Physical Properties of the Amino-Terminal Precursor-Specific Portion of Type I Procollagen†

Jürgen Engel,* Peter Bruckner, Udo Becker,‡ Rupert Timpl, and Bea Rutschmann

ABSTRACT: Fragments from the amino end of the pα1(I) chain of sheep procollagen were isolated after digestion with cyanogen bromide (peptide CB 0.1) or collagenase (peptide Col 1). Peptide CB 0.1 contains the whole precursor-specific segment which consists mainly of a collagenous and a noncollagenous domain. It is primarily the noncollagenous portion which is retained in Col 1. Physical properties of these fragments were studied by circular dichroism and analytical ultracentrifugation. Peptide Col 1 showed no tendency to aggregate in solution whereas peptide CB 0.1 formed trimers at low temperature. This association occurred in the collagenous region and gave rise to a triple-helical structure. The noncollagenous portion in CB 0.1 had no or only little effect on the coil to triple-helix conversion. Peptide Col 1 showed a high frictional coefficient indicating an elongated shape. It resembled in its CD spectrum proteins containing aperiodic structural elements and β structure. Reduction of the five disulfide bridges in Col 1 caused unfolding of the fragment. Only small and reversible changes were induced in native Col 1 by guanidine hydrochloride, dodecyl sulfate, LiCl, elevated temperature, and extreme pH. High concentrations of trifluoroethanol produced a sharp transition from the native structure of Col 1 to a conformation with a considerable proportion of α helix. This type of conformational change was more readily achieved with reduced Col 1. A loss of native conformation was also obtained by tryptic cleavages of bonds in two disulfide loop regions of Col 1. Proteolytic removal of about 30% of the peptide sequences from the amino and carboxyl end of Col 1, however, produced little change in the CD spectrum, indicating the presence of randomly coiled structures in the terminal portions of the molecule.

Procollagen, the precursor form of collagen, contains additional peptide segments not found in collagen at both ends of the three polypeptide chains that comprise the molecule (Bornstein, 1974; Martin et al., 1975; Prockop et al., 1976; Tanzer et al., 1974; Byers et al., 1975; Fessler et al., 1975; Olsen et al., 1976). Intermediate forms of procollagen containing only the amino or only the carboxyl precursor-specific peptides have been identified. It is commonly referred to such intermediate forms as p collagens and to the constituent chains as pα chains (Martin et al., 1975). In dermatosparactic animals, the intermediate with the amino-terminal extension accumulates due to the absence of the enzyme in this disorder which cleaves this peptide (Kohn et al., 1974). Type I p collagen from dermatosparactic animals has been the major source of the amino-terminal peptide for chemical studies (Furthmayr et al., 1972; Becker et al., 1976). Little is known about the carboxyl-terminal segments of procollagen.

We have previously isolated a peptide (pα1-CB 0.1)[1] of about 165 amino acid residues from the pα1(I) chain that included the entire precursor-specific segment plus 19 amino acid residues from the amino end of the collagen α1(I) chain (Becker et al., 1976). Unexpectedly it was found that this peptide contained 8 residues of hydroxyproline, suggesting that the peptide contained a collagen-like part. By digestion of pα1-CB 0.1 or of the entire pα1(I) chain with bacterial collagenase, we obtained a peptide pα1-CB 0.1 Col 1 which contained about 100 amino acid residues and lacked hydroxyproline. Further studies established that the collagenase-resistant domain contained five intrachain disulfide bridges and was located at the amino end of the precursor-specific segment. The following 40 to 50 amino acids contained hydroxyproline as well as sufficient glycine to account for glycine in every third position. This part of the extension therefore strongly resembled the type of sequence found in the triple-helical portion of the collagen α chains. Finally, the peptide contained a short non-triple-helical segment connecting it with the α1(I) chain (Fietzek and Kühn, 1976).

The observation of a noncollagenous and a collagenous region in the amino-terminal segment of procollagen raised the question whether these regions are folded into separate domains with different conformations. The present work was undertaken to study this question by circular dichroism and analytical ultracentrifugation.

Materials and Methods

Materials. Dermatosparactic sheep skin which was used to isolate procollagen peptides was kindly supplied by Dr. O. Helle, University of Oslo. Bacterial collagenase (grade CLSPA) and trypsin treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone (Tos-PheCH$_2$Cl-trypsin) were obtained from Worthington. Protease from *Staphylococcus aureus* strain V8 was obtained from Miles Laboratories. All other reagents were of analytical grade (Merck, Darmstadt).

† From the Department of Biophysical Chemistry, Biozentrum der Universität Basel, Switzerland, and the Max-Planck-Institut für Biochemie, Martinsried, Munich, West Germany. *Received January 12, 1977.* These studies were supported by Grant Ti95/2 from the Deutsche Forschungsgemeinschaft.

‡ Present address: Behring Werke, Marburg/Lahn, West Germany.

[1] Abbreviations used: CB 0.1 and Col 1, fragments from the amino end of dermatosparactic sheep procollagen pα1(I) chain derived by cyanogen bromide treatment and collagenase treatment, respectively [the same peptides were designated pα1 CB 0.1 and pα1-CB 0.1 by Becker et al. (1976)]; Col 1 DI and CB 0.1 DI, peptide Col 1 and peptide pα1-CB 0.1 completely reduced with dithioerythritol and S-carboxymethylated; Col 1 T and Col 1 SP, core fragments of Col 1 after treatment with trypsin and staphylococcal protease, respectively; CD, circular dichroism; Tos-PheCH$_2$Cl, L-1-tosylamido-2-phenylethyl chloromethyl ketone; Tris, tris(hydroxymethyl)aminomethane; DEAE, diethylaminoethyl; Gdn-HCl, guanidine hydrochloride; UV, ultraviolet; NMR, nuclear magnetic resonance; NaDodSO$_4$, sodium dodecyl sulfate.

TABLE I: Amino Acid Composition of Col 1 of Its Core Proteins Obtained by Treatment with Trypsin (Col 1 T) and with Staphylococcal Protease (Col 1 SP) and of the Individual Chain Fragments Produced from the Core Proteins by Reduction and Alkylation.[a]

			Chain fragments in Col 1 T				Chain fragments in Col 1 SP		
	Col 1[b]	Col 1 T	T 1	T 3-4	T 5	Col 1 SP	SP 1	SP 2	SP 3
Cys	10	10	1[c]	6[c]	3[c]	10	6[c]	1[c]	3[c]
Asp	14	12	2	8	2	13	11	3	
Thr	6	5	1		5	3	1	1	
Ser	2	2			1				
Glu	20	20	10	3	8	8	5	1	1
Pro	9	9	2	3	3	8	4	2	2
Gly	9	7	3	1	3	4	3		
Ala	1	1		1		1		1	
Val	9	9	2	5	3	8	7	1	1
Ile	2	2	1	1		2	2		
Leu	3	3	1	2		3	2	1	
Tyr	1					1	1		
His	1					1	1		
Lys	4	4		3	1	3	1	2	
Arg	3	1	1			2	2		
Trp	1	1		1		ND[d]	1		
Total	95	86	24	34	29	67	47	13	7

[a] Data are given in residues per peptide chain rounded to the nearest whole number. If no value is given, this denotes less than 0.2 residue. [b] Taken from Furthmayr et al. (1972) and Becker et al. (1976). [c] As carboxymethylcysteine. [d] ND, not determined.

Preparation of Precursor-Specific Segments from the Amino End of Procollagen Pα1(I) Chain. The procollagen peptides were isolated after cleavage with cyanogen bromide (peptide CB 0.1) or with bacterial collagenase (peptide Col 1) and purified as described previously (Becker et al., 1976). In this procedure the materials were exposed to 6–8 M urea. In order to reform a triple-helical conformation in the collagenous portion of peptide CB 0.1, it was dissolved in 8 M urea (1–2 mg/mL) and dialyzed against 0.05 M sodium phosphate buffer, pH 7.0, during 4 days at 6–8 °C.

In addition we prepared Col 1 by a method which avoided the use of urea or other denaturants. Collagen was extracted from the skin of a dermatosparactic sheep with 1 M sodium chloride solution and precipitated by dialysis against 0.02 M sodium phosphate buffer, pH 8.6, containing 15% potassium chloride. The precipitate was dissolved in 0.05 M Tris-HCl buffer, pH 7.4, containing 0.5 M sodium chloride and 5 mM calcium chloride and treated with collagenase at 37 °C for 4 h. The enzyme substrate ratio was 1:100. Small peptides were removed by dialysis against 0.1 M acetic acid and Col 1 was purified by chromatography on a Bio-Gel P4 column (1.5 × 105 cm) with 0.1 M acetic acid at the eluent. This was followed by chromatography on phosphocellulose (column size 1 × 8 cm) in 1 mM sodium acetate buffer, pH 3.6, with a linear gradient from 0 to 0.3 M sodium chloride (100/100 mL). Details of these procedures were described by Furthmayr et al. (1972).

Reduction and Alkylation. Peptides (5 mg) were dissolved in 1 mL of 0.08 M Tris-HCl buffer, pH 8, containing 8 M urea and 0.02 M dithioerythritol or in 1 mL of 0.1 M sodium phosphate buffer, pH 8, with the same amount of dithioerythritol. The reaction was allowed to proceed at 37 °C for 4 h. After cooling to room temperature, 16.7 mg of sodium iodoacetate was added. After 1 h the reagents were removed by gel filtration on a Bio-Gel P2 column with 0.1 M formic acid as the eluent. The amount of carboxymethylcysteine was determined by amino acid analysis. A complete alkylation of all five disulfide bridges in Col 1 was observed irrespectively whether reduction was carried out in the presence or absence of urea. This material is designated Col 1 DI.

Proteolytic Modification of Peptide Col 1. For the preparation of protease-resistant cores from Col 1, 20 mg of native peptide was dissolved in a 2 mL 0.2 M ammonium bicarbonate solution, pH 7.9, and was incubated with 0.2 mg of Tos-PheCH$_2$Cl-trypsin at 37 °C for 24 h. Another sample of the peptide was treated with staphylococcal protease at an enzyme–substrate ratio of 1:50 for 20 h at 37 °C. This enzyme is known to cleave mainly at the carboxyl site of glutamate residues (Houmard and Drapeau, 1972). Both digests were then separated on a Bio-Gel P10 column (1.5 × 105 cm) equilibrated in 0.2 M ammonium bicarbonate, pH 8.5. In each case a single large fragment and several small fragments were obtained. The large fragments were desalted by repeated lyophilization. Examination by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (Furthmayr and Timpl, 1971) showed that each fragment moved as a single band which had a higher mobility than the original peptide Col 1. Amino acid analysis of the trypsin-resistant core (Col 1 T) and the core peptide resistant to staphylococcal protease (Col 1 SP) showed that they are composed of about 86 and 67 amino acid residues, respectively (Table I). They contained all the cysteines found in peptide Col 1.

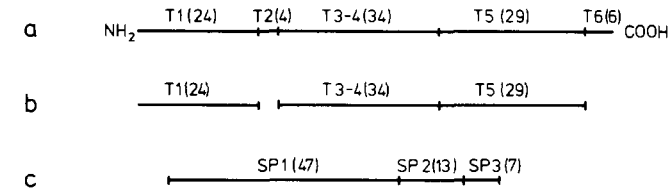

FIGURE 1: Order and size of the chain fragments comprising the core peptides Col 1 T (b) obtained by tryptic cleavage and Col 1 SP (c) obtained by digestion with staphylococcal protease. For comparison data are included on the tryptic peptides obtained from reduced Col 1 (a) (Rohde, Bruckner, Helle, and Timpl, in preparation). The three chain fragments of Col 1 T and Col 1 SP are held together by disulfide bridges. Bracketed numbers refer to the amino acid number of single peptides.

The structure of the protease-resistant core peptides is shown in Figure 1 and was determined by reduction and alkylation of the peptides and separation of the chain fragments either on DEAE-cellulose (Col 1 T) or on Bio-Gel P-10 (Col 1 SP). The size of the fragments was determined by amino acid analysis (see Table I) and their position within Col 1 assigned by comparison with fragments obtained from reduced and alkylated Col 1 treated either with trypsin or staphylococcal

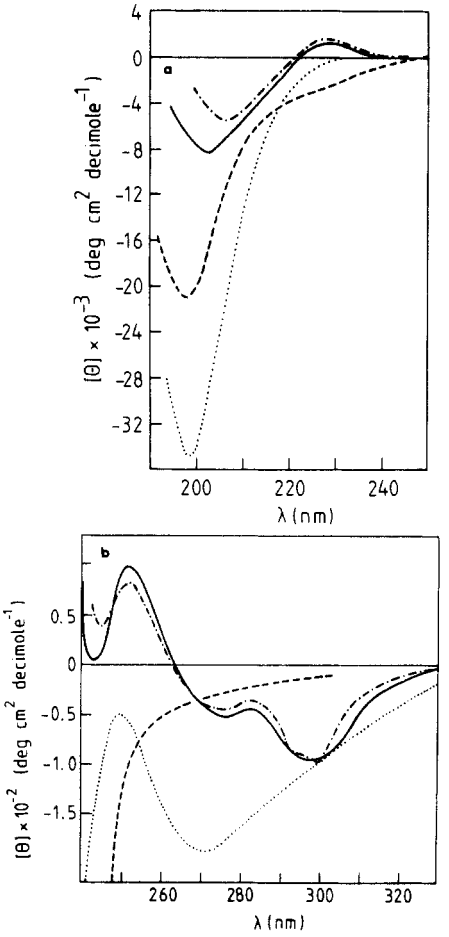

FIGURE 2: Circular dichroism (CD) spectra in the low (a) and near (b) UV region of the peptide Col 1 in its native form (—), after complete reduction and alkylation (Col 1 DI) (- - -) and after modification with trypsin (Col 1 T) (• • •) or with staphylococcal protease (Col 1 SP) (- • - •). All measurements were made in 0.025 mM sodium phosphate buffer, pH 7, at 20 °C. Molar ellipticity is expressed per amino acid residues.

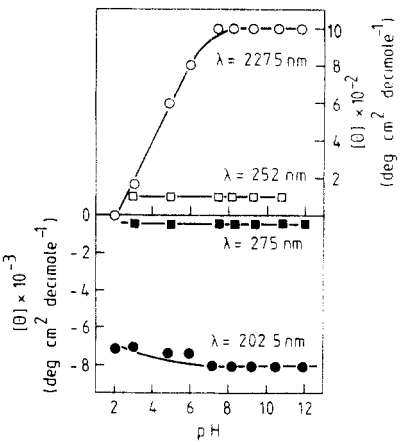

FIGURE 3: Variation of the CD signals of peptide Col 1 on pH. T = 20 °C.

protease (Rohde, Bruckner, Helle, and Timpl, in preparation). The data indicate that Col 1 T consists of three chain fragments of similar length held together by disulfide bridges. The chain fragment T3-4 contains the single tryptophan. Compared with Col 1 the core peptide Col 1 T was shorter by a hexapeptide T6 located at the carboxyl end of Col 1 and a tetrapeptide T2 containing the single histidine and tyrosine residue. Peptide Col 1 SP also consisted of three polypeptide chains and the three aromatic residues were located in fragment SP 1. As compared with Col 1, Col 1 SP lacked about 25 amino acid residues from the carboxyl end and about 6 amino residues from the amino end.

Determination of Protein Concentration. Aliquots of the solutions (0.01 to 0.05 mL) containing 1 to 5 nmol of peptide were hydrolyzed in 6 M hydrochloric acid at 110 °C for 24 h. The amino acid content of the hydrolyzates was determined on a Durrum D-500 amino acid analyzer.

Measurements of Circular Dichroism. A Cary 61 spectropolarimeter equipped with thermostated cells of 0.2, 1, and 10 mm path length (Perkin-Elmer) was employed. The pH dependence of CD spectra was measured in a rectangular cell of 1 cm path length. The pH of a solution containing 0.025 M sodium phosphate was adjusted to the desired value by addition of dilute NaOH or HCl and measured by a microelectrode in the cell before and after recording each spectrum. For measurements at high absorbance in concentrated solutions of LiCl or GdnHCl, an assembly cell (Beckman Instruments) with spacers of approximately 15 μm thickness was used in a thermostated cell holder. The exact path length was determined by calibration with an alkaline sodium bichromate solution after each measurement.

Analytical Ultracentrifugation. All measurements were performed at 20 °C with a Spinco Model E centrifuge (Beckman Instruments) equipped with a scanner. Sedimentation velocity runs were performed at 60 000 rpm with single-sector cells and equilibrium runs at 32 000 rpm with double-sector cells.

Results

CD Spectra of the Noncollagenous Segment Col 1 and of the Chemically and Proteolytically Produced Derivatives. The CD spectrum of the native extension protein Col 1 is compared in Figure 2 with that of the reduced and alkylated form of this protein and with the spectra of the core proteins Col 1 T and Col 1 SP obtained by digesting native Col 1 with trypsin or with staphylococcal protease, respectively. Completely identical spectra were obtained for Col 1 isolated in the presence or absence of urea (see Materials and Methods). This proves that 6 M urea does not irreversibly destroy the conformation of Col 1. The low UV part of the spectrum (Figure 2a) shows a shallow trough at 202.5 nm with $[\theta]_{min}$ of only -8200 deg cm^2 dmol^{-1} and a small positive maximum at 227.5 nm. In the near UV at least three distinct bands are observed (Figure 2b). Reduction and alkylation of the disulfide bridges in Col 1 changes the CD spectrum drastically. In the far UV a large minimum at 198 nm with $[\theta]_{min} = -21\,000$ deg cm^2 dmol^{-1} is observed, and in the near UV the spectrum is completely featureless. Again no difference in the CD spectra of materials reduced in the presence or absence of urea was observed.

The core protein Col 1 SP exhibits a similar spectrum to native Col 1, whereas that of Col 1 T is completely different. The latter shows a large negative trough at about 198 nm with $[\theta]_{min} = -35\,000$ deg cm^2 dmol^{-1}. No positive maximum is seen near 230 nm, and in the near UV only a single negative band near 270 nm shows up.

Dependence of the CD Spectrum of Col 1 on pH and Temperature. Figure 3 shows the pH dependence of the CD signals of the unmodified extension protein Col 1 at 275, 252, 227.5, and 202.5 nm. Only the band at 227.5 nm exhibits a significant pH dependence. The ellipticity at this wavelength decreases almost linearly below pH 8. No plateau region is observed at low pH. Moreover, the pH dependence is smooth and extends over a region of more than four pH units. It is therefore not indicative of a cooperative conformational transition of the

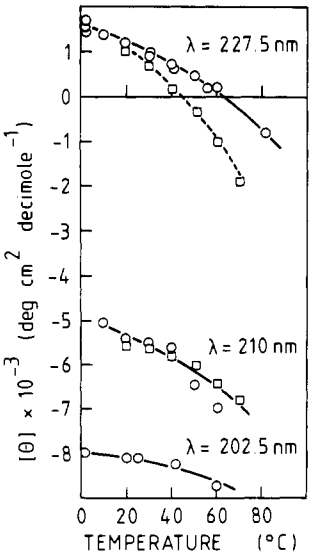

FIGURE 4: Temperature dependence of the CD signals of Col 1 at 227.5 nm (positive extremum), at 210 nm and at 202.5 nm (trough) in 0.025 M sodium phosphate buffer, pH 7, without (○) and with 4.8 M guanidine hydrochloride (□).

TABLE II: CD Signals of Peptide Col 1 in Various Solvents.

Solvent	T (°C)	$[\theta]_{227.5}$	λ_{min} (nm)	$[\theta]_{min}$
Buffer:	20	1200	202.5	−8200
0.025 M sodium	40	700	202.5	−8300
phosphate, pH 7	60	0	202.5	−8700
	80	−800	202.5	−9000
Sulfolan:buffer	21	600	202.5	−6000
2:1 (v/v)	40	0	202.5	−6500
	61	−830	202.5	−6800
4.8 M Gdn-HCl	20	1000		
in buffer	40	200		
	60	−1000		
	70	−1900		
6 M LiCl	20	0	~205	−6200
in buffer				
10% $NaDodSO_4$ in	20	+1400	202.5	−10 800
buffer	70	0	205	−11 200
20% $NaDodSO_4$ in	21	0	202.5	−12 700
buffer	70	0	205	−13 900

[a] $[\theta]$ in deg cm² dmol⁻¹ expressed per amino acid residue.

protein. Also the invariance of the size of the peptide band at 202.5 shows that the gross structure of the protein is unaltered over the pH range of 2 to 11. At pH <2 and pH >11 slow and irreversible changes occur which are probably due to hydrolysis of peptide bonds. Also the temperature dependence of the CD signals is relatively small (Figure 4) and fully reversible. There was no indication for a sharp temperature-induced conformational transition.

Conformation of Peptide Col 1 in the Presence of Denaturants. Because of the absence of any temperature or pH-induced conformational change, an attempt was made to unfold the structure of Col 1 by high concentrations of Gdn-HCl, $NaDodSO_4$, LiCl, and sulfolan (Table II). Surprisingly little changes were observed. In sulfolan/buffer 2:1 (v/v), the spectrum was essentially the same as in buffer. In 4.8 M Gdn-HCl, the spectrum could not be determined below 210

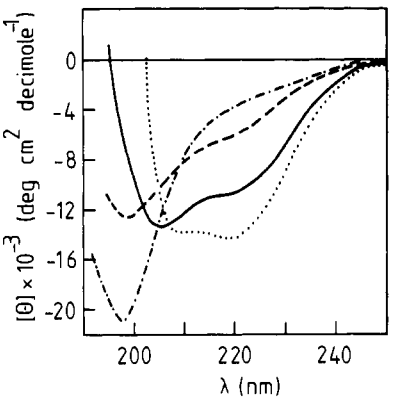

FIGURE 5: CD spectra of the completely reduced and alkylated peptide Col 1 DI in mixtures of water and trifluoroethanol at 20 °C. The volume fraction of trifluoroethanol are 0.3 (- - -), 0.5 (—), and 0.96 (• • •). The spectrum in 0.025 M phosphate buffer, pH 7, is shown for comparison (- • - •).

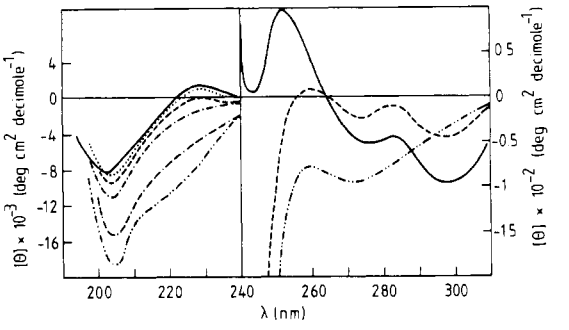

FIGURE 6: CD spectra of native peptide Col 1 in mixtures of water and trifluoroethanol at 20 °C. The volume fractions of trifluoroethanol are 0 (—), 0.5 (• • •), 0.7 (- - -), 0.8 (- • - •), 0.9 (- - -), and 0.96 (- • - •).

nm because of the high absorption of the reagent. Above 210 nm the spectrum was very similar to that measured in buffer. The temperature dependence at 227.5 nm is somewhat larger in Gdn-HCl than in buffer whereas the values at 210 nm are identical (see Figure 4). In 6 M LiCl the positive band at 227.5 nm disappeared but the negative band in the far UV did not change very much. In the presence of $NaDodSO_4$ the positive band disappeared at a detergent concentration of 20% and the minimum at 202.5 nm was somewhat more negative than in buffer. A small temperature dependence of the ellipticity around 202.5 nm was always in the direction of decreasing the amplitude of the negative band (see Table II). Dramatic changes could not be induced by any of the detergents and there are no indications for a transition to a different conformation.

Conformational Transitions Induced in Col 1 by Trifluoroethanol. When trifluoroethanol is added to solutions of Col 1 or Col 1 DI, dramatic conformational changes are induced in both proteins. For the reduced and alkylated chain (Col 1 DI), a completely new spectrum with features characteristic of the α helix (troughs at 208 and 220 nm) develops by addition of trifluoroethanol (Figure 5). A spectrum with similar qualitative features is observed for the native protein Col 1 upon addition of the alcohol (Figure 6). As shown in Figure 7, the transition of Col 1 takes place at a much higher trifluoroethanol to water ratio (midpoint above a ratio of 9:1) than the transition of Col 1 DI which transforms with a midpoint of the transition at an alcohol to water ratio of about 0.8:1. Features characteristic of α-helical elements appear in the CD spectrum of Col 1 DI already in water when the ionic strength

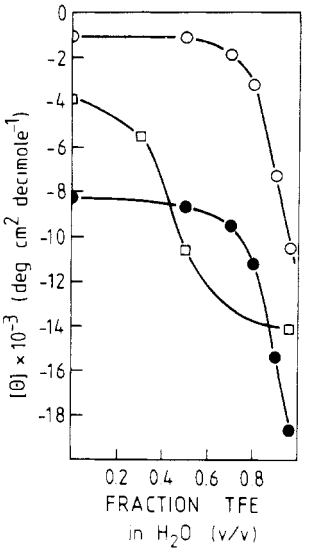

FIGURE 7: Conformational transition of the native (○) and of the reduced and alkylated (□) peptide Col 1 in water/trifluoroethanol (TFE) as followed by the CD signals at 202.5 (filled symbols) and 220 nm (open symbols). Values are taken from Figures 5 and 6.

TABLE III: Molecular Weights (M) and Sedimentation Coefficients (s) for the Native Extension Protein (Col 1), the Reduced and Alkylated Protein (Col 1 DI), the Tryptic Core Protein (Col 1 T) and the Fragment Obtained by Treatment with Staphylococcal Protease Col 1 SP as Determined by Analytical Ultracentrifugation.[a]

Material	$\bar{v}$[b] (mL/g)	M_{found}[a]	M_{calcd}[b]	$s \times 10^{13}$[a] (s)	f/f_{min}[c]
Col 1	0.698	10 700	10 410	1.15	1.68
Col 1 DI	0.698	11 000	11 030	1.45	1.38
Col 1 T	0.701	10 700	9 240	1.25	1.42
Col 1 SP	0.706	7 200	7 299		

[a] Protein concentration 0.1 to 0.7 mg/mL in 0.1 M sodium phosphate buffer, pH 7, at 20 °C. [b] Calculated from the amino acid composition (Table I). [c] Ratio of experimental frictional coefficient f and the frictional coefficient for unsolvated spheres of same M and $\bar{v}$.

is reduced whereas the spectrum of Col 1 is completely invariant under the same conditions (not shown).

Molecular Weights and Frictional Coefficients. The molecular weights of Col 1, of its reduced and alkylated form Col 1 DI, and of the core peptides Col 1 T and Col 1 SP were determined in the ultracentrifuge in order to see whether the proteins tend to aggregate in solution. Partial specific volumes ($\bar{v}$) were calculated from the amino acid composition in Table I (Cohn and Edsall, 1943) (Table III). Good agreement of the molecular weights determined by sedimentation equilibrium and the molecular weights calculated from the amino acid composition was observed (Table III), indicating that the peptides exist as monomers in solution. In the concentration range of 0.1 to 0.5 mg/mL no significant dependence of the apparent molecular weight and sedimentation coefficients on protein concentration was found. At a concentration of 1 mg/mL some association of Col 1 DI was indicated by an upward curvature of the ln c vs. r^2 plot of the sedimentation equilibrium data and by a 20% increase in the sedimentation coefficient.

Sedimentation coefficients were determined in order to estimate the shape of the molecules in solution. The minimum frictional coefficients f_{min} for unsolvated rigid spheres were

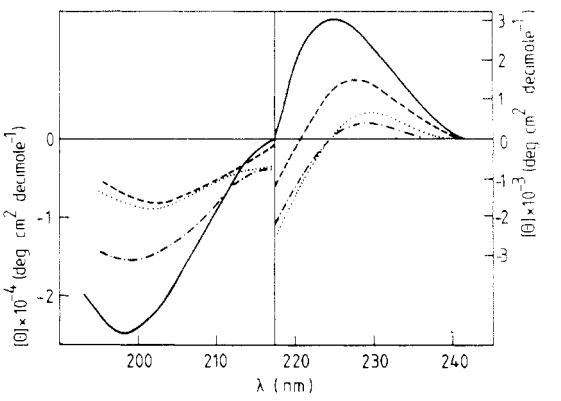

FIGURE 8: CD spectra of peptides CB 0.1 (—) and Col 1 (- - -) at 1 °C and of CB 0.1 (- · -) and Col 1 (· · ·) at 45 °C in 0.025 M sodium phosphate buffer, pH 7.

calculated from M and $\bar{v}$ (Tanford, 1961). The experimental frictional coefficients f were calculated from s, M, $\bar{v}$, and the viscosity of the solvent (Tanford, 1961). The ratios f/f_{min} are given in Table III. The native protein shows a significantly higher f/f_{min} ratio than the derivatives.

Conformation of the Collagen-Like Region in the Peptide CB 0.1. Peptide CB 0.1 comprises Col 1 plus additional collagen-like sequences. Information on the conformation of this part of the molecule was obtained by subtraction of the CD spectrum of Col 1 from that of CB 0.1. Since peptide CB 0.1 was isolated under conditions known to disrupt triple-helical structure, we dialyzed a solution of this peptide in 8 M urea at 6–8 °C against phosphate buffer, pH 7.0, for several days. As judged by sedimentation equilibrium, CB 0.1 formed trimers during dialysis. An apparent molecular weight of 46 000 was found at a protein concentration of 0.5 mg/mL compared with a molecular weight of 17 000 calculated for monomeric CB 0.1 (Becker et al., 1976).

The CD spectrum of trimeric CB 0.1 at 1 °C is compared with that of Col 1 in Figure 8. From these data we constructed the spectrum of the collagen-like portion in CB 0.1. Assuming additivity, the molar ellipticity per residue of the collagen-like part is given by

$$[\theta] = ([\theta]_{CB0.1} - w[\theta]_{Col1}) \frac{1}{1 - w} \quad (1)$$

where $w = 10\,410/16\,950 = 0.612$ is the weight fraction of Col 1 in CB 0.1. The resulting spectrum is shown in Figure 9 and closely resembles that of triple-helical collagen (Brown et al., 1972). Melting of the collagen-like structure in CB 0.1 was observed between 35 and 45 °C at a peptide concentration of 0.8 mg/mL. From the spectra of CB 0.1 and Col 1 at 45 °C (Figure 7), a spectrum was calculated by eq 1 which corresponded to that of denatured collagen (Figure 9). The melting curves did not depend on the disulfide bonds in CB 0.1. Identical results were obtained with CB 0.1 with intact disulfide bridges or with reduced and alkylated CB 0.1. In reduced CB 0.1 peptide, the Col 1 part was found to be in an unfolded state. This followed from the observation that the CD spectra of reduced CB 0.1 (not shown) could be constructed with eq 1 from the CD spectra of Col 1 DI and collagen. As in the case of intact CB 0.1 a good fit of the calculated and observed spectra was obtained for 4 °C with the spectrum of native collagen and at 45 °C with that of denatured collagen.

Kinetic Studies on Triple-Helix Formation. The kinetics of triple helix reformation from the collagen-like portions in peptides CB 0.1 and CB 0.1 DI was followed after quenching samples from 45 to 1 °C. A kinetically unresolved increase of

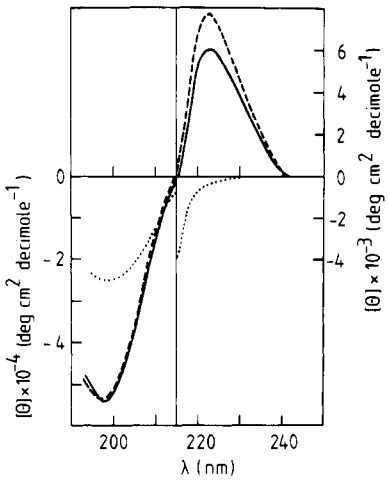

FIGURE 9: CD spectrum of the collagen-like fraction in the peptide CB 0.1 at 1 °C (—) and at 45 °C (···) as calculated from the spectra of CB 0.1 and Col 1 (see Figure 8) by eq 1. The spectrum of native collagen (Brown et al., 1972) is given for comparison (- - -).

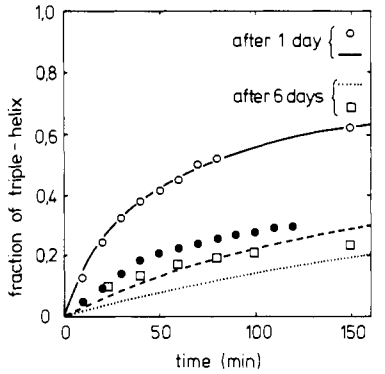

FIGURE 10: Kinetics of triple helix formation by the collagen-like portions in peptides CB 0.1 and CB 0.1 DI in 0.025 M sodium phosphate buffer, pH 7 at 1 °C. The concentrations of CB 0.1 were 7.2×10^{-5} M (— ○ —) and 2.9×10^{-5} M (- - - ● - - -) and that of CB 0.1 DI was 2.35×10^{-5} M (··· □ ···). The values in the upper right corner were measured for the same samples of CB 0.1 and CB 0.1 DI after 1 and 6 days, respectively. The theoretical progress curves were calculated by eq 4 with $\vec{k} = 2 \times 10^4$ M^{-2} s^{-1} for the above peptide concentrations.

the CD signal at 227.5 nm was observed within the first 3 min, the time needed to cool the polarimeter cell. This fast change corresponded to the temperature dependence of the CD signal of the Col 1 portion (see Figure 4). Only the much slower following increase of ellipticity was attributed to triple helix reformation. The degree of triple helix at time t was calculated according to

$$F(t) = \frac{\Delta[\theta]_{slow}(t)}{[\theta]_1 - [\theta]_{45} - \Delta[\theta]_{fast}} \quad (2)$$

where $\Delta[\theta]_{slow}(t)$ is the change of ellipticity in the slow phase, $\Delta[\theta]_{fast}$ is the total change in the unresolved phase, $[\theta]_{45}$ is the ellipticity at 45 °C, $[\theta]_1$ is the ellipticity which the sample exhibited before unfolding after equilibration at 1 °C for several days.

The increase of helicity F with time is plotted in Figure 10. An extremely large concentration dependence of the rate of re-formation is evident. This is expected for the recombination of three chains and was indeed observed for a small peptide (α1-CB 2) from the triple helical part of collagen (Piez and Sherman, 1970) and for synthetic model peptides $(Pro\text{-}Pro\text{-}Gly)_n$ (Sutoh and Noda, 1974). The kinetics of the coil $\rightleftharpoons$ triple helix conversion of the latter systems was described by the all or none mechanism in which intermediates between coiled chains C and fully ordered triple helices T were neglected

$$3C \underset{\bar{k}}{\overset{\vec{k}}{\rightleftharpoons}} T \quad (3)$$

$\vec{k}$ and $\bar{k}$ are overall rate constants of association and dissociation. Although this mechanism allows only an approximate description of the full mechanism (Weidner and Engel, 1974), the data in Figure 10 are analyzed by mechanism 3 in order to allow a comparison with the published data on α1-CB 2 and $(Gly\text{-}Pro\text{-}Pro)_n$. Under our experimental conditions the dissociation of T into C can be neglected at 1 °C and the rate equation for mechanism 3 reads

$$\frac{dF}{dt} = 3\vec{k}c^2(1 - F)^3 \quad (4)$$

where c is the total concentration of peptide chains. With the initial condition $F = 0$ at $t = 0$ integration yields

$$\frac{1 - (1 - F)^2}{2(1 - F)^2} = 3\vec{k}c^2t \quad (5)$$

The theoretical curves in Figure 10 were calculated by eq 5 with $\vec{k} = 2 \times 10^4$ M^{-2} s^{-1}. This value was obtained by a best fit to the data at the highest concentration of peptide CB 0.1 and reasonable agreement is also obtained for the lower concentrations. No significant difference is seen between peptides CB 0.1 and CB 0.1 DI. It has to be emphasized that the kinetic data for CB 0.1 are less accurate than data for collagen-like peptides because of the fact that the optical signal which corresponds to the triple-helix coil transition comprises only a fraction of the total signal.

Discussion

Conformational properties of the amino-terminal region in procollagen pα1(I) chain were studied using two related peptide fragments. The larger fragment CB 0.1 consists of a noncollagenous and a collagenous portion and comprises the whole precursor-specific segment. As shown by ultracentrifugation and CD spectra, CB 0.1 occurs as a trimeric structure with collagenous segments folded into a triple-helical structure. This triple helix shows a remarkable thermal stability presumably due to the presence of eight triplets of the sequence Gly-Pro-Hyp (Becker et al., 1976) which are known to occur in a very stable triple-helical conformation (Sakakibara et al., 1973; Engel et al., 1977). Our observation does not necessarily imply that the same triple-helical conformation exists in the amino-terminal segment of type I procollagen which is composed of two pα1(I) and one pα2 chain. The presence of a similar triple helix in type I procollagen is likely since the precursor-specific segment of pα2 consists mainly of a collagenous sequence (Becker et al., 1977). This may be a general feature of procollagen since a collagenous structure has also been found in the amino-terminal, precursor-specific segment of type III procollagen (Nowack et al., 1976).

Degradation of CB 0.1 by bacterial collagenase liberates a fragment Col 1 which consists primarily of the noncollagenous sequences found in CB 0.1. The CD spectrum of Col 1 is different from that observed for the collagen helix. It does not assume an α helix but consists of an aperiodic structure which may contain large portions of β structure. The spectrum observed in the far UV range is very similar to that of β trypsin (Rosenkranz, 1974), a protein with a defined and rigid but largely aperiodic structure containing up to 50% distorted β structure (Bode and Schwager, 1975).

In the wavelength region of 220 to 330 nm, Col 1 shows several distinct bands. The positive band centered at 252 nm

may be clearly assigned to the optical activity of the disulfide bridges. This assignment is mainly based on the striking similarity of this part of the spectrum with that of neurophysin (Breslow, 1970), a protein of molecular weight 10 000 which contains 6 to 7 disulfide bridges and only one tyrosine and three phenylalanine residues. Bands around 250 nm have been observed also for L-cystine and for other peptides or model compounds containing disulfide bridges (Beychock, 1968, 1973). The positive sign of this band indicates that the majority of the disulfide bridges in Col 1 are in a left-handed skewed conformation. The molar ellipticity per cystine at 252 nm is 1900 as compared with 2300 deg cm^2 $dmol^{-1}$ in neurophysin (Breslow, 1970). According to the spectra of neurophysin and of other cystine-containing compounds, another CD band of opposite sign should originate from the disulfide bridges in the region of 275 to 280 nm. This band is indeed observed with native Col 1 at 275 nm. Similar spectral contributions of disulfide bonds were recently observed for trypsin inhibitor of Adzuki beans (Yoshida et al., 1976) which contains six disulfides but only one aromatic residue (phenylalanine).

The single tyrosine which is present in Col 1 would be expected to contribute a small positive or negative band with $4 < [\theta] < 10$ deg cm^2 $dmol^{-1}$ in the region of 275 to 280 nm (Beychok, 1973). In the CD spectrum of Col 1 this band cannot be clearly separated from the contribution of disulfides. The band at 297.5 nm can be clearly attributed to the single tryptophan residue in Col 1. With the normally observed molar ellipticities of tryptophan residues in proteins of -2000 to -4000 deg cm^2 $dmol^{-1}$ (Beychock, 1973), only about $\frac{1}{3}$ of the observed trough at 297.5 can be explained, but a prominent near-UV tail of the disulfide contributions is often observed (Breslow, 1970; Beychok, 1968, 1973).

Due to the occurrence of a distinct band for tryptophan in the spectrum of Col 1, we conclude that the aromatic portion is fixed in a rigid ordered structure of protein. Since the band at 275 nm does not change with pH, it is likely that the band is entirely made up of disulfide contributions or that tyrosine is not deprotonated. The latter explanation is favored by preliminary NMR studies (Wüthrich, personal communication) which indicate that the tyrosine residue in Col 1 cannot be titrated up to a pH of 12. A rigid and ordered structure of large portions of Col 1 rather than a random coil is suggested by the features of the aromatic residues discussed above and by the invariance of the CD spectrum with changes of environment. Supporting evidence is given by recent hydrogen exchange studies (Wüthrich, personal communication) which indicate that 10 to 20% of the amide hydrogens exchange very slowly in Col 1 while all hydrogens exchange rapidly in the reduced peptide. The existence of conformational antigenic determinants in Col 1 which are lost upon reduction (Rohde et al., 1976) is consistent with these data.

The conformation of Col 1 was found to be stable even under extreme conditions of temperature and pH, and in the presence of denaturants. A loss of structure was only observed at trifluoroethanol to water ratios higher than 8 (v/v). Under these conditions, a CD band at 220 nm appeared indicating that a new conformation was formed containing α-helical elements (Beychok, 1973). This suggests that hydrophobic interactions are essential for maintaining the native conformation of Col 1 and that their disruption by alcohol favors hydrogen-bond formation in a new structure. Col 1 is unfolded upon reduction and alkylation of its five disulfide bridges since the CD spectrum of reduced Col 1 resembles that found for certain randomly coiled peptide chains (Beychok, 1973). In accordance with the flexibility of the chain, no bands were found of the aromatic amino acid residues. The spectrum of the reduced peptide was found to be very sensitive to changes of solvent conditions and in particular to changes in ionic strength. The reduced peptide readily converts to a new structure at lower concentrations of trifluoroethanol than required for a similar conversion of native Col 1. The transition of the reduced peptide is not as sharp as that observed with native peptide and produces a structure which contains about 50% α helix (Beychok, 1973).

It is interesting to note that the core protein Col 1 T which still contains about 90% of the residues of Col 1 is completely unfolded in aqueous solutions as indicated by the far-UV CD spectra and from the absence of a tryptophan band in the near-UV. The breakdown of the structure is apparently caused by three cleavages occurring in two disulfide loops associated with the loss of a tetrapeptide containing the histidine and tyrosine residue of Col 1. In contrast, the removal of about 23 residues from the carboxyl end and about 6 residues from the amino end of Col 1 in Col 1 SP does not significantly alter the CD spectrum. This indicates that flexible regions exist in the amino and carboxyl part of Col 1 and agrees with immunological studies demonstrating sequential antigenic determinants shared by the native and reduced form of the peptide (Rohde et al., 1976). Indeed, these determinants were localized at the amino and carboxyl ends of Col 1 (Rohde and Timpl, unpublished). Flexibility in the carboxyl terminal region in Col 1 is also in accordance with the observation that the collagen-like part in CB 0.1 associates to form a triple-stranded collagen-like helix. As expected, the CD spectra of this region and of Col 1 are apparently additive.

It follows from the high ratio of frictional coefficients f/f_{min} of 1.68 that native Col 1 has an elongated shape in solution with an axial ratio of about 10. Here we assume a prolate ellipsoide as the hydrodynamic equivalent and a degree of solvation of 0.3. The f/f_{min} ratio of about 1.4 for Col 1 DI and for Col 1 T may be assigned to an unfolded randomly coiled state and is in accordance with the information obtained from the CD spectra. For disulfide-containing proteins in the unfolded state, the frictional coefficient increases further when the disulfide bridges are split (Tanford et al., 1967). In contrast, Col 1 exhibits a higher f/f_{min} ratio than its reduced form which again indicates that native Col 1 is not randomly coiled. Assuming a rigid structure, a length of about 140 Å may be estimated for the native Col 1 fragment from the axial ratio of 10 and with the partial specific volume and molecular weight given in Table III. The length of the precursor-specific triple-helical stretch (about 50 amino acid residues) may be estimated to be 150 Å. There is sufficient space in the gap region of native collagen fibers to accommodate the entire amino terminal procollagen extension (Hodge and Petruska, 1963; Doyle et al., 1974). Recently the presence of the amino-terminal extensions in the gap regions of collagen fibrils has been demonstrated by immune electronmicroscopy in dermatospractic sheep skin (Olsen, Wick, and Timpl, unpublished). Previously a faint band was observed in negatively stained fibers which filled about one-third of the gap region (Fjølstad and Helle, 1974).

Several functions have been suggested for the precursor-specific peptides in procollagen (Bornstein, 1974; Martin et al., 1975; Prockop et al., 1976). They may promote association of the three polypeptide chains in the rough endoplasmic reticulum and provide a triple-helical nucleus which would allow a fast propagation of the helical structure along the collagen α chains. It is, however, evident from the kinetic data that the collagen-like regions in CB 0.1 combine to a triple helical structure in a similar way as other single chain peptides. The value obtained for $\tilde{k} = 2 \times 10^4$ M^{-2} s^{-1} at 1 °C is only little

higher than $\vec{k} = 8 \times 10^3\ M^{-2}\ s^{-1}$ observed for $(Pro\text{-}Pro\text{-}Gly)_{15}$ at the same temperature (Sutoh and Noda, 1974), but higher than the value of $\vec{k} = 300\ M^{-2}\ s^{-1}$ found for the collagen peptide α1-CB 2. These differences between the kinetic constants appear to be due to differences in the collagen-like sequences. Most importantly the kinetics is not influenced by the state of the Col 1 portion. The same rate is observed when this part of the peptide is in its native conformation or reduced and alkylated and therefore unfolded.

The failure to detect an association of Col 1, as well as the kinetic data which indicate that the triple-helix formation in CB 0.1 is not faster than that observed for synthetic collagenous peptides of similar size and not influenced by the conformational state of the Col 1 portion, seems to indicate that the amino-terminal extension proteins are not directly involved in the initiation of the triple helix. It is possible that this function requires the proper interaction of precursor-specific segments from the amino end of both the pα1(I) and pα2 chain. Since triple-stranded molecules composed of three pro-α1(I) chains are formed by cultured cells under certain conditions (Müller et al., 1974; Mayne et al., 1975), it does not indicate a crucial role for the pα2 chain segment. The role of the precursor-specific segment at the carboxyl end of procollagen remains still to be explored and recent data indeed suggest that it may have an essential function in the helix formation (Rosenbloom et al., 1976).

References

Becker, U., Helle, O., and Timpl, R. (1977), *FEBS Lett. 73,* 197.

Becker, U., Timpl, R., Helle, O., and Prockop, D. J. (1976), *Biochemistry 15,* 2853.

Beychok, S. (1968), *J. Biol. Chem. 243,* 151.

Beychok, S. (1973), Circular Dichroism, in Physical Principles and Techniques of Protein Chemistry, Part C, Leash, S. J., Ed., London, Academic Press, p 459.

Bode, W., and Schwager, P. (1975), *J. Mol. Biol. 98,* 693.

Bornstein, P. (1974), *Annu. Rev. Biochem. 43,* 567.

Breslow, E. (1970), *Proc. Natl. Acad. Sci. U.S.A. 67,* 493.

Brown, F. R., Di Corato, A., Lorenzi, G. P., and Blout, E. R. (1972), *J. Mol. Biol. 63,* 85.

Byers, P. H., Click, E. M., Harper, E., and Bornstein, P. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 3009.

Cohn E. J., and Edsall, J. T. (1943), *ACS Monogr. 90,* 370.

Doyle, B. B., Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A., and Woodhead-Galloway, J. (1974), *Proc. R. Soc. London, Ser. B 187,* 37.

Engel, J., Chen, H. T., Prockop, D. J., and Klump, H. (1977), *Biopolymers 16,* 601.

Fessler, L. I., Morris, N. P., and Fessler, J. H. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 4905.

Fietzek, P. P., and Kühn, K. (1976), *Int. Rev. Connect. Tiss. Res. 7,* 1.

Fjølstad, M., and Helle, O. (1974), *J. Pathol. 112,* 183.

Furthmayr, H., and Timpl, R. (1971), *Anal. Biochem. 41,* 510.

Furthmayr, H., Timpl, R., Stark, M., Lapiere, C. M., and Kühn, K. (1972), *FEBS Lett. 28,* 247.

Hodge, A. J., and Petruska, J. A. (1963), in Aspects of Protein Structure, Ramachandran, G. N., Ed., New York, N.Y., Academic Press, p 289.

Houmard, J., and Drapeau, G. R. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 3506.

Kohn, L. D., Isersky, C., Zupnik, J., Lenaers, A., Lee, G., and Lapiere, C. M. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 40.

Martin, G. R., Byers, P. H., and Piez, K. A. (1975), *Adv. Enzymol. 42,* 167.

Mayne, R., Vail, M. S., and Miller, E. J. (1975), *Proc. Natl. Acad. Sci. U.S.A. 73,* 1674.

Müller, P. K., Meigel, W. N., Pontz, B., and Raisch, K. (1974), *Hoppe-Seyler's Z. Physiol. Chem. 355,* 985.

Nowack, H., Olsen, B. R., and Timpl, R. (1976), *Eur. J. Biochem. 70,* 205.

Olsen, B. R., Hoffmann, H.-P., and Prockop, D. J. (1976), *Arch. Biochem. Biophys. 175,* 341.

Piez, K. A., and Sherman, M. R. (1970), *Biochemistry 9,* 4134.

Prockop, D. J., Berg, R. A., Kivirikko, K. I., and Uitto, J. (1976), in Biochemistry of Collagen, Ramachandran, G. N., and Reddi, A. H., Ed., New York, N.Y., Plenum Press, p 163.

Rohde, H., Becker, U., Nowack, H., and Timpl, R. (1976), *Immunochemistry 13,* 967.

Rosenbloom, J., Endo, R., and Harsch, M. (1976), *J. Biol. Chem. 251,* 2079.

Rosenkranz, H. (1974), in Proteinase Inhibitors, Fritz, H., et al., Ed., Springer, Berlin, p 458.

Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., and Prockop, D. J. (1973), *Biochim. Biophys. Acta 303,* 198.

Sutoh, K., and Noda, H. (1974), *Biopolymers 13,* 2477.

Tanford, Ch. (1961), Physical Chemistry of Macromolecules, New York, N.Y., Wiley, pp 317–456.

Tanford, Ch., Kawakara, K., and Lapanje, S. (1967), *J. Am. Chem. Soc. 89,* 729.

Tanzer, M. L., Church, R. L., Yaeger, J. A., Wampler, D. E., and Park, E. D. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 3009.

Weidner, H., and Engel, J. (1974), in Peptides, Polypeptides and Proteins, Blout, E. R., Bovey, F. A., Goodman, M., and Lotan, N., Ed., New York, N.Y., Wiley, p 419.

Yoshida, Ch., Yoshiwaka, M., and Tagaki, T. (1976), *J. Biochem. (Tokyo) 80,* 449.