amino acid	α1-(III)-CB8	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15[b]	C1	V2	total[c]
Hyp	14.2	1.7	1.2	0.2	0.8	1.0	0.9			0.7		2.8	0.9	1.7	1.6		8.7	4.6	15
	(14)	(2)	(1)		(1)	(1)	(1)			(1)		(3)	(1)	(2)	(2)		(9)	(5)	
Asp	7.4	1.0	0.4	1.2		0.4	0.2	0.1				1.9	1.2	0.1	0.9	(1)	4.9	2.0	7
	(7)	(1)		(1)								(2)	(1)		(1)		(5)	(2)	
Thr	0.2		0.1										0.1				0.3		
Ser	1.2	0.1	0.1		0.1		0.2					0.8					1.1	0.1	1
	(1)											(1)					(1)		
Glu	9.8	0.2	1.1	1.1	1.3	0.1	1.1	1.0	0.8			2.0	0.9	0.9	0.2		6.8	2.9	10
	(10)		(1)	(1)	(1)		(1)	(1)	(1)			(2)	(1)	(1)			(7)	(3)	
Pro	8.9	0.2			0.3	2.1	0.9	1.2				0.3	0.3	2.2	2.2		4.2	4.7	8
	(9)					(2)	(1)	(1)						(2)	(2)		(4)	(5)	
Gly	44.1	4.0	4.2	2.1	3.2	3.4	2.1	2.1	1.1	2.0	1.1	6.1	4.1	4.3	5.1	(1)	28	17	44
	(44)	(4)	(4)	(2)	(3)	(3)	(2)	(2)	(1)	(2)	(1)	(6)	(4)	(4)	(5)		(28)	(17)	
Ala	18.4	0.2	1.2	0.2	1.8	2.2	0.2	0.9		1.0	1.0	3.1	2.1	1.2	3.0	(1)	12	5.9	18
	(18)		(1)		(2)	(2)		(1)		(1)	(1)	(3)	(2)	(1)	(3)		(12)	(6)	
Val	1.7	0.8										0.1	0.8				1.0	0.7	2
	(2)	(1)											(1)				(1)	(1)	
Ile	2.1												0.8	0.9			1.1	0.9	2
	(2)												(1)	(1)			(1)	(1)	
Leu	2.4											0.9			1.1		2.2		2
	(2)											(1)			(1)		(2)		
Phe	0.9									1.1							1.1		1
	(1)									(1)							(1)		
Hyl[d]	0.2													0.1			1.2	0.3	1
	(0.2)																(1)		
Lys[d]	4.6		1.0	0.9	0.2								0.7	0.8		(1)	2.8	0.8	5
	(4.6)		(1)	(1)									(1)	(1)			(3)	(1)	
Arg	9.3				1.0	1.2	1.0	1.1	1.0	1.2	1.0	0.9			1.0		4.6	3.7	9
	(9)				(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)			(1)		(5)	(4)	
Hse[e]	0.9	1.2																0.8	1
	(1)	(1)																(1)	
total	125	9	8	5	8	9	6	6	3	6	3	19	12	12	15	4	80	45	125

[a] Values expressed as residues per peptide. No entry indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values. [b] Assumed based on amino acid sequence analysis of α1(III)-CB8. [c] Total of tryptic peptides above plus the Asn-Gly-Ala-Lys which was not isolated. [d] Partial lysine hydroxylation was noted. [e] Includes homoserine lactone.

terminus. Automated Edman degradation of C2 through 39 residues provided the information necessary to align T13-7-5-4-1 (Figure 2). Edman degradation of the individual tryptic peptides confirmed the results obtained from the sequence analysis of C2.

Sequential degradation of intact α1(III)-CB8 and C1 through 48 residues enabled the positioning of T10-14-2-15-6-8-12 in that order (Figure 2). That T3 and T13 follow T12 in the alignment was determined by isolation and characterization of the tryptic cleavage products obtained from maleylated α1(III)-CB8. Phosphocellulose chromatography (supplementary material) of a tryptic digest of maleylated α1(III)-CB8 yielded a major peak, which upon subsequent gel filtration on Sephadex G-50 SF separated into two peaks: the larger one eluting at the position corresponding to an estimated molecular weight of approximately 3400 contained MT1, and a smaller one corresponding to an estimated molecular weight of 2000 contained a mixture of T(2-15-6) and T(13-7). The larger peptide, MT1, had an amino acid composition consistent with it being T(12-3-11) (residues 40–75). Automated Edman degradation of demaleylated MT1 established the alignment and the sequence T12-13-11. In summary, then, the order of the tryptic peptide in C1 is T-10-14-2-15-6-8-12-13-11. Considered together with the alignment of tryptic peptides in C2, T9 must follow T11 by exclusion.

The postion of T9 was confirmed by isolation and characterization of a hydroxylamine cleavage product, C1-HA4, from C1. A Sephadex G-50 SF chromatography (supplementary material) of a hydroxylamine-treated C1 yielded several fragments, including one which eluted in a position corresponding to 1200 molecular weight (C1-HA4). Automated sequence analysis of C1-HA4 confirmed the peptide alignment of T11-9. The entire amino acid sequence of α1(III)-CB8 was then confirmed by determining the internal sequences of the individual tryptic peptides as summarized in Figure 2.

Amino Acid Sequence of α1(III)-CB4. The peptide α1(III)-CB4 was obtained by CM-cellulose chromatography at pH 3.8 (supplementary material) followed by CM-cellulose chromatography at pH 4.8 (supplementary material). Its amino acid composition is presented in Table III and an automated amino acid sequence determination provided identification of the initial 50 residues (Figure 3). Attempts to isolate all of the tryptic peptides by phosphocellulose chromatography were unsuccessful due to the large number of peptides having nearly equal charge to mass ratios. However, eight of the smaller peptides (T1–T8) were isolated by automatic peptide analysis on PA-35 resin (supplementary material). Maleylated α1(III)-CB4 with subsequent cleavage at each Arg residue yielded three peaks having approximate molecular weights of 8000, 2500, and 1000 by Sephadex G-50 SF gel filtration (supplementary material). The two larger peaks, MT1 and MT2, were later found to be homogeneous peptides. Their amino acid composition is presented in Table III. The smaller ultraviolet absorbing peak was separated into two components

TABLE III: Amino Acid Composition of α1(III)-CB4, Tryptic, and Thermolysin Peptides of Type III Collagen from Human Cirrhotic Liver.[a]

amino acid	α1-(III)-CB4	T1	T2	T3	T4	T5	T6	T7	T8	MT1	MT2	MT3	MT4	Th1	Th2	Th3	total[b]
Hyp	18.8	2.4	2.1	0.8	0.7	1.6		0.6	0.9	11.7	3.7	1.6	1.5	6.7	7.6	3.9	20
	(19)	(3)	(2)	(1)	(1)	(2)		(1)	(1)	(12)	(4)	(2)	(2)	(7)	(8)	(4)	
Asp	9.1	0.3	1.0	0.2	1.2		1.3	1.7		6.2	0.3		3.1	6.2	3.2	0.2	9
	(9)		(1)		(1)		(1)	(2)		(6)			(3)	(6)	(3)		
Thr	4.8				0.2					3.6	1.2		0.4	3.3	0.8		5
	(5)									(4)	(1)			(3)	(1)		
Ser	0.3									0.3	0.2			0.2			
Glu	11.4	1.2		0.8	1.3	0.2	1.0	0.3	0.2	8.3	2.2	0.2	1.2	4.7	4.3	0.9	11
	(11)	(1)		(1)	(1)		(1)			(8)	(2)		(1)	(5)	(4)	(1)	
Pro	17.1	2.2	0.2	2.4	0.2	1.3		0.3	1.4	10.1	4.4	1.4	1.4	9.4	3.2	4.1	16
	(17)	(2)		(2)		(1)			(1)	(10)	(4)	(1)	(1)	(9)	(3)	(4)	
Gly	54.8	7.4	4.8	5.7	3.0	4.1	1.2	2.8	2.0	34.3	13.4	4.3	6.0	23.8	18.4	12.8	57
	(55)	(7)	(5)	(6)	(3)	(4)	(1)	(3)	(2)	(34)	(13)	(4)	(6)	(24)	(18)	(13)	
Ala	12.5	0.2	3.1	1.2	2.1	2.0	0.2	0.9		5.3	4.6	2.2	1.1	1.2	4.9	4.8	13
	(13)		(3)	(1)	(2)	(2)		(1)		(5)	(5)	(2)	(1)	(1)	(5)	(5)	
Val																	
Ile	0.2									0.2					0.2		
Leu	4.6		0.2			1.6				1.7	1.1	1.7		0.3	1.9	0.6	5
	(5)					(2)				(2)	(1)	(2)			(2)	(1)	
Phe	0.9								0.8				0.7	0.9			1
	(1)								(1)				(1)	(1)			
Hyl[c]	0.7		0.2						0.4	0.5			0.4	0.3			
Lys[c]	7.4	1.1	0.8	1.0	0.2		0.2	0.7	0.6	4.3	0.9		1.7	3.8	2.7	0.8	8
	(8)	(1)	(1)	(1)				(1)	(1)	(5)	(1)		(2)	(4)	(3)	(1)	
Arg	3.4				0.9	1.0	1.1	0.2		1.1		1.1	1.0	1.2	1.1	0.9	3
	(3)				(1)	(1)	(1)			(1)		(1)	(1)	(1)	(1)	(1)	
Hse[d]	1.0										1.1						1
	(1)										(1)						
total	149	14	12	12	9	12	4	8	6	85	33	12	18	61	48	30	149

[a] Values expressed as residues per peptide. No entry indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values. [b] Total of MT1, MT2, MT3, and MT4. [c] Partial lysine hydroxylation was noted. [d] Includes homoserine lactone.

by phosphocellulose chromatography, supplementary material (MT3 and MT4). Their amino acid composition is also shown in Table III.

The amino acid composition of MT4 was consistent with it being the 18 residue NH_2-terminal maleylated tryptic peptide. Automated Edman degradation of MT4 through the COOH terminus yielded an identical sequence as the degradation of intact α1(III)-CB4 establishing its position at the NH_2 terminus (Figure 3). MT2 contained the single residue of Hse and, therefore, must be the COOH terminus. Sequential degradation established its internal sequence. A comparison of the amino acid sequence of intact α1(III)-CB4 through the initial 50 residues with the sequence data obtained for MT4 (Figure 3) established the position of MT1 to follow MT4. By exclusion, then, MT3 must follow MT1. In summary, the alignment of the maleylated tryptic peptides was deduced to be MT4-1-3-2.

Thermolytic cleavage of α1(III)-CB4 yielded three major peptides which eluted from Sephadex G-50 SF in the positions corresponding to 6000, 4500, and 2500 mol wt (Th1, Th2, and Th3, respectively); see supplementary material. The amino acid composition of Th1 was consistent with the amino acid sequence presented for the initial 61 residues of α1(III)-CB4 and indicated Gly-Leu (residues 61–62) as a major cleavage site for thermolysin. This was confirmed by amino acid sequence analysis of Th2 which began with Leu-Gln-Gly-Leu-Hyp, the second Leu being followed by Hyp and hence apparently resistant to thermolytic hydrolysis. Further sequence analysis of Th2 yielded identification of residues 62–108. Similarly, the NH_2-thermal sequence of Th3 was identified as Leu-Arg (pos 116, 117) followed by the previously established sequence of MT2 (Figure 3).

Several futile attempts were made to identify positions 110–115. Reexamination of the tryptic peptides previously isolated confirmed the presence of a 12 residue Arg-containing peptide (T5, residues 106–117) not previously positioned in the sequence analysis of the other peptides. Its amino acid sequence (Figure 3) confirmed the presence of two Leu residues, the latter being followed by a COOH terminal Arg. This tryptic peptide, therefore, represents the connecting link between Th2 and Th3 (pos 110–115) and provides the necessary information on complete amino acid sequence of α1(III)-CB3.

The proposed alignment of these type III collagen peptides with the corresponding region of α1(I) was determined by comparison of amino acid sequence homology using the various possible positions of these peptides. The following analysis was made without glycine since, in each collagen chain, glycine occurs at every third residue. The 12-residue peptide, α1(III)-CB1, was located as the amino-terminal peptide primarily because of its similarity with the corresponding sequence of the α1(I) chain. Five of the eight residues, excluding glycine, were identical. Additionally, two large hydropholic residues, leucine and isoleucine, were homologous with positions having isoleucine and phenylalanine, respectively, in the α1(I) chain.

Two of the remaining four peptides also begin with glycine and, hence, are possible candidates for the amino terminal peptide of this collagen segment. The α1(III)-CB8, a 125 amino acid peptide, was selected as the second peptide and has

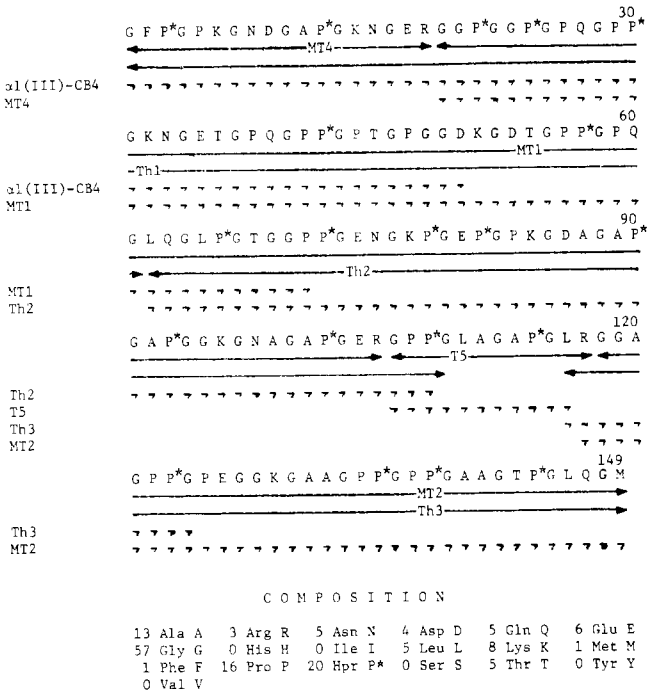

FIGURE 3: The complete amino acid sequence of α1(III)-CB4 of human liver type III collagen. The trypsin and thermolysin-derived peptides isolated are indicated by long arrows (↔). Short arrows (→) indicate the extent of Edman degradation of each peptide.

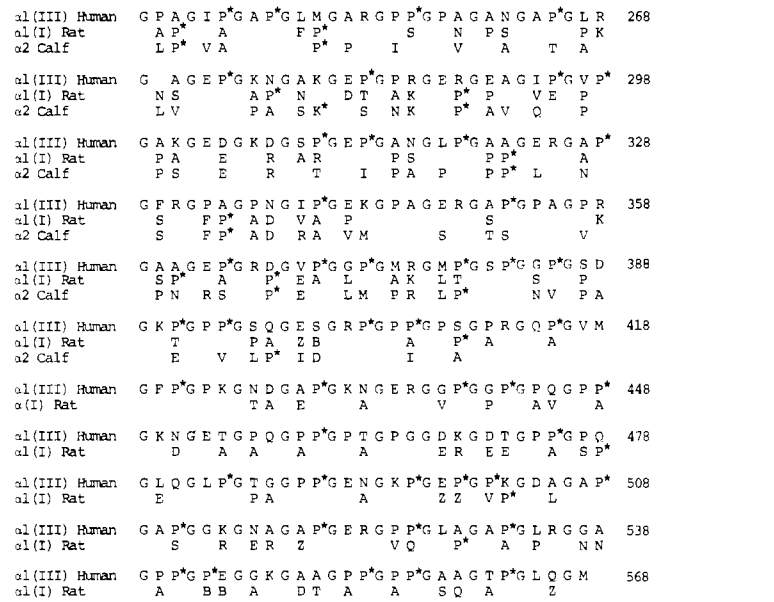

FIGURE 4: Comparison of the amino acid sequence of residues 239–567 of pepsin-solubilized human liver type III collagen with the identical regions from rat α1 and known calf α2 collagen chains. α1(III)-CB1-8-10-2 represents the carboxyl-terminal portion of α1(I)-CB8, while α1(III)-CB4 corresponds to the homologous peptide α1(I)-CB3. The same region corresponds to the carboxyl-terminal portion of α2-CB4, all of α2-CB2, and the known NH_2-terminal sequence of α2-CB3. Identical residues in the same position of each chain are indicated by blank spaces except for residues 410–567 in the calf α2 collagen chain which has not been determined.

a similar sequence with the corresponding position of the α1(I) collagen chain. With this alignment, 49 of the possible 83 amino acids were identical. Should the α1(III)-CB8 be located as the first peptide instead of α1(III)-CB1, only 11 of a possible 83 amino acids were similar. Alternately, if α1(III)-CB4, the remaining peptide beginning with glycine, were positioned as the initial peptide in place of α1(III)-CB1 and α1(III)-CB8, only 17 of 99 amino acids would be identical with the α1(I) chain compared with 55 of 99 residues identical when α1(III)-CB4 is located at the COOH-terminal peptide of this segment. The remaining possibility would be to switch α1(III)-CB4 with α1(III)-CB8. If α1(III)-CB4 followed α1(III)-CB1, the level of identity would be only 15 of 99 residues.

The remaining two peptides, α1(III)-CB10 and α1(III)-CB2, contain 3 and 40 amino acids, respectively. Both begin with a Y amino acid of the traditional Gly-X-Y triplet, thus requiring a methionine in the X position. The tripeptide, α1(III)-CB10, contains too few amino acids to allow comparison, but, with the proposed alignment of CB10-2, the level of identity was 23 of 28 residues; with a switch in positions with α1(III)-CB2 preceding α1(III)-CB10, the level of identity would be reduced to only 12 of 28.

Considering, therefore, the identity level of these five peptides, the best possible alignment must be α1(III)-CB1-8-10-2-4. The peptide segment α1(III)-CB1-8-10-2, therefore, corresponds to the COOH-terminal region of α1(I)-CB8 and α1(III)-CB4 with α1(I)-CB3. The latter two CNBr peptides each have 149 residues. The amino-terminal portion of α1(I)-CB8 has previously been shown to be homologous with α1(III)-CB6 (Seyer & Kang, 1977), thereby making the segment of type III collagen containing CB6-1-8-10-2 homologous with the 279 residue peptide, α1(I)-CB8.

Discussion

Type III collagen was obtained by limited pepsin digestion of human cirrhotic liver. Ten distinct peptides were obtained by CNBr digestion of the α1(III) collagen chains. Three peptides from the NH_2-terminal portion of the molecule (7-3-6) have previously been characterized and their complete amino acid sequence determined (Seyer & Kang, 1977). In the present report, five additional CNBr peptides were isolated by gel filtration and ion-exchange chromatography. Their estimated molecular weights were as follows: 300 for α1(III)-CB10, 1000 for α1(III)-CB1, 4000 for α1(III)-CB2, 13 500 for α1(III)-CB4, and 12 000 for α1(III)-CB8. Together they represent 329 amino acids or nearly one-third of the type III collagen molecule and are summarized in Figure 4.

The smallest peptide, which we have tentatively designated as α1(III)-CB10, has not been previously identified in studies with type III collagen, presumably because of its small molecular weight. Its amino acid sequence was identified as Arg-Gly-Hse and tentatively positioned at residues 376–378 for reasons cited below.

The amino acid sequence of the 149 residue peptide α1(III)-CB4 was determined utilizing automated protein sequence methods combined with thermolytic and tryptic digestion, the latter before and after maleylation. Sufficient sequence information was obtained to establish overlapping peptides needed for the complete amino acid sequence. Homology of the NH_2 terminus of this peptide, α1(III)-CB4, with α1(I)-CB3, has previously been established with type III collagen of calf aorta (Fietzek & Kuhn, 1976). The data presented in Figure 4 provide support that both the peptides contain 149 residues and that they are indeed homologous in both size and sequence.

The amino acid sequences of the three remaining peptides, α1(III)-CB1, α1(III)-CB2, and α1(III)-CB8, were determined, the former two with intact peptides to their penultimate residues and the latter by a combination of tryptic, chymotryptic and hydroxylamine cleavage (Figure 6). No peptides of these particular sizes and sequences exist in type I collagen,

and, therefore, their position in the collagen chain was more difficult to establish. However, α1(III)-CB5 has previously been suggested as being homologous with α1(I)-CB7 (Fietzek & Kuhn, 1976; Miller et al., 1976) and α1(III)-CB9 with the COOH terminus of α1(III) (Chung et al., 1974). Together these latter two peptides, α1(III)-CB5 and CB9, contain approximately 450 amino acids and represent the COOH half of the α1(III) collagen chain. The only remaining portion of the type III molecule in which α1(III)-CB1, 2, 8, and 10 peptides may exist is, therefore, in the region occupied by α1(I)-CB8 (279 residues) in the α1(I) collagen chain. Previous sequence analysis of α1(III)-CB6 (Seyer & Kang, 1977) suggested this 99 residue peptide to be homologous with the NH_2 terminus of α1(I)-CB8 leaving 180 amino acids undetermined. A comparison of the position of the basic amino acids, lysine and arginine, in the peptides analyzed here with those remaining 180 residues of α1(I)-CB8 suggests the most likely peptide alignment to be α1(III)-CB1-8-10-2. This alignment would also provide the fewest number of amino acid substitutions between the α1(I) and α1(III) collagen chains. The tripeptide, α1(III)-CB10, was located to follow α1(III)-CB8, since this alignment maximizes the homology between α1(III) and α1(I). Furthermore the overlap of α1(III)-CB5-9 has been identified (Miller et al., 1976) and the COOH-terminal sequence of α1(III)-CB9 established (Fietzek et al., 1977). No other positions are available in which this tripeptide could be inserted.

The primary structure of the segment of the α1(III) collagen chain as obtained in this study remains consistent with several principles previously found in studies with other collagen chains (Hulmes et al., 1973; Gallop & Pas, 1975; Fietzek & Kuhn, 1976; Piez, 1976). Glycine is present at every third residue in the Gly-X-Y triplet sequence, and eight Gly-Gly-Y sequences were found (in agreement with the higher quantities of glycine in type III collagen). In four of these instances, Hyp was present in the Y position while two sequences each of Gly-Gly-Lys and Gly-Gly-Ala were identified. A similar Gly-Gly-Lys sequence had been found previously in α1(III)-CB6 (Seyer & Kang, 1977). Two triplets, in contrast, contained Gly in the Y position (residues 466, 489) suggesting glycine to be acceptable in any position in the primary structure of the type III collagen. Interestingly, only one Gly-Gly sequence exists in the α1(I) chain (Bornstein, 1969), which in the case of α1(III) is altered with a Gly → Ala substitution in the X position (pos 327) in α1(III). Proline in the Y position was always totally or partially hydroxylated as 4-Hyp. Lysyl residues in the Y positions were generally not hydroxylated to any significant amount, the only exception being residue 6 of α1(III)-CB4, which was approximately 40% hydroxylated.

The present data also allow a comparison of the sequence of the homologous region of α1(I) from rat (Bornstein, 1969; Butler, 1969; Butler & Ponds, 1971). Thirty-two percent of the amino acid residues were substituted when comparing the collagen chains. Most of the amino acid substitutions were conservative with the changes involving neutral amino acids. A total of 5 neutral amino acid residues were substituted for basic residues, 5 acidic for neutral residues, 5 neutral for acidic residues, 2 basic for neutral residues, and 1 basic for acidic residue, thus, giving an overall net charge increase of 1 basic residue within the entire 329 amino acid sequence presented here. Comparison of human α1(III)-CB4 with the 31 NH_2 terminal residues of bovine aorta α1(III)-CB4 (Fietzek & Kuhn, 1976) indicated only one substitution (Hyp → Ala) at position 448 indicating very little interspecies variation between human and calf type III collagen.

In summary, the amino acid sequence of 566 residues of human type III collagen has thus far been elucidated. A comparison of this segment of the α1(III) collagen chain with either α1(I) or α2 from type I collagen suggests a large amount of homology. The differences between α1(III) and α1(I) are similar to those established between α1(I) and α2 and α1(III) and α2, thereby leaving no clue as to any possible preferential sequence relationship among the three chains. More charged amino acid differences were noted between this segment of α1(III) and α(I) or α2 than normally found with comparisons of α(I) and α2. This, theoretically, may explain the difference in banding pattern between SLS aggregates of type III collagen and type I collagen when examined by electron microscopy after staining.

Acknowledgments

The authors are grateful to Dr. J. M. Young of this Veterans Administration Hospital for pathologic examinations of tissue samples and for valuable discussions. We also acknowledge the excellent technical assistance of T. Marshall, P. Hill, and E. Rowland.

Supplementary Material Available

Sixteen figures (isolation and separation of peptides by various chromatographic methods) and one table (quantitation of Pth derivatives after automatic Edman degradation of peptides from human liver type III collagen) containing additional data as mentioned throughout text (18 pages). Ordering information is available on any current masthead page.

References

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

Bornstein, P. (1969) *Biochemistry 8,* 63.

Butler, P. J. G., Harris, J. T., Hartley, B. S., & Leberman, R. (1969) *Biochem. J. 112,* 679.

Butler, W. T. (1969) *Biochemistry 8,* 44.

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

Chung, E., & Miller, E. J. (1974) *Science 183,* 1200.

Chung, E., Keele, E. M., & Miller, E. J. (1974) *Biochemistry 13,* 3459.

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

Dixit, S. D., Seyer, J. M., & Kang, A. H. (1977a) *Eur. J. Biochem. 81,* 599.

Dixit, S. D., Seyer, J. M., & Kang, A. H. (1977b) *Eur. J. Biochem 73,* 213.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1,* 80.

Epstein, E. H., Jr. (1974) *J. Biol. Chem. 249,* 3225.

Fietzek, P. P., & Kuhn, K. (1976) *Int. Rev. Conn. Tiss. Res. 7,* 1.

Fietzek, P. P., & Rautherberg, J. (1975) *FEBS Lett. 49,* 365.

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

Fietzek, P. P., Allmann, H., Rauterberg, J., & Wachter, E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 84.

Foster, J. A., Bruenger, C. L., Hu, C. L., Albertson, K., & Franzblau, C. (1973) *Biochem. Biophys. Res. Commun. 53,* 70.

Gallop, P. M., & Paz, M. A. (1975) *Physiol. Rev. 55,* 418.

Gallop, P. M., Blumenfield, O. O., & Seifter, S. (1972) *Annu. Rev. Biochem. 41,* 617.

Hulmes, D. J. S., Miller, A., Parry, D. A., Piez, K. A., & Woodhead-Galloway, J. (1973) *J. Mol. Biol. 70,* 137.

Kang, A. H. (1972) *Biochemistry 11,* 1828.
Kang, A. H., & Gross, J. (1970) *Biochemistry 9,* 796.
Miller, E. J., & Lunde, L. G. (1973) *Biochemistry 12,* 3153.
Miller, E. J., Harris, E. D., Chung, E., Finch, E. J., McCroskery, P. A., & Butler, W. T. (1976) *Biochemistry 15,* 787.
Piez, K. A. (1968) *Anal. Biochem. 26,* 305.
Piez, K. A. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N., & Redd, A. H., Eds.) Plenum Press, New York, N.Y.
Pisano, J. J., & Bronzert, T. J. (1969) *J. Biol. Chem. 244,* 5597.
Seyer, J. M., & Kang, A. H. (1977) *Biochemistry 16,* 1158.
Seyer, J. M., Hutcheson, E. T., & Kang, A. H. (1976) *J. Clin. Invest. 57,* 1498.
Seyer, J. M., Hutcheson, E. T., & Kang, A. H. (1977) *J. Clin. Invest. 59,* 241.
Traub, W., & Piez, K. A. (1971) *Adv. Protein Chem. 25,* 243.
Zimmerman, C. L., Pisano, J. J., & Appella, E. (1973) *Biochem. Biophys. Res. Commun. 55,* 1220.

Structural Studies on Rabbit Transferrin: Isolation and Characterization of the Glycopeptides†

Dudley K. Strickland and Billy G. Hudson*

ABSTRACT: The structure of rabbit transferrin was investigated with regard to number, size, and composition of the heteropolysaccharide units and their relative location on the polypeptide chain. The composition and molecular weight of the Pronase glycopeptides revealed that rabbit transferrin contains two heteropolysaccharide units, each composed of 2 sialic acid residues, 2 galactose residues, 3 mannose residues, and 4 *N*-acetylglucosamine residues. The composition and molecular weight of the tryptic glycopeptides further substantiated the existence of two identical heteropolysaccharide units and revealed that both units have identical amino acid residues in the immediate vicinity of the carbohydrate attachment sites to the polypeptide chain, suggesting a sequence homology surrounding the two glycosylation sites. Characterization of the cyanogen bromide fragments from rabbit transferrin indicated that both heteropolysaccharide units are located within a single polypeptide fragment representing approximately one-third of the molecule.

The transferrins represent a group of iron binding proteins that play a vital role in iron metabolism (Aisen & Brown, 1977). This role includes binding, transport, and transfer of iron to a variety of tissues such as the erythrocyte precursor where this iron is utilized in hemoglobin biosynthesis. The mechanism of iron transfer from rabbit transferrin to the rabbit reticulocyte has been extensively studied (Aisen & Brown, 1977; Morgan, 1974). The initial event involves binding of transferrin to specific receptors located on the reticulocyte surface (Jandl et al., 1959; Jandl & Katz, 1963; Baker & Morgan, 1971). These receptors have recently been solubilized from the rabbit reticulocyte (Leibman & Aisen, 1977) and from the differentiated Friend erythroleukemic cell (Hu et al., 1977). The mechanism of subsequent release of iron from transferrin to the reticulocyte is at this time poorly understood.

The critical importance of species variability in assessing the mechanism of iron transfer from transferrin to the reticulocyte was recently demonstrated (Harris & Aisen, 1975a). In the homologous system of rabbit transferrin and rabbit reticulocytes (Harris & Aisen, 1975a) or human transferrin and human reticulocytes (Harris & Aisen, 1975b), the two iron binding sites of transferrin are functionally equivalent. However, in the heterologous system involving human transferrin and rabbit reticulocytes, the sites are functionally nonidentical, possibly indicating an artifact of the heterologous system (Harris & Aisen, 1975a).

Ultimately, the elucidation of the detailed mechanism of iron transfer, using the homologous system of rabbit transferrin and rabbit reticulocytes, will require a detailed knowledge of the chemical structure of rabbit transferrin. At present the amino acid and carbohydrate composition (Hudson et al., 1973), molecular weight and single chain nature (Hudson et al., 1973; Palmour & Sutton, 1971; Green & Feeney, 1968), iron binding properties (Hudson et al., 1973) and preliminary crystallographic data (Al-Hilal et al., 1976) have been reported for rabbit transferrin. The purpose of this investigation was to determine the number, composition and size of the heteropolysaccharide units of rabbit transferrin and to define their relative location on the polypeptide chain. The results of this study show that rabbit transferrin contains two heteropolysaccharide units that are identical with respect to size and monosaccharide composition. In addition, each carbohydrate unit contains identical amino acid residues in the immediate vicinity of the two glycosylation sites, and, finally, both units are located in a single polypeptide segment representing about one-third of the intact transferrin molecule. These results provide further support for the existence of sequence homology within the transferrin molecule.

Experimental Procedures

Materials

Urea was purchased from Aldrich and was recrystallized and deionized prior to use. Ethylenimine and cyanogen bro-

† From the Department of Biochemistry, University of Kansas Medical Center, Kansas City, Kansas 66103. *Received March 7, 1978.* This work was supported by the Herman Frasch Foundation and by a grant from the National Institutes of Health (AM 18381).

0006-2960/78/0417-3411$01.00/0 © 1978 American Chemical Society