Detection of cis and trans X-Pro peptide bonds in proteins by 13C NMR: Application to collagen
(proline cis–trans isomerism/protein folding/collagen/13C NMR)

SUSANTA K. SARKAR*, P. E. YOUNG‡, C. E. SULLIVAN*, and D. A. TORCHIA*

*Mineralized Tissue Research Branch, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20205; and ‡Department of Natural Sciences, York College, City University of New York, Jamaica, NY 11451

Communicated by F. A. Bovey, April 20, 1984

ABSTRACT Heretofore the complexity of natural abundance spectra has precluded the use of 13C NMR to detect cis peptide bonds in proteins. We have incorporated [4,13C]proline into chicken calvaria collagen and report here well-resolved C\(^{\gamma}\) signals, arising from cis and trans X-Pro and X-Hyp peptide bonds (where X is any amino acid residue) in the 13C NMR spectrum of the thermally unfolded protein. Measurement of 13C signal areas shows that 16\% of the X-Pro and 8\% of X-Hyp bonds are cis in the unfolded collagen. These results strongly support the conclusion drawn from kinetic studies that cis–trans isomerization of peptide bonds is the rate-limiting step in helix propagation after nucleation. Our method can be applied to other proteins as well and should aid in testing the generality of the hypothesis of Brandts, Halvorson, and Brennan that cis–trans isomerization is the rate-limiting step in protein folding when proline is present.

Brandts et al. (1) have suggested that cis–trans isomerization is the rate-limiting step in protein folding when proline is present. Kinetic studies of the folding of a type III procollagen peptide (2, 3), type III collagen and type III p\(^{-}\)collagen (4), type I procollagen (5), and ribonuclease A (6–9) support this hypothesis. However, other investigations (10–14) indicate that globular protein folding is not necessarily rate-limited by proline isomerization. In order to extend these important studies and to check the generality of the hypothesis of Brandts et al. (1), it is necessary to have a method for measuring the number of cis and trans X-Pro peptide bonds, where X is any amino acid residue, in a wide variety of proteins. In principle, high resolution 13C NMR spectroscopy is such a method. It is well known that cis–trans isomerization of the peptide bond is slow on the NMR timescale (15), and that separate signals of the cis and trans isomers are resolved in spectra of peptides (15–17). In particular, the chemical shift of the prolyl \(\gamma\)-carbon (C\(^{\gamma}\)) resonance is diagnostic for a cis or trans X-Pro peptide bond (18). In spite of the success that the NMR method has achieved with peptides, NMR has not been useful for proteins because of the difficulty of assigning the cis and trans X-Pro signals in the complex spectrum of a protein. In this study we solved the problem of spectral assignment by incorporating [4,13C]proline (19) into collagen. We observed well resolved C\(^{\gamma}\) signals arising from cis and trans X-Pro and X-hydroxyproline (X-Hyp) peptide bonds in the spectrum of the thermally unfolded protein. [A post-translational enzymatic modification converts almost one half of the proline residues in each collagen chain into hydroxyproline residues (23).] We found that 16\% of the X-Pro and 8\% of the X-Hyp bonds are cis in unfolded collagen.

METHODS
Soluble collagen was prepared from rat-tail tendons by the procedure of Chandrakasan et al. (20). Collagen was disolved in 0.5 M acetic acid, centrifuged at 40,000 rpm, lyophilized, and dissolved in 0.05 M acetic acid for NMR experiments. The 125.75 MHz 13C NMR spectrum (see Fig. 1) was obtained on a Nicolet NT-500 spectrometer; 40,000 time-domain signals were collected in quadrature with 32,768 points per channel, a 30-kHz spectral window, and a delay time of 1 s between the 20-\(\mu\)s 90° pulses. Broad-band proton decoupling was used during data acquisition; otherwise low-level proton irradiation was used to obtain an Overhauser enhancement. The NMR sample contained 50\% \(\text{H}_2\text{O}\), which served as an internal deuterium lock. Sample temperature was maintained at 60 \(\pm\) 2°C with the spectrometer temperature controller.

[4,13C]Proline, synthesized by the method of Young and Torchia (19), was incorporated at a level of 30\% into lathyritic chicken calvaria collagen in culture. The labeled protein was prepared and characterized as described by Jelinski and Torchia (21), and the protein was hydrolyzed in 6 M HCl at 106°C for 24 hr; 62.98-MHz 13C spectra were obtained on a home-built spectrometer (22). Twenty thousand (see Fig. 2a) and twelve thousand (see Fig. 2b) time-domain signals were collected in quadrature with 2048 points per channel, a 10-kHz spectral window, and delay times of 0.7 s (see Fig. 2a) and 4.0 s (see Fig. 2b) between the 8-\(\mu\)s 90° pulses. The signal areas of Fig. 2a did not change when larger delay times were used. Broad-band proton decoupling was used during data acquisition; otherwise either low-level or no proton irradiation was used depending on whether or not an Overhauser enhancement was desired. NMR tubes (5-mm) containing 0.2 ml of solution were used in a solenoid coil. Samples were not spun, and a field-frequency lock was not used. Each solution contained ca. 0.25% 13C urea (90\% enriched with 13C at the methyl carbon), and the methyl resonance (1 ppm, relative to external tetramethylsilane) served as an internal reference. Sample temperature was maintained constant within \(\pm\)1°C by using a Varian temperature controller.

RESULTS AND DISCUSSION
Collagen was chosen for this initial study for the following reasons. First, almost one-quarter of the 1000 amino acid residues in each collagen chain are either prolyl or hydroxyprolyl (23). Second, after nucleation, the rate-limiting step in helix formation has been assigned to cis–trans isomerization (3–5). Third, 13C-labeled amino acids can be incorporated into collagen by using chicken calvaria tissue culture (21), and a convenient synthesis of [4,13C]proline has been reported recently (19).

On the basis of studies of peptides (15–17, 24, 25), we expect that at least 80\% of the X-Pro and X-Hyp bonds in unfolded collagen are trans. This expectation is confirmed by the prominent signals observed at 24 ppm (primarily due to trans X-Pro C\(^{\gamma}\)) and 69.5 ppm (due to trans X-Hyp C\(^{\gamma}\)) in the 125.75-MHz natural abundance spectrum of thermally un-

Abbreviations: Hyp, hydroxyproline; X, any amino acid residue.

4800
folded rat-tail-tendon collagen (Fig. 1). The peptide studies (15-17, 24, 25) show that the cis X-Pro C\(^\gamma\) signal overlaps signals from lysine, leucine, and isoleucine residues at ca. 22 ppm. In similar fashion, the cis X-Hyp C\(^\gamma\) signal and the threonine C\(^\delta\) signal overlap at ca. 67.5 ppm. Therefore, in spite of the fact that the natural abundance spectrum was obtained at the highest available field strength, we were unable to determine if unfolded collagen contains a small percentage of cis peptide bonds.

In contrast with this result, the four signals in the spectrum (Fig. 2a) of the \(^{13}\)C-labeled thermally unfolded chicken calvaria collagen were readily assigned to the X-Pro and X-Hyp cis and trans isomers using the peptide data (15-17, 24, 25). Signals from cis and trans isomers will not be observed in a spectrum obtained after peptide bonds are hydrolyzed. Therefore, the observation of one C\(^\gamma\) resonance for proline and one for hydroxyproline in the spectrum of the acid-hydrolyzed protein (Fig. 2b) strongly supports the C\(^\gamma\) assignments in the spectrum of the unfolded protein.

From the measured areas of the four signals in Fig. 2a, we calculate that 8 \(\pm\) 2\% of the X-Hyp and 16 \(\pm\) 2\% of the X-Pro bonds are cis in unfolded collagen at pH 3.5 and 43\(^\circ\)C. Within experimental error, the same results were obtained at pH 7 and 60\(^\circ\)C and when spectra were obtained without a nuclear Overhauser enhancement. The percentage of cis peptide bonds found in unfolded collagen agrees with results reported for several model polypeptides (16), and the direct observation of cis peptide bonds in unfolded collagen strongly supports the conclusion of the kinetic studies (3-5) that cis–trans isomerization is the rate-limiting step in helix propagation.

From the percentages of cis X-Pro and X-Hyp peptide bonds measured in the NMR experiments, the equilibrium constant, \(K = [\text{trans}] / [\text{cis}]\) per tripeptide (3), was calculated to be 11. This value is 2-6 times larger than the values of \(K\) obtained from the kinetic studies of collagen helix formation (3-5).

This discrepancy is not as serious as it appears because the kinetic studies of collagen helix formation accurately determine the product of \(K\) and the rate constant, \(k\), of cis–trans isomerization. Except in the case of a procollagen peptide (3), the individual values of \(k\) and \(K\) obtained from the kinetic studies have large uncertainties (4-5). The proposed model (3-5) of helix formation—zipper-like propagation from a single nucleus—could be tested more thoroughly with accurate values of \(k\) and \(K\). The method described herein provides an accurate value of \(K\), and it may be possible to measure \(k\) by saturation-transfer (17), since \(T_1\) values for the

---

**Fig. 1.** Natural abundance 125.75-MHz \(^{13}\)C NMR spectrum of rat-tail-tendon collagen (40 mg/ml) in 0.05 M acetic acid at 60\(^\circ\)C. C\(^\gamma\) assignments: trans-Pro, 24.0 ppm [overlaps signals of other amino acid sidechains (15–17, 24, 25)]; trans-Hyp, 69.5 ppm; cis-Pro and -Hyp signals, if any, are obscured by signals of other amino acids (see text). Chemical shifts are relative to external tetramethylsilane.
C\(^+\) carbons in unfolded collagen exceed 0.2 s in the 50–90°C temperature range.

The NMR procedure used herein is widely applicable to unfolded proteins because \(^{13}\)C-labeled proline can be incorporated into many proteins by a variety of biosynthetic methods (26–28). The present work indicates that in a protein containing 3% proline, a signal from 5% cis X-Pro peptide bonds can be detected with a signal-to-noise ratio of 10 in an overnight data accumulation on 10 mg of protein 50% enriched in \(^{13}\)Cproline. Linewidths in native proteins increase as molecular weight increases, and this limits application of the present method to native proteins having molecular weights less than ca. 10\(^5\). Because of its high molecular weight (3 \times 10^5) and low solubility, we have not obtained spectra of helical collagen, which has all peptide bonds trans (23). Signal assignments in native proteins will be difficult if the local environment of the proline sidechain strongly perturbs the chemical shift of the C\(^+\). This point can be checked by obtaining spectra of labeled native proteins having known three-dimensional structures. The available polypeptide data (15–17, 24, 25) indicate that assignments in labeled unfolded proteins should be straightforward as found with collagen.

Biophysics: Sarkar et al.  


