Characterization of Collagen-like Peptides Containing Interruptions in the Repeating Gly-X-Y Sequence†

Cynthia Gwynne Long,‡ Emory Braswell,§ Dan Zhu,§ Josefa Apigo,‡ Jean Baum,‖ and Barbara Brodsky*,‡

Department of Biochemistry, UMDNJ—Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, National Analytical Ultracentrifugation Facility, University of Connecticut, Storrs, Connecticut 06269, and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855

Received June 1, 1993; Revised Manuscript Received August 19, 1993®

ABSTRACT: Glycine is found as every third residue along the entire length of triple helices in fibrillar collagens, but the triple-helix regions of nonfibrillar collagens and other proteins usually contain one or more interruptions in this repeating pattern. A set of four peptides was designed to model the effect of interruptions in the $(Gly\text{-}X\text{-}Y)_n$ repeating pattern on triple-helix formation, stability, and folding. Into the middle of the stable triple-helical peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$, an interruption was introduced representing one of the four possible categories: a glycine deletion, a deletion of a hydroxyproline (Y position), an alanine insertion, or a glycine to alanine substitution. As shown by sedimentation equilibrium, NMR, and CD studies, the introduction of an interruption still allowed formation of trimers in solution, but with a marked decrease in stability. The degree of destabilization and the thermodynamic basis for the loss of stability depended on the type of interruption. The glycine substitution and alanine insertion were the least disruptive, followed by the hydroxyproline deletion, with the glycine deletion being the most destabilizing. Our results suggest that the breaks in these peptides affect both the triple-helical conformation and the monomer conformation. These studies provide a basis for considering the structural and functional consequences of different kinds of interruptions in collagen.

The triple-helical conformation is one of the basic recurring structural motifs in proteins. This structure has long been recognized as a major component of all collagens, and more recently, a number of other unrelated molecules have been found to contain triple-helical domains. Included among these are the macrophage scavenger receptor (Kodama et al., 1990), surfactant protein A (Benson et al., 1985), and the mannose binding protein (Oka et al., 1987). The most characteristic feature of the triple-helical conformation is the presence of glycine as every third residue (Traub & Piez, 1971). This generates a strict Gly-X-Y repeating amino acid sequence where X and Y can be almost any amino acid, but are frequently proline and hydroxyproline, respectively. The high imino acid content promotes the formation of an extended polyproline II like helix in the individual chains, and the three chains then wrap around each other to form a right-handed supercoil (Rich & Crick, 1955; Ramachandran & Kartha, 1955). The glycine residues at every third position are densely packed near the central axis of the supercoil with their amide groups involved in interchain hydrogen bonds. It has been suggested that any residue larger than glycine would not be able to pack in the center, therefore making the Gly-X-Y repeat a requirement for the formation of a stable triple helix (Rich & Crick, 1955; Ramachandran & Kartha, 1955).

More than 15 different genetic types of collagen have been identified in recent years, characterized by the presence of large triple-helical domains (van der Rest & Garrone, 1991; Linsenmayer, 1991). Although some collagens do strictly conform to the $(Gly\text{-}X\text{-}Y)_n$ requirement, this is not the case for all collagen types. The 5 members of the fibrillar collagen family (types I, II, III, V, and XI), which aggregate into the D-periodic fibrils of tendon, bone, and other connective tissues, contain long triple-helical domains of about 1000 amino acid residues with glycine as every third residue. In the triple-helical domains of this family, the Gly-X-Y pattern has been conserved throughout evolution, and when mutations are found that interrupt the pattern, they result in diseases of the extracellular matrix (Kuivaniemi et al., 1991). However, there are a growing number of triple helix containing extracellular matrix molecules that do not form D-periodic fibrils, and are classified as nonfibrillar collagens. These molecules are found to associate into a variety of supramolecular structures (van der Rest & Garrone, 1991). In contrast to the fibrillar collagens, most of the nonfibrillar collagens contain one or more interruptions in the Gly-X-Y sequence. For example, in type IV collagen, which forms a networklike structure in basement membranes, the 1300 amino acid triple-helical domain of the human $\alpha 1$ chain contains 21 interruptions in the Gly-X-Y amino acid sequence (Brazel et al., 1987; Soininen et al., 1987). The triple-helical domain (454 residues) of type VIII collagen, which may play a role in cellular organization, contains 8 interruptions per chain (Yamaguchi et al., 1989; Sage & Iruela-Arispe, 1990), and the triple-helical domain (336 residues) of type VI collagen, which is found in microfibrils, contains 2 interruptions per chain (Chu et al., 1988). Among noncollageneous molecules, the triple helices found in C1q (Reid et al., 1972), mannose binding protein (Oka et al., 1987), and surfactant protein A (Benson et al., 1985) each contain a single interruption near the center of their triple-helical domains. Thus, outside of the fibrillar class of collagens, it is common to find one or more breaks in the Gly-X-Y pattern of triple-helical domains.

† This work was supported by NIH Grants AR19626 (B.B.) and GM45302 (J.B.) and by NSF Grant DIR 8717034 (National Ultracentrifugation Facility).

* To whom correspondence should be addressed.
‡ UMDNJ—Robert Wood Johnson Medical School.
§ University of Connecticut.
‖ Rutgers University.

® Abstract published in *Advance ACS Abstracts,* October 1, 1993.

0006-2960/93/0432-11688$04.00/0

Very little is known about what role interruptions play in the normal structure and function of nonfibrillar collagens and other triple helix containing molecules. One possibility is that they may serve a primarily structural role by introducing kinks or sites of local flexibility into the rigid triple helix. For example, in the C1q component of complement, the presence of an interruption in the triple helix is proposed to facilitate the formation of a 50° kink (Porter & Reid, 1979; Schumaker et al., 1981), and flexible sites in type IV collagen detected by rotary shadowing correlate fairly well with the positions of interruptions in the sequence (Hofmann et al., 1978; Saus et al., 1989; Dölz et al., 1988). It is also possible that structural perturbations resulting from interruptions serve as recognition sites for binding enzymes, cells, or other extracellular matrix components (Koliakos et al., 1989).

The types of interruptions found in the Gly-X-Y sequence of triple-helical domains can be grouped into four broad categories depending on the nature of the sequence involved. Sequences such as Gly-X-Y-Ala-X-Y can be classified as a glycine substitution, while Gly-X-Y-X-Y can be classified as a glycine deletion, and Gly-X-Gly-X-Y as a deletion of a Y (or X) residue. The fourth type of interruption involves sequences such as Gly-X-Y-$(X)_n$-Gly-X-Y which can be classified as an insertion of n residues. In the nonfibrillar collagens, insertions and X or Y deletions are the most frequent interruptions found, while glycine substitutions or deletions are less common.

Studies on model peptides present opportunities for examining the effect of interruptions in triple helices that are not easily approached using whole collagen molecules, due to their large size and the presence of multiple interruptions. In model peptides, the impact of a single interruption can be studied in a defined amino acid sequence environment. The small size of the peptides makes them amenable to studies by nuclear magnetic resonance (NMR),[1] and the reversibility of the helix–coil transitions allows the effects of a single amino acid change on thermodynamic parameters to be quantitated. We have generated a set of peptides, each containing one of the four classes of interruptions found in various triple helices. Each interruption is introduced into the peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$, which forms a very stable triple helix. CD, NMR, and equilibrium sedimentation are used to determine whether the peptides are able to form trimers and adopt a standard triple-helical conformation. Comparison of peptides with different types of interruptions can clarify whether the effect on the triple helix depends on the type of interruption.

MATERIALS AND METHODS

Peptides. The peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ was purchased from Peptides International, Louisville, KY. All other peptides were synthesized by the Protein Microchemistry Laboratory, at the Center for Advanced Biotechnology and Medicine, Piscataway, NJ. The synthesis was carried out on an Applied Biosystem 430A peptide synthesizer by stepwise solid-phase procedures on *t*-Boc-L-Gly-PAM resin. The *t*-Boc group was used for the amino terminus. The side chain protecting group for hydroxyproline was *O*-Bzl. All amino acids were double-coupled, and the final coupling percents as assayed by a ninhydrin test were greater than 99%. The peptides were cleaved from the resin using hydrogen fluoride, extracted, and redissolved in 0.1% trifluoroacetic acid. The purity of all the peptides was greater than 95% by reverse-phase HPLC on a Vydac C-18 column.

Mass Spectrometry. Liquid secondary ion mass spectrometry was conducted using a VG Analytical ZAB-T instrument at the Center for Advanced Food Technology, New Brunswick, NJ. The cesium gun anode current was 2 mA, and the voltage was 30 kV. The mass range was 200–3000 daltons, and the scan time was 2 s/decade. The instrument accelerating voltage was 8 kV. The samples were prepared from powder in magic bullet matrix.

Equilibrium Sedimentation. Prior to sedimentation, lyophilized peptides were dissolved in 0.1 M acetic acid at a concentration of 1 mg/mL. For studies at neutral pH, the solutions were dialyzed in Spectra/Por dialysis tubing with a molecular weight cutoff of either 500 or 1000 against phosphate buffer (0.01 M phosphate/0.15 M sodium chloride, pH 7.0) for 18–24 h at 4 °C. The samples were packed on ice and shipped overnight to the National Analytical Ultracentrifuge Facility at the University of Connecticut, Storrs, CT. All samples were centrifuged in a Beckman Model E analytical ultracentrifuge at speeds (40 000–52 000 rpm) designed to produce a gradient equivalent to a standard deviation (σ) for the monomer (Johnson et al., 1981) of at least 2 for a 6–9-mm solution column. Runs were carried out on all peptides at both 10 °C and 30 °C. The concentration resulting from the sedimentation ranged from about 10 μg to 1–3 mg/mL. The data were acquired through a real-time TV system which records the fringe displacement as a function of the distance from the center of rotation with a typical error of about ±0.01–0.02 fringe. Under these conditions, this corresponds to a concentration error of ±2–4 μg/mL. The data were spaced at ca. 10 μM intervals down the cell. Data were taken every 2–4 h to test for equilibrium. Equilibrium was determined to have been reached when no change in the gradient could be observed after subtraction of successive interferograms from each other. Blanks were run (H_2O vs H_2O) both before and after the experimental run at the same rotor speeds as the experimental run, in order to correct for optical distortions. The final equilibrium data were subsequently corrected by subtracting the blank.

Data analysis was performed using a nonlinear least-squares program (Johnson et al., 1981) assuming various models. The criteria for goodness of fit are that a reasonable value of the monomer be produced, that the rms error be small (usually less than 0.02 fringe, which is approximately 10 μg/mL), and that the error be random, i.e., little systematic error in the residuals. To convert the values of σ obtained into molecular weight, one needs to know the value of the specific volume (v) of the molecule in its solvent and the density of the solvent. The values of v were calculated from the amino acid composition data (Cohn & Edsall, 1943) and were found to range from 0.692 to 0.694 for all the peptides. The density of the solvent was measured with a Par DMA 602 density meter and found to be 1.0057 g/mL at 20 °C, which, when corrected to 10 and 30 °C, becomes 1.0073 and 1.0032 g/mL, respectively.

Circular Dichroism. Circular dichroism (CD) spectra were recorded on an Aviv Model 60DS or 62DS spectropolarimeter. Samples were prepared at a concentration of 2 mg/mL in 0.1 M acetic acid and equilibrated at 4 °C for at least 24 h prior to recording spectra. Cells of path length 0.2–2.0 mm were used, and the temperature in the cell was controlled using a Hewlett Packard Peltier thermoelectric temperature controller. For wavelength scans, the signal was averaged from 3 to 5 s,

[1] Abbreviations: CD, circular dichroism; NMR, nuclear magnetic resonance; 1D, one dimensional; o.i., osteogenesis imperfecta; standard one-letter and three-letter abbreviations are used for the common amino acids, with hydroxyproline denoted by O (one-letter code) and Hyp (three-letter code); thus, $(POG)_{10}$ is used for $(Pro\text{-}Hyp\text{-}Gly)_{10}$.

Table I: Molecular Weights and Association Constants of the Four Interrupted Peptides

peptide	sequence	calculated MW	mass spec MW	ln K, 10 °C[a]	ln K, 30 °C[a]
$(Pro\text{-}Hyp\text{-}Gly)_{10}$	$(Pro\text{-}Hyp\text{-}Gly)_{10}$	2691.0	2691.0	only trimers	very large
Gly→Ala	$(Pro\text{-}Hyp\text{-}Gly)_4\text{-}Pro\text{-}Hyp\text{-}Ala\text{-}(Pro\text{-}Hyp\text{-}Gly)_5$	2705.0	2705.3	+6.0 ± 2.0	−2.2 ± 0.5
Ala^+	$(Pro\text{-}Hyp\text{-}Gly)_4\text{-}Pro\text{-}Hyp\text{-}Ala\text{-}Gly\text{-}(Pro\text{-}Hyp\text{-}Gly)_5$	2762.1	2762.2	+8 ± 1	−2.9 ± 0.1
Hyp^-	$(Pro\text{-}Hyp\text{-}Gly)_4\text{-}Pro\text{-}Gly\text{-}(Pro\text{-}Hyp\text{-}Gly)_5$	2577.9	2578.3	+1.5 ± 0.7	−2.9 ± 1.3
Gly^-	$(Pro\text{-}Hyp\text{-}Gly)_4\text{-}Pro\text{-}Hyp\text{-}(Pro\text{-}Hyp\text{-}Gly)_5$	2633.9	2635.0	−2.5 ± 0.3	−9.6 ± 1.6

[a] Association constant for monomer to trimer, measured in 0.1 M acetic acid by equilibrium ultracentrifugation.

and data points were taken every 0.5 nm. For equilibrium melting transitions, the temperature in the cell was increased at a rate of 0.1 °C/min, or by increasing the temperature in increments of 0.3 °C and equilibrating for 3 min at each temperature before collecting a data point. Data were recorded at a constant wavelength of 227 nm. The fraction folded was determined by the equation:

$$F = \frac{\theta_{observed} - \theta_{monomer}}{\theta_{helical} - \theta_{monomer}}$$

where $\theta_{observed}$ is the observed mean residue ellipticity, $\theta_{helical}$ is the ellipticity when the peptide is fully associated, and $\theta_{monomer}$ is the ellipticity of the monomer. The monomer ellipticity exhibits a strong temperature dependence (Engel et al., 1977). The value for the ellipticity of the monomer at low temperatures was determined by extrapolating the slope of the monomer ellipticity at high temperature to lower temperatures in the transition region (Engel et al., 1977). For each peptide, the slope of the monomer transition is determined by collecting data points every 0.3 °C for more than 10 °C above the trimer to monomer transition. This procedure will yield accurate values for the thermodynamic parameters and is believed to account for disparities with previously published data. The rate of triple-helix formation was monitored by first denaturing the samples in a 2-mm water-jacketed circular cell at 85 °C for 10–15 min. Folding was initiated by switching to a water bath adjusted so that the final temperature in the cell was 10 °C. The ellipticity in the cell was monitored at 230 nm, and data points were collected every 5 s with an averaging time of 1 s.

One-Dimensional ^1H NMR. One-dimensional ^1H NMR spectra were obtained on a Varian XL400 spectrophotometer. Samples were prepared in 0.1 M deuterated acetic acid, and equilibrated at 4 °C for at least 24 h before spectra were recorded. A spectral width of 5000 Hz was used, and 8192 data points were collected per scan. Two hundred scans were averaged for the final spectrum, with 1-s delay time between consecutive scans. For equilibrium melting transitions, the temperature in the probe was increased at intervals varying from 1 to 5 °C. The samples were equilibrated for 10 min/°C or a maximum of 20 min per interval. The spectra were plotted and integrated by cutting out the peaks and weighing them. An estimate for the change in the triple-helix content of the peptides was obtained by calculating the ratio of the triple-helical proline $C_\delta H$ peak (3.2 ppm) to the sum of the denatured glycine $C_\alpha H$ peak (4.2 ppm) plus the helical proline $C_\delta H$ peak (Li, 1993; Brodsky et al., 1992). For the Gly→Ala peptide, the loss of one glycine residue was taken into account. The fraction folded at each temperature was calculated by

$$F = \%\ TH_t / \%\ TH_{max}$$

where $\%\ TH_t$ is the percent triple helix at temperature t and $\%\ TH_{max}$ is the percent triple helix when the peptide is fully associated. The values of percent triple helix calculated by this method represent an approximation due to the error associated with determining the base line for integration of the peaks.

Calculation of Thermodynamic Parameters. The equilibrium melting transitions were fit to a two-state monomer to trimer transition, and the T_m and $\Delta H°$ were determined by curve-fitting with the equation (Engel et al., 1977; Marky & Breslauer, 1987):

$$K = \exp\left[\frac{\Delta H°}{RT}\left(\frac{T}{T_m} - 1\right) - \ln\,(0.75\ cm^2)\right]$$

The equation:

$$T_m = \frac{\Delta H°}{\Delta S° + R\ln(0.75\ cm^2)}$$

was used to calculate $\Delta S°$ and to correct the T_m values to a common concentration of 1 mm. The equation:

$$\Delta G° = -RT\ln K = \Delta H° - T\Delta S°$$

was used to calculate $\Delta G°$ at 25 °C.

RESULTS

Model Peptide Design. The peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ has been studied by ultracentrifugation (Sakakibara et al., 1973), NMR (Li et al., 1993), X-ray fiber diffraction (Rogulenkova et al., 1964; Andreeva et al., 1961), and optical methods (Sakakibara et al., 1973; Engel et al., 1977), and was shown to adopt a triple-helical conformation similar to that observed for collagen. The tripeptide sequence Pro-Hyp-Gly is the most commonly occurring tripeptide found in the triple-helical domains of collagen, representing about 10% of the tripeptides in fibrillar collagens and in type IV collagen (Traub & Piez, 1971). Of the synthetic polytripeptides that have been studied, this sequence is observed to form the most stable triple helix and is proposed to be of essential importance in the overall stability of the triple helix (Sakakibara et al., 1973; Engel et al., 1977; Burjanadze et al., 1982; Privalov & Tiktopulo, 1970). Thus, the peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ was taken as the triple-helical prototype into which interruptions were introduced.

A homologous set of peptides was synthesized to investigate the influence of the four different classes of interruptions in the Gly-X-Y pattern of the triple helix (Table I). These peptides differ from each other by the insertion, substitution, or deletion of one residue in the center of the peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$. The Gly→Ala peptide was designed to investigate the impact of a single glycine substitution, and preliminary results on this peptide have been previously reported (Long et al., 1992; Brodsky et al., 1992). To model the effects of a deletion, a single glycine (Gly^-) or hydroxyproline (Hyp^-) residue is deleted. In type IV collagen, insertions have been found that range in length from 1 to more than 20 residues. As a first attempt to model this category of interruption, a single alanine residue is inserted into the center of the peptide $(Pro\text{-}Hyp\text{-}Gly)_{10}$ (referred to as Ala^+ peptide). The molecular weights of these peptides as determined by mass spectrometry were as expected on the basis of the amino acid sequence (Table I).

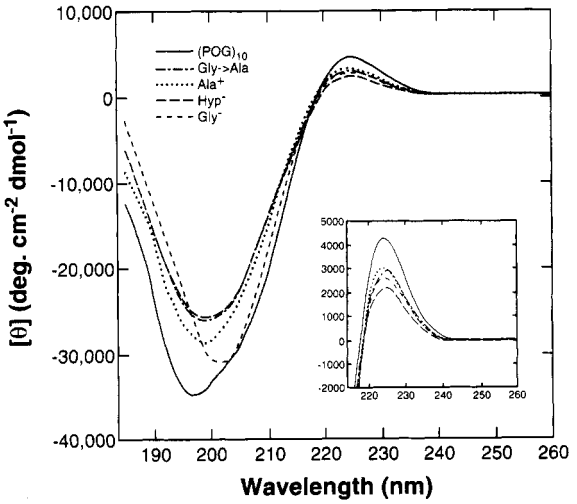

FIGURE 1: Circular dichroism spectra of (Pro-Hyp-Gly)$_{10}$, Gly→Ala, Ala$^+$, Hyp$^-$, and Gly$^-$ at 2 mg/mL in 0.1 M acetic acid, 10 °C. The maximum near 225 nm is shown using an expanded scale in the insert.

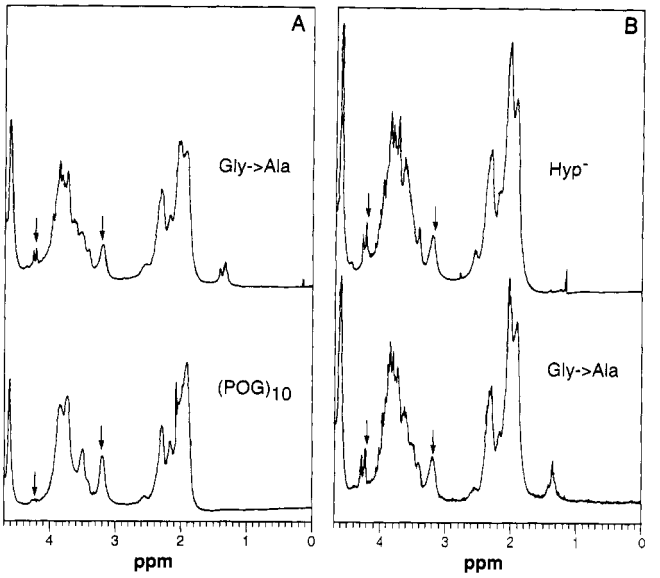

FIGURE 2: One-dimensional NMR of peptides, with the peaks denoting helical Pro $C_\delta H$ (3.2 ppm) and nonhelical Gly $C_\alpha H$ (4.2 ppm) marked with arrows. (A) Comparison of (Pro-Hyp-Gly)$_{10}$ (2 mg/mL) and the Gly→Ala peptide (11 mg/mL) at concentrations where both are more than 99% in the trimer form. (B) Comparison of the Gly→Ala peptide (1 mg/mL) and the Hyp$^-$ peptide (26 mg/mL) at concentrations where both are 95% associated as trimers, as calculated from the sedimentation equilibrium data.

Equilibrium Sedimentation. To investigate whether the interrupted peptides would form trimers, equilibrium sedimentation was done at 10 and 30 °C. The peptide (Pro-Hyp-Gly)$_{10}$ in 0.1 M acetic acid is completely associated into trimers at 10 °C. At 30 °C, it is still almost completely trimeric, with only a small monomer population, consistent with published results (Sakakibara et al., 1973). Results at neutral pH were very similar to those in acid solution, but the small amount of nonideality seen at low pH was eliminated. Monomer–trimer association is the only model consistent with the sedimentation data for all of the peptides, except for the Ala$^+$ peptide at neutral pH. The neutral pH data of Ala$^+$ at low temperature fit a monomer–tetramer model, while the acid pH data of Ala$^+$ fit well to a monomer–trimer model. Since the CD and NMR experiments in this study were done at acidic pH, only monomer–trimer association is considered further for Ala$^+$, as well as for the other peptides.

All of the interrupted peptides showed shifts in the equilibrium constant toward the monomer compared with (Pro-Hyp-Gly)$_{10}$ (Table I). At low temperature, the Gly→Ala and Ala$^+$ peptides are very highly associated into trimers, while Hyp$^-$ is significantly less associated, and the Gly$^-$ peptide is predominantly shifted to the monomer form. For instance, at 10 °C, at a concentration of 2 mg/mL, the Gly→Ala and Ala$^+$ peptides are about 97% trimer, the Hyp$^-$ peptide is 75% trimer, and the Gly$^-$ peptide is 15% trimer. At 30 °C, the Gly→Ala, Ala$^+$, and Hyp$^-$ peptides are largely monomeric, with only a small trimer population, while for the Gly$^-$ peptide only monomer is detected.

CD and NMR Spectra. CD and NMR were used to investigate whether the trimers formed by the interrupted peptides contained the characteristic triple-helical conformation. A triple helix, such as (Pro-Hyp-Gly)$_{10}$, exhibits a very characteristic CD spectrum with a positive peak near 220–225 nm and a large negative peak near 197 nm (Heidemann & Roth, 1982; Piez & Sherman, 1970). Upon denaturation, the intensities of these peaks are decreased. The CD spectra for all the interrupted peptides at 2 mg/mL, 10 °C, exhibit spectral characteristics of the triple-helical conformation, but the magnitudes of the maxima and minima are decreased (Figure 1). For Gly$^-$, the minimum is also shifted, to 201.5 nm. The decreased magnitude of the peaks is consistent with a loss of triple-helical conformation in the interrupted peptides. This loss could be a perturbation of the trimer resulting from the introduction of an interruption in the middle of the chain or be due to an increase in the monomer population as a result of the shift in the equilibrium constant. For the Gly→Ala peptide and Ala$^+$ peptides, with a high degree of association and no shift in the wavelength of the peaks, the decrease in the magnitude of the peaks is likely to result from some loss of triple-helical content with each trimer. The shift of the negative peak for Gly$^-$ may reflect the presence of polyproline II like conformation, which has a minimum at 205 nm and a maximum at 225 nm, as a result of the high proportion of monomer present in this peptide.

One-dimensional NMR can be used to estimate the triple-helical content (Li, 1993; Brodsky et al., 1992; Long et al., 1992). At 10 °C, the spectra of the interrupted peptides have features similar to that of (Pro-Hyp-Gly)$_{10}$ (Figure 2). When recorded at the same concentration, all of the interrupted peptides have a decrease in the intensity of the peak at 3.2 ppm, which is assigned to the proline $C_\delta H$ in the triple-helical conformation, together with a corresponding increase in the peak at 4.2 ppm, which is assigned to the nonhelical glycine $C_\alpha H$ (data not shown). As in the CD studies, the observed loss of triple helix implied by these intensity changes could reflect the change in the equilibrium constant, resulting in more monomer and less trimer, or a perturbation and loss of helix within each trimer. Comparison of peptides at comparable degrees of association will eliminate the equilibrium effect and clarify if there is loss of structure in the trimers due to the interruption. The 1D spectra of the Gly→Ala peptide (11 mg/mL) and (Pro-Hyp-Gly)$_{10}$ (2 mg/mL) were compared at concentrations where the association of each peptide is greater than 99% (Figure 2A), and the differences in the intensities of the 3.2- and 4.2-ppm peaks suggest a loss of about 25% of the triple-helical content as a result of the glycine substitution. A similar helical loss is found for the Ala$^+$ peptide (data not shown). The lower equilibrium constant of Hyp$^-$ made it impractical to reach 99% association, but comparison of the Hyp$^-$ peptide (26 mg/mL) and Gly→Ala peptide (1 mg/mL) was made at concentrations where each is calculated

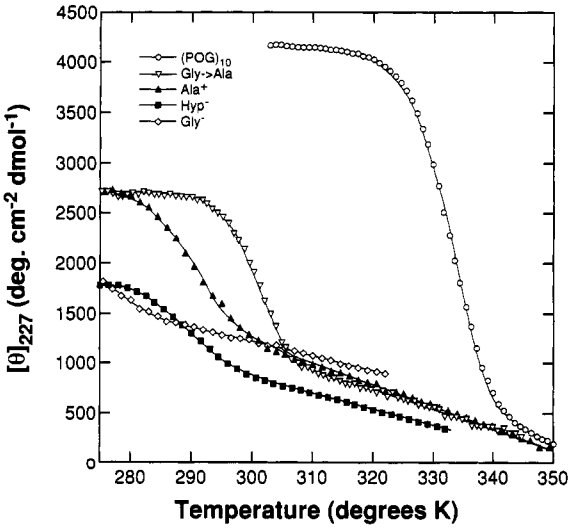

FIGURE 3: Thermal equilibrium melting curves of $(Pro\text{-}Hyp\text{-}Gly)_{10}$, Gly→Ala, Ala^+, Hyp^-, and Gly^-, monitored by circular dichroism at a wavelength of 227 nm. All peptides are at a concentration of 2 mg/mL in 0.1 M acetic acid. The solid lines represent the theoretical curves determined by curve-fitting, except in the case of Gly^-. Because the transition for Gly^- is not complete even at low temperature, it was not possible to fit the data theoretically, but a line was drawn for ease of viewing.

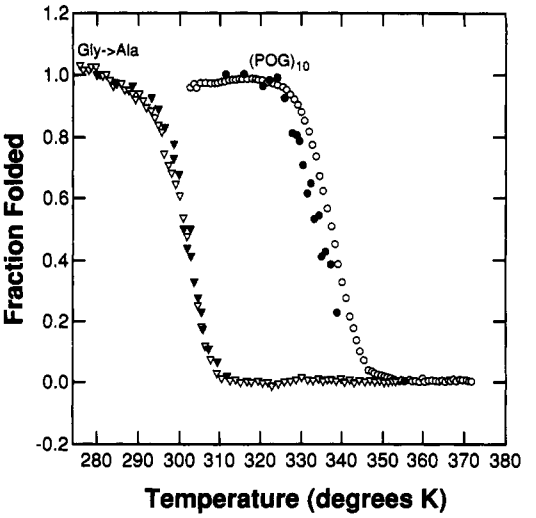

FIGURE 4: Comparison of thermal equilibrium melting curves monitored by CD (open symbols) and 1D NMR (closed circles) for $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and the Gly→Ala peptide. Samples are at a concentration of 2 mg/mL in 0.1 M deuterated acetic acid. CD melting curves are monitored at 227 nm. It is interesting to note that in deuterated solvent, the thermal stability of the peptides is increased compared to the nondeuterated solvent, and the mean residue ellipticity is increased. A similar effect has also been observed for $(Pro\text{-}Pro\text{-}Gly)_{10}$ in deuterated solvent (Bhatnagar et al., 1990).

on the basis of equilibrium constants to reach 95% trimer. The ratio of the helical 3.2-ppm peak to the total peak (nonhelical plus helical) was very similar for both peptides (Figure 2B), suggesting that Hyp^- has a similar helix content to the Gly→Ala peptide. Thus, the glycine substitution, the alanine insertion, and the hydroxyproline deletion all appear to have an estimated 25% of the helix perturbed. The trimer association of Gly^- as determined by equilibrium sedimentation was too low to make any similar comparison.

Thermal Stability by CD and NMR. The thermal stabilities of the four interrupted peptides were compared to $(Pro\text{-}Hyp\text{-}Gly)_{10}$, which has a melting transition with a T_m of about 60 °C (Sakakibara et al., 1973; Engel et al., 1977). The presence of interruptions in this peptide results in a dramatic decrease in the thermal stabilty of the trimer (Figure 3, Table II). The T_m is decreased to 29 °C for the Gly→Ala peptide, to 19 °C for the alanine insertion peptide, and to 17 °C for the hydroxyproline deletion peptide. The deletion of a glycine residue is the most destabilizing interruption, with the melting transition not complete at 2 °C. All of the peptides, except Gly^-, showed a sharp trimer to monomer thermal transition which is followed, at higher temperature, by a linear temperature dependence of the ellipticity of the monomer (Li, 1993). The slope and intercept of this linear change varied in the different peptides (Table II).

To analyze the CD melting curves, a two-state model must be assumed for the monomer to trimer transition. Support for a two-state model can be obtained by determining that the thermal transition is the same when monitored by two or more methods that reflect different structural elements of the molecule (Pace et al., 1989; Marky & Breslauer, 1987). CD melting curves of the triple helix, monitored at 225 nm, arise from the structure of the peptide bond, and can be compared with melting curves derived from the 1D NMR spectra, which monitor the environment of the side-chain protons of the peptide (Li, 1993; Long et al., 1992). Equilibrium melting curves of the peptides using these two methods show sharp melting transitions with similar values (Figure 4), supporting a two-state transition. This result is consistent with calorimetric studies that proved a two-state model for a similar peptide, $(Pro\text{-}Pro\text{-}Gly)_{10}$ (Engel et al., 1977).

Thermodynamic parameters for $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and the three interrupted peptides with complete thermal transitions (Gly→Ala, Ala^+, and Hyp^-) were calculated by analyzing the shape of the CD melting curves using a two-state trimer to monomer model (Table II; Marky & Breslauer, 1987). The theoretical curves show an excellent fit to the real data (Figure 3). Compared with $(Pro\text{-}Hyp\text{-}Gly)_{10}$, the Gly→Ala peptide has an enthalpy value which is similar, but slightly higher, and there is a significant increase in the entropy term. For the Hyp^- and Ala^+ peptides, the interruptions result in a decrease in both the enthalpy and entropy terms.

The thermodynamic effect of interruptions on the monomer chain is also considered. Since the amino acid sequence of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ is composed of only the most flexible amino acid residue (glycine) and the most rigid residues (proline and hydroxyproline), it is possible to predict what effect single

Table II: Thermodynamic Parameters and Monomer Slopes Calculated from CD Equilibrium Melting Transitions of Interrupted Peptides

peptide	T_m[a] (K)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/mol)	$\Delta G°$ (kcal/mol)	monomer slope[b]	monomer intercept[c]
$(Pro\text{-}Hyp\text{-}Gly)_{10}$	334.7	125	345.7	22.1	−26.1	2184
Gly→Ala	302.3	129	400.3	10.2	−19.6	1641
Ala^+	291.7	74	226.4	6.7	−22.4	1821
Hyp^-	289.8	71	217.9	6.3	−16.3	1289
Gly^-	NA[e]	NA	NA	NA	−15.6	1638

[a] T_m values are corrected to a common concentration of 1 mM. [b] The slope for the temperature-dependent decrease in ellipticity of the monomer; units are deg·cm^{-2}·dmol^{-1}·K^{-1}. [c] The y intercept obtained by extrapolation of the linear temperature dependence of ellipticity of the monomer; units are deg·cm^{-2}·dmol^{-1}.

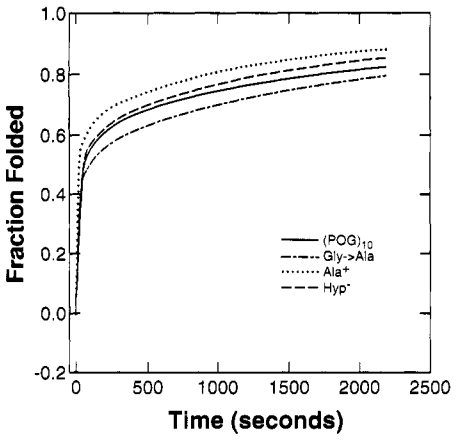

FIGURE 5: Folding curves of (Pro-Hyp-Gly)$_{10}$, Gly→Ala, Hyp$^-$, Ala$^+$, and Gly$^-$ monitored by circular dichroism at 230 nm. The folding is expressed as the fraction of the final mean residue ellipticity as a function of time. The samples are at a concentration of 2 mg/mL in 0.1 M acetic acid.

amino acid changes will have on the configurational entropy of the monomer chains (Nemethy et al., 1966). The replacement of a glycine by any other residue will decrease the configurational entropy since all residues are more rigid than glycine. Likewise, a glycine deletion will also decrease the configurational entropy since it results in a net increase in the imino acid content while decreasing the glycine content. A Hyp deletion and an Ala insertion, on the other hand, decrease the imino acid contribution and will have an increased configurational entropy.

Folding. Circular dichroism spectroscopy was also used to investigate the effect of the interruptions on the rate of triple-helix formation. Due to the temperature-dependent ellipticity of the monomer chains and technical difficulties in reducing the temperature in the CD cell within a short enough time span, it was not possible to obtain rate constants for the folding process. However, qualitative comparisons of the interrupted peptides, calculating the percent of the triple-helical ellipticity that is reached after a specific time of refolding, showed the folding rates to be similar (Figure 5).

DISCUSSION

Studies on collagen and triple helices have stressed the uniform nature of the structure and the repetitiveness of the Gly-X-Y sequence, but recent evidence suggests that interruptions in the Gly-X-Y repeat and resulting structural imperfections may be important in the structure and function of nonfibrillar collagens and other triple-helix-containing proteins. The studies on model peptides presented here focus on the effect that different interruptions have on the triple helix. Previously, polypeptides where every glycine residue is replaced with an alanine residue, such as (Ala-Pro-Pro)$_n$, were not found to adopt a triple-helical conformation (Rapaka et al., 1976). Our studies show that introduction of a single interruption in the very stable environment of Pro-Hyp-Gly repeating sequences allowed the formation of trimers in solution. The features of these trimers depend upon the kind of interruption present. The peptides with a glycine substitution, alanine insertion, and hydroxyproline deletion all formed stable trimers with a high triple-helical content and sharp thermal transitions. The peptides containing a glycine to alanine substitution or an alanine insertion were least perturbed in terms of the equilibrium constant of the monomer–trimer association, the ellipticity of their CD spectrum, and their thermal stability, while the hydroxyproline deletion peptide was substantially less associated. The Gly$^-$ peptide was capable of forming some trimer, but was mostly in the monomer form even at low temperatures. It was clear that the deletion of a glycine was by far the most disruptive interruption of the group studied here.

CD and NMR data indicate that the interruptions have led to a perturbation of the triple-helical conformation in the trimer. The perturbation appears to be relatively local based on the fact that peptides that are only 30 residues in length are still able to form triple helices, and NMR data suggest a disturbance in the range of 6–9 residues per chain. The conformational space available to the residues in these imino acid rich peptides excludes many possible conformations (such as α helix or β sheet), but allows a polyproline II helix, as found in the triple helix, or various β turns. The two likely possibilities at the interruption site are a distorted triple helix and a structure with β bends looping out from the helix. Computer models for both of these possible conformational changes have been generated for the Gly→Ala peptide (Long et al., 1992), and two-dimensional NMR (Fan, Long, Baum, and Brodsky, unpublished data) and X-ray crystallographic studies (Bella, Brodsky, and Berman, unpublished data) are in progress to define the perturbation present in this peptide. The deletion of a residue may be more restrictive than a substitution or an insertion, and thus cause the greater disruptive effects seen for the Hyp$^-$ and Gly$^-$ peptides. The Gly$^-$ peptide contains a stretch of four imino acids at the interruption site, and it may be difficult to accommodate this rigid region into a triple-helical environment.

The breaks in the Gly-X-Y repeating pattern result in a dramatic decrease in the thermal stability of all peptides compared with (Pro-Hyp-Gly)$_{10}$. The thermodynamic analysis suggests that the factors contributing to the stabilization of the interrupted peptides are very different from each other and from (Pro-Hyp-Gly)$_{10}$. In the case of the Ala$^+$ and Hyp$^-$ peptides, there are substantial decreases in both the enthalpy and entropy terms, suggesting the involvement of a number of different factors in the destabilization which are difficult to sort out. However, for the Gly→Ala peptide, the $\Delta H°$ of folding is slightly greater than for (Pro-Hyp-Gly)$_{10}$, indicating that the destabilization is likely to arise from the large entropic change. Therefore, the glycine to alanine substitution is leading to a larger difference in entropy between the monomer and trimer forms than seen for (Pro-Hyp-Gly)$_{10}$. Such an increased entropy difference could arise from a change in the monomer conformation such that it is less structured, from a change in trimer conformation such that it is more ordered, or from solvent-mediated effects. All of these factors may play a role. The CD data suggest that there are conformational differences among the different monomers, but analysis of configurational entropy suggests a more ordered, rather than less ordered, Gly→Ala monomer chain. Changes are detected experimentally in the trimer conformation of the Gly→Ala peptide, but are suggestive of disruption of the triple helix rather than a more ordered conformation. These considerations of just the monomer and trimer do not justify the observed entropic destabilization. Therefore, solvent is likely to play a very important role in the entropy change, and water is known to be an integral part of the triple-helical structure (Privalov, 1982).

In basement membrane type I collagen, there are more than 20 breaks of various kinds in the Gly-X-Y repeating pattern. Despite this large number of interruptions, type IV collagen has a thermal stability and folding rate comparable to those of collagens with no interruptions (Davis et al., 1989).

Most of the sites of interruptions are found at the same location in both chains of type IV collagen [α1(IV) and α2(IV)] (Saus et al., 1989), and some sites are conserved in different species as well (Pettett & Kingston, 1991). The conservation of location supports a role for these breaks in molecular conformation, supramolecular organization, or binding of other molecules. While the locations of breaks along the triple helix are often conserved, the nature of the interruption at a given site usually varies between α1(IV) and α2(IV) chains. Our results indicate that different categories of interruptions can have different effects on stability and conformation, with the glycine deletion being by far the most disruptive. It is interesting to note that there are few cases of glycine deletions found in type IV collagen, other nonfibrillar collagens, and other triple-helix containing molecules, so this type of break may play a special role. The peptides containing interruptions other than a glycine deletion still form trimers. Their thermal stability is substantially decreased from $(Pro\text{-}Hyp\text{-}Gly)_{10}$, but is not very different from that of uninterrupted peptides of similar length with a $(X\text{-}Y\text{-}Gly)_n$ pattern but with several imino acid poor triplets (Li et al., 1993). The presence of a small insertion, a glycine substitution, or an X or Y deletion in type IV collagen might, therefore, not be expected to introduce a much greater destabilization than found for Gly-X-Y stretches with no imino acids. However, the results reported here indicate that these interrupted regions have some conformational perturbation not seen for imino acid poor regions (Li et al., 1993), which could manifest itself in either flexibility or a looping-out that could be important for functions such as binding.

The studies of interrupted peptides may also clarify the role of glycine substitutions in fibrillar collagen diseases. Glycine substitutions are the only interruptions that have been observed in fibrillar collagens, and these are seen only as mutations in pathological conditions (Kuivaniemi al., 1991). The best characterized case is osteogenesis imperfecta, where more than 60 cases of glycine substitutions in the α1 and α2 chains of type I collagen have been reported (Kuivaniemi et al., 1991). Our studies show that the single glycine to alanine substitution is the least disruptive of all kinds of interruptions. Collagens with a single glycine substitution isolated from osteogenesis imperfecta (o.i.) fibroblasts often show a decrease in thermal stability of several degrees. If the total entropy of a typical human fibrillar collagen molecule were decreased by that found for the Gly→Ala peptide (63.5 kcal/mol), and the enthalpy were to remain the same (as observed for the Gly→Ala peptide within experimental error), the thermal stability of a collagen molecule with a glycine substitution would be decreased by about 2 °C. This is close to that observed for a number of mutant o.i. collagens. The observation of variations in the clinical severity of similar glycine substitutions at different sites in type I collagen chains has led to the suggestion that the local environment of the interruption may be important. For instance, Bächinger and Davis (1991) suggested that substitutions occurring in very stable regions of the triple matrix may lead to greater destabilization than those occurring in regions that are somewhat less stable. Our peptide studies address the effect of breaks in the most stable triple-helical environment in a homotrimeric molecule, but it will be important to clarify the impact of context as well as having two or three distinct chains for the interruptions to understand collagen diseases and nonfibrillar collagens.

In addition to direct physical effects of the glycine substitution on the triple helix, the pathological consequences of o.i. collagens may result from abnormal folding (Bonadio & Byers, 1985). A recent report of an abnormally strong binding of monomer o.i. collagen with a large deletion to protein disulfide-isomerase (Chessler & Byers, 1992) supports the involvement of a folding aberration. Although the *in vitro* folding of our peptides is affected little by the interruptions, changes appear to take place in monomer conformation as a result of the interruption, in addition to changes in trimer conformation. Such alterations in monomer structure as a consequence of a substitution or deletion could affect its interactions with other molecules in the *in vivo* folding process. Investigations are in progress to define the monomer conformation of both normal and interrupted peptides.

ACKNOWLEDGMENT

We thank Karen Hempel for assistance in carrying out CD studies and Ming Hua Li for discussion of her results.

REFERENCES

Andreeva, N. S., Debabov, V. A., Millionova, M. I., Shibnev, V. A., & Chirgadze, Y. N. (1961) *Biophysics 6*, 85–86.

Bächinger, H. P., & Davis, J. M. (1991) *Int. J. Biol. Macromol. 13*, 152–156.

Benson, B., Hawgood, S., Schilling, J., Clements, J., Bamm, D., Cordell, B., & White, R. T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6379–6383.

Bhatnagar, R. S., Scaria, P. V., & Sorensen, K. R. (1990) Fourth Symposium of the Protein Society, M16, San Diego, CA.

Bonadio, J., & Byers, P. H. (1985) *Nature 316*, 363–366.

Brazel, D., Oberbäumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzman, R., & Kühn, K. (1987) *Eur. J. Biochem. 168*, 529–536.

Brodsky, B., Li, M. H., Long, C. G., Apigo, J., & Baum, J. (1992) *Biopolymers 32*, 447–451.

Burjanadze, T. V. (1982) *Biopolymers 21*, 1587–1595.

Chessler, S. D., & Byers, P. H. (1992) *J. Biol. Chem. 267*, 7751–7757.

Chu, M. L., Conway, D., Pan, T. C., Baldwin, C., Mann, K., Deutzmann, R., & Timpl, R. (1988) *J. Biol. Chem. 263*, 18601–18606.

Cohn, E. J., & Edsall, J. T. (1943) in *Proteins, Amino Acids, & Peptides*, Reinhold Publishing, New York.

Davis, J. M., Boswell, B. A., & Bächinger, H. P. (1989) *J. Biol. Chem. 264*, 8956–8962.

Dölz, R., Engel, J., & Kühn, K. (1988) *Eur. J. Biochem. 178*, 357–366.

Engel, J., Chen, H. T., Prockop, D. J., & Klump, H. (1977) *Biopolymers 16*, 601–622.

Heidemann, E., & Roth, W. (1982) *Adv. Polym. Sci. 43*, 143–203.

Hofmann, H., Fietzek, P. P., & Kühn, K. (1978) *J. Mol. Biol. 125*, 137–165.

Johnson, M. L., Correia, J. J., Halvorson, H. R., & Yphantis, D. A. (1981) *Biophys. J. 36*, 575–583.

Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P., & Krieger, M. (1990) *Nature 343*, 531–535.

Koliakos, G. G., Kouzi-Koliakos, K., Furcht, L. T., Reger, L. A., & Tsilibary, E. (1989) *J. Biol. Chem. 264*, 2313–2323.

Kuivaniemi, H., Tromp, G., & Prockop, D. J. (1991) *FASEB J. 5*, 2052–2060.

Li, M. H. (1993) Ph.D. Thesis, Rutgers University.

Li, M. H., Fan, P., Brodsky, B., & Baum, J. (1993) *Biochemistry 32*, 7377–7387.

Linsenmayer, T. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., Ed.) pp 7–44, Plenum Press, New York.

Long, C. G., Li, M. H., Baum, J., & Brodsky, B. (1992) *J. Mol. Biol. 225*, 1–4.

Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers 26*, 1601–1620.

Nemethy, G., Leach, S. J., & Scheraga, H. A. (1966) *J. Phys. Chem. 70*, 998–1004.

Oka, S., Nobuyuki, I., Kawasaki, T., & Yamashina, I. (1987) *J. Biochem. 101*, 135–144.

Pace, C. N., Shirley, B. A., & Thomson, J. A. (1989) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 311–330, IRL Press, Oxford.

Pettett, J., & Kingston, I. B. (1991) *J. Biol. Chem. 266*, 16149–16156.

Piez, K. A., & Sherman, M. R. (1970) *Biochemistry 9*, 4129–4133.

Porter, R. R., & Reid, K. M. B. (1979) *Adv. Protein Chem. 33*, 1–71.

Privalov, P. L. (1982) *Adv. Protein Chem. 35*, 1–103.

Privalov, P. L., & Tiktopulo, E. I. (1970) *Biopolymers 9*, 127–139.

Ramachandran, G. N., & Kartha, G. (1955) *Nature 176*, 593–595.

Rapaka, R. S., & Bhatnagar, R. S. (1976) *Int. J. Pept. Protein Res. 8*, 371–377.

Reid, K. B. M., Lowe, D. M., & Porter, R. R. (1972) *Biochem. J. 130*, 749–763.

Rich, A., & Crick, F. H. C. (1955) *Nature 176*, 915–916.

Rogulenkova, V. N., Millionova, M. I., & Andreeva, N. S. (1964) *J. Mol. Biol. 9*, 253–254.

Sage, H., & Iruela-Arispe, M. L. (1990) *Ann. N.Y. Acad. Sci. 580*, 17–31.

Sakakibara, S., Inouye, K., Shudo, K., Yasua, K., Kobayashi, Y., & Prockop, D. J. (1973) *Biochim. Biophys. Acta 303*, 198–202.

Saus, J., Quinones, S., MacKrell, A., Blumberg, B., Muthukumaran, G. Pihlajaniemi, T., & Kurkinen, M. (1989) *J. Biol. Chem. 264*, 6318–6324.

Schumaker, V. N., Poon, P. H., Seegan, G. W., & Smith, C. A. (1981) *J. Mol. Biol. 148*, 191–197.

Soininen, R., Hakaarisku, T., Prockop, D. J., & Tryggvason, K. (1987) *FEBS Lett. 255*, 188–194.

Traub, W., & Piez, K. A. (1971) *Adv. Protein Chem. 25*, 243–352.

van der Rest, M., & Garrone, R. (1991) *FASEB J. 5*, 2814–2823.

Yamaguchi, M., Benya, P. D., van der Rest, M., & Ninomiya, Y. (1989) *J. Biol. Chem. 264*, 16022-16029.