# NMR of Silk Fibroin. 3. Assignment of Carbonyl Carbon Resonances and Their Dependence on Sequence and Conformation in *Bombyx mori* Silk Fibroin Using Selective Isotopic Labeling

**Tetsuo Asakura,* Yasuo Watanabe, and Takuro Itoh**

*Department of Biopolymer Engineering, Faculty of Technology, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184, Japan.* *Received March 1, 1984*

ABSTRACT: The sequential and conformational analyses of *Bombyx mori* silk fibroin have been performed by $^{13}C$ NMR spectroscopy. Sixteen peaks were observed in the carbonyl region of the silk fibroin spectrum. These peaks were classified into the carbonyl carbons of five kinds of amino acid residues from a comparison of the spectra among [1-$^{13}C$]Ala-labeled, [1-$^{13}C$]Gly-labeled, and unenriched silk fibroins. The observation of satellite peaks due to the $^{13}C$–$^{15}N$ direct spin coupling in the carbonyl region of [$^{15}N$]Gly-labeled fibroin indicated the presence of the -X-Gly- sequence, where X = Ala or Ser residue. A further detailed assignment was performed from a comparison of the silk fibroin spectrum with that of the crystalline fraction, i.e., the precipitated fraction, after chymotrypsin hydrolysis of the fibroin. Most of the peaks were assigned to the primary sequence (Table I in the text). The dispersions of the chemical shifts due to sequence were as large as 0.7 ppm for the carbonyl carbons of Ala and Gly residues. In addition, the internal rotation angles, $\phi$, around the N–$C^\alpha$ bonds of the Ala and Ser residues were determined from the long-range coupling constant $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$ in the Gly carbonyl resonances assigned to the *–Gly–*Ala– or *–Gly–*Ser– sequences in the aqueous solution. The use of [1-$^{13}C$]Gly-labeled silk fibroin and careful NMR observation made it possible to determine the coupling constant directly with an experimental error of ca. 0.1 Hz for silk protein. After correcting $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$ for the doublet spacings in the case of overlapping components, the values were in fair agreement with the corresponding coupling values for Ac-Ala-NDMe and Ac-Ser-NDMe in deuterium oxide. These facts lead to the conclusion that the flexibility of the main chain of random coil fibroin originates essentially from the possibilities of rotation about the N–$C^\alpha$ and $C^\alpha$–C′ single bonds adjacent to the $\alpha$-carbons in aqueous solution.

## Introduction

In the class Insecta, many species produce long silken filaments to form substantial cocoons in which they pupate. Wide variations in amino acid composition and structure have been found among different silks[1] and their study offers a unique possibility for obtaining an understanding of the effect of composition and sequence on conformation in proteins.

Silk fibroins from *Bombyx mori* and *Philosamia cynthia ricini* yield very sharp $^{13}C$ NMR spectra in aqueous solution[2] and also in the gellike state (liquid silk stored in the middle silk gland of the silkworm)[3,4] despite a fairly high molecular weight.[5] This is due to the high mobility of the chain, which is characterized by a very small correlation time, on the order of $10^{-10}$ s, for the segmental motion at 40 °C.[2] In addition, the amino acid composition determined from the NMR spectra was very close to that for silk fibroins determined from amino acid analysis,[5] indicating that both crystalline and noncrystalline domains in the chain were observable simultaneously in the NMR spectra.[2] This spectral behavior presents a striking contrast to other fibrous proteins such as collagen[6–10] or elastin.[7,11,12]

The conformation of *P. c. ricini* silk fibroin was analyzed on the basis of the "doublet peak" observations which were assigned to $\alpha$-helix and random coil for the $C^\alpha$, $C^\beta$, and carbonyl carbons of the Ala residue.[4] In addition, the absence of an $\alpha$-helical portion in *B. mori* silk fibroin was deduced from a spectroscopic comparison of *P. c. ricini* and *B. mori*.[4] It is striking that high-resolution $^{13}C$ NMR observation was possible directly on the silk fibroin stored in the silk gland of intact silkworm, as well as in the solution[2] and solid states (the latter by CP/MAS NMR).[13,14] It should be emphasized that NMR on the intact gland is a novel tool for conformational characterization when the conformational change occurs readily under weak external forces such as dryness or weak mechanical stress in the sample preparation.[15] Thus NMR study of silk fibroins yields unique and detailed insight concerning the relationship between the amino acid composition, primary sequence, and conformation of proteins.

In this paper, sequential and conformational analyses of *B. mori* silk fibroin were performed by $^{13}C$ NMR spectroscopy. To begin with, the carbonyl peaks are assigned to the primary sequence from biosynthetic labeling of the carbonyl carbon and from a change in the spectrum after chymotrypsin hydrolysis. Although it has been pointed out that the sequence effects in the $^{13}C$ spectra of peptides are very small, except for the proline residue,[16] most of the carbonyl peaks were assigned to the sequence (mainly pentapeptide) in the *B. mori* silk fibroin spectrum. The chemical shift dispersions due to the sequence were as large as 0.7 ppm for Ala and Gly carbonyl carbons.

The determination of the internal rotation angles, $\phi$, around the N–$C^\alpha$ bonds of the Ser and Ala residues was attempted from the observation of the long-range coupling constant between $^{13}C$ and $^1H$ nuclei, $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$, in the $^1H$-coupled carbonyl peaks assigned to the sequence of [1-$^{13}C$]Gly-labeled silk fibroin of *B. mori*. This was based on the fact that the coupling constant $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$ yields data for the angle $\phi$ from the Karplus-like nature of its angular dependence.[17–19] The use of isotopically labeled silk fibroin and the careful NMR observation make it possible to determine the coupling constant with an experimental error of ca. 0.1 Hz for silk protein directly. A comparison of the angles $\phi$ determined for the Ala and Ser residues of *B. mori* silk fibroin with those of the dipeptide model compounds Ac-Ala-NDMe and Ac-Ser-NDMe indicates that the flexibility of the main backbone of the random coil fibroin chain originates essentially from the possibilities of rotation about the N–$C^\alpha$ and $C^\alpha$–C′ single bonds adjacent to the $\alpha$-carbons.

## Experimental Section

**Materials.** A [1-$^{13}C$]Ala-labeled cocoon from *B. mori* was prepared biosynthetically by feeding [1-$^{13}C$]alanine (90% enrichment, Dai-ichi Pure Chemicals Co., Tokyo) in addition to an

0024-9297/84/2217-2421$01.50/0 

artificial diet (Silk Mate 2M, Nippon Nosan Kogyo Co., Tokyo) to silkworm larvae (a hybrid between strains Nichi 140 and Shi 140) of the fifth instar for a whole day. [1-$^{13}$C]Gly-labeled and [$^{15}$N]Gly-labeled cocoon samples from *B. mori* were obtained in a manner similar to that for the [1-$^{13}$C]Ala-labeled cocoon (enrichments were 92% and 95% for [1-$^{13}$C]glycine and [$^{15}$N]glycine, respectively, Shoko Tsusho Co., Tokyo). The regenerated silk fibroin solution was prepared as described in a previous paper.[2] The crystalline fraction ($C_p$ fraction) of *B. mori* silk fibroin was prepared as follows.[20] Chymotrypsin (40 mg, Seikagaku Kogyo Co., Tokyo) dissolved in a few milliliters of water, was added to an aqueous solution of about 4 g of fibroin buffered at pH 7.8 with $Na_2HPO_4 \cdot 12H_2O$ and $NaH_2PO_4 \cdot 2H_2O$. The solution (200 mL) was incubated at 40 °C for 24 h, and the precipitate that formed ($C_p$ fraction) was separated by centrifuging at 10000 rpm followed by washing with 0.03 N HCl to inactivate the enzyme reaction. Then the precipitate was washed several times with distilled water, ethyl alcohol, and ethyl ether. Finally, the precipitate was dried in vacuo (ca. 55% of the original fibroin). Lithium bromide, both extrapure grade and superpure grade, was purchased from Wako Chemicals Co., Oosaka, and Merck Co., Inc. Rahway, NJ, respectively. The $(Ala\text{-}Ala\text{-}Gly)_n$ sample was a generous gift from Dr. A. Shoji of the College of Technology, Gunma University, Kiryu, Gunma, and the synthetic process was described elsewhere.[14] Ac-Ala-NHMe and Ac-Ser-NHMe were synthesized as described in a previous paper.[2]

**$^{13}$C NMR Measurement.** $^{13}$C NMR spectra were recorded with a JEOL FX-200 NMR spectrometer operating at 50.3 MHz. All spectra were taken at a probe temperature of 40 °C, maintained by a variable-temperature unit. The sample concentration was 2.2–10.0% (w/v) and a small amount of $D_2O$ (5%) was added to the solution. Before the observation of $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$ the aqueous solution of regenerated silk fibroin was dialyzed against $D_2O$ for 20 h in order to exchange the labile NH protons with solvent deuteron completely. The pH value of the solution was determined by using a TP-101 pH meter (Toko Chemical Laboratories, Tokyo) with a Toa CE-103 C combination micro glass electrode. For most $^{13}$C NMR observations, spectral conditions were the following: 3000–13000 pulses, 45° pulse angle (10 μs), 4.2–6.0-s delay between pulses, 1000-Hz spectral width, 8K data points, $^1$H noise decoupling. The $^1$H-coupled $^{13}$C spectra for the $^3J_{C'\text{-}N\text{-}C^\alpha\text{-}H}$ determination were recorded with gated proton decoupling (gated *off* during the data acquisition time but *on* during the pulse delay). For other observations the spectral conditions were as follows: 550–600 pulses, 90° pulse width (20 μs), 10.6-s delay between pulses, 1000-Hz spectral width, 8K data points supplemented with 8K zero points (a "resolution" of 0.12 Hz/point). Chemical shifts were measured relative to external $(CH_3)_4Si$.

## Results and Discussion

**Assignment of the Carbonyl Carbon Resonances.** In our previous paper,[2] fine structure had been observed for the carbonyl peaks of the Gly and Ala residues of *B. mori* silk fibroin. These splittings were considered to be reflections of the amino acid sequence of the silk fibroin in the random coil state since no significant changes in the spectra were observed on adding urea to the solution, contrary to the case of *P. c. ricini* silk fibroin.[4] Detailed assignment of the carbonyl peaks of *B. mori* silk fibroin due to sequence is performed here as follows.

Figure 1 shows the carbonyl region of the $^{13}$C NMR spectra of *B. mori* silk fibroin (spectrum A) together with the isotopically labeled (spectra B–D) silk fibroins in aqueous solution. There are at least 16 peaks, which are labeled a–p with increasing field in spectrum A. Peaks c and n were observed as shoulders on peaks b and m, respectively. In addition, considerable broadening and some further splitting were observed for the peaks o and p. The chemical shifts and relative intensities[21] are summarized in Table I. As shown in Figure 1, a comparison of the spectrum of [1-$^{13}$C]Ala-labeled silk fibroin, spectrum B, and that of unenriched fibroin, spectrum A, indicates clearly that peaks a–f are attributable to Ala residues.

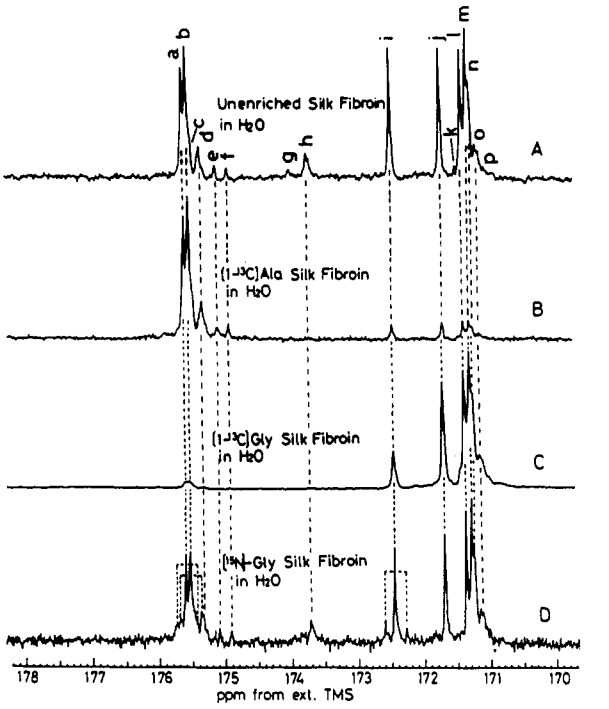


**Figure 1.** Comparison of $^{13}$C NMR spectra of the carbonyl region among (A) unenriched (sample concentration 8.1% (w/v), 6870 pulses), (B) [1-$^{13}$C]Ala-labeled (sample concentration 4.4% (w/v), 90 pulses), (C) [1-$^{13}$C]Gly-labeled (sample concentration 3.7% (w/v), 150 pulses), and (D) [$^{15}$N]Gly-labeled silk fibroins (sample concentration 2.8% (w/v) 13000 pulses) of *B. mori* in aqueous solution at 40 °C. The peaks were labeled from a to p with increasing field in spectrum A.

**Table I**
**Chemical Shift, Relative Intensity, and Assignment of the Carbonyl Carbons of *B. mori* Silk Fibroin**

| peak[a] | chem shift obsd[b] | chem shift ref[c] | rel intens[d] | residue | sequence |
|---|---|---|---|---|---|
| a | 175.59 | 176.3 | 101 | Ala | SGAGA[e] |
| b | 175.53 | | 101 | Ala | AGAGS[e] |
| c | 175.5 | | 39 | Ala | AGAGT |
| d | 175.32 | | 40 | Ala | TGAGA |
| e | 175.09 | | 10 | Ala | TGAGV[e] |
| f | 174.90 | | 9 | Ala | VGAGT[e] |
| g | 173.98 | 174.8 | 7 | Val | AGVGA |
| h | 173.70 | | 40 | Tyr | AGTGA |
| i | 172.45 | 173.1 | 100 | Ser | AGSGA |
| j | 171.69 | 172.7 | 96 | Gly | GAGSG |
| k | 171.45 | | 3 | Gly | *f* |
| l | 171.37 | | 104 | Gly | GSGAG |
| m | 171.28 | | 100 | Gly | GAGAG |
| n | 171.24 | | 76 | Gly | GTGAG[e] |
| o | 171.12 | | 57 | Gly | GAGTG[e] |
| p | 170.95 | | 21 | Gly | *f* |

[a] The carbonyl peaks were labeled a–p toward higher field in spectrum A (Figure 1). [b] Ppm from external $(CH_3)_4Si$. [c] Gly-Gly-X-Gly-Gly, where X = Gly, Ala, Ser, and Val. The chemical shifts were calibrated from internal $CS_2$ and corrected to internal $(CH_3)_4Si$ reference.[25] [d] The intensity of the Ser C=O peak, i, was assumed to be 100. [e] The assignments were exchangeable with each other. [f] These peaks were attributable to the Gly carbonyl carbons in the noncrystalline domain, but further assignments were not performed here.

Peaks j–p are attributable to Gly residues from a similar spectroscopic comparison of spectrum C with spectrum A.[27] Thus each resonance of the Ala and Gly carbonyl carbons is spread over 0.7 ppm due to the primary sequence effect. Peaks g, h, and i are attributable to the Val, Tyr, and Ser residues, respectively, from both the relative intensities and a comparison of the chemical shifts of peaks g and i

with those of the pentapeptide Gly-Gly-X-Gly-Gly, where X = Ser and Val,[23-25] as listed in Table I. The relative intensities agree with the amino acid composition of *B. mori* silk fibroin[2] and the chemical shift data of the carbonyl carbons are close to those of the pentapeptides.[26]

In the spectrum of [$^{15}$N]Gly-labeled silk fibroin, spectrum D, one-bond spin coupling between $^{13}$C and $^{15}$N nuclei, $^1J_{^{13}C-^{15}N}$, is expected to be observed in the carbonyl resonance of the X residue if the X–Gly sequence is present in the chain[17,27] and if the $^{15}$N enrichment is more than 24%.[28] Actually, the main peaks, Ala carbonyl carbon peaks a and b and Ser carbonyl peak i, decrease in intensity at the expense of the satellite peaks due to spin coupling with the amide $^{15}$N of the Gly residue. The ratio of the sum of the areas of two satellite peaks to the area of the total peak is approximately 30%. The $^1J_{^{13}C-^{15}N}$ values are 15 and 16 Hz for peaks a and b of the Ala carbonyl carbon, respectively, and 16.5 Hz for the Ser carbonyl carbon. These values are in the usual range for peptide bonds.[27] Thus the presence of the Ala-Gly and Ser-Gly sequences in *B. mori* silk fibroin is demonstrated clearly from the $^{13}$C NMR observation of the [$^{15}$N]Gly-labeled fibroin. Peaks a and b are attributable to the Ala carbonyl carbons of the *Ala*-Gly sequence. Similarly, peak i is attributable to the Ser carbonyl carbon of the *Ser*-Gly sequence. In addition, it is clear that peaks j, l, and m, assigned to the Gly carbonyl carbons, do not involve *Gly*-Gly sequences. These assignments are consistent with those made on the basis of a comparison of the spectrum of the original fibroin with that of the crystalline fraction after chymotrypsin hydrolysis, as mentioned below. However, the assignment of small peaks is impossible from the [$^{15}$N]Gly-labeled fibroin spectrum, spectrum D, because of the poor signal-to-noise ratio.

The sequence has been reported for the crystalline fraction ($C_p$ fraction), i.e., the precipitated fraction after chymotrypsin hydrolysis of *B. mori* silk fibroin and ca. 55% of the original fibroin, as Gly-Ala-Gly-Ala-Gly-Ser-Gly-Ala-Ala-Gly-[Ser-Gly-(Ala-Gly)$_n$]$_8$-Tyr, where $n$ is usually 2.[29] A further detailed peak assignment of the carbonyl region of the *B. mori* silk fibroin spectrum due to the sequence is possible by reference to the sequence analysis data of the Cp fraction. Since the Cp fraction precipitated immediately in the dialysis of the LiBr solution against water, the spectroscopic comparison of original fibroin and the $C_p$ fraction was performed in 9 M LiBr solution. Figure 2 shows the carbonyl region of the $^{13}$C NMR spectra of *B. mori* silk fibroin in aqueous solution, spectrum A (this spectrum is the same as spectrum A in Figure 1), and in 9 M LiBr, spectrum E, together with those of the $C_p$ fraction, spectrum F (pH 6.5), and a sequential copolymer (Ala-Ala-Gly)$_n$, spectrum G, in 9 M LiBr. The individual peaks in the carbonyl region of the original silk fibroin could be followed as represented by a dotted line (spectra A and E) from the observation of the chemical shift dependence on LiBr concentration.

To begin with, the assignment of main peaks (peaks a, b, j, l, and m) was performed. In the spectrum of the $C_p$ fraction, spectrum F, the relative intensities of the three Gly peaks, j, l, and m, were equal and close to that of the Ser peak, i. In addition, the intensity of each peak was approximately half that of the main peak, a + b, of the Ala carbonyl carbon. Thus the Ala peak, a + b, is attributable to the Gly-*Ala*-Gly sequence and the three Gly peaks, j, l, and m, are attributable to three kinds of Gly carbonyl carbon, taking into account the sequence of the $C_p$ fraction, approximately (*Gly*-Ser-*Gly*-Ala-*Gly*-Ala)$_n$. Among the three kinds of Gly carbonyl carbons, it is expected that the difference in the chemical shifts of the Gly-Ser-*Gly*-Ala-Gly and Gly-Ala-*Gly*-Ala-Gly sequences will be relatively small. In addition, the chemical shift of the peak m is close to that of the Ala-*Gly*-Ala carbonyl carbon of a sequential copolymer (Ala-Ala-Gly)$_n$. Thus, peaks j, l, and m are attributable to the Gly-Ala-*Gly*-Ser-Gly, Gly-Ser-*Gly*-Ala-Gly, and Gly-Ala-*Gly*-Ala-Gly sequences, respectively. Similarly, it is clear that the Ser peak, i, is attributable to the Ala-Gly-*Ser*-Gly-Ala sequence. Since there are no differences in the chemical shifts of these main peaks between the original silk fibroin and the $C_p$ fraction in 9 M LiBr, the assignments made in the $C_p$ fraction spectrum are essentially retained in the spectrum of *B. mori* silk fibroin in 9 M LiBr. Moreover, the same assignments also apply to the original silk fibroin in aqueous solution except for the assignment of the Ala carbonyl carbon. The Ala singlet peak, a + b, observed in 9 M LiBr splits into a doublet, a and b, with equal intensities in aqueous solution (spectra A and E). Thus each peak is attributable to either Ala-Gly-*Ala*-Gly-Ser or Ser-Gly-*Ala*-Gly-Ala sequences in the chain. These assignment agree with the conclusion derived from [$^{15}$N]-Gly-labeled silk fibroin.

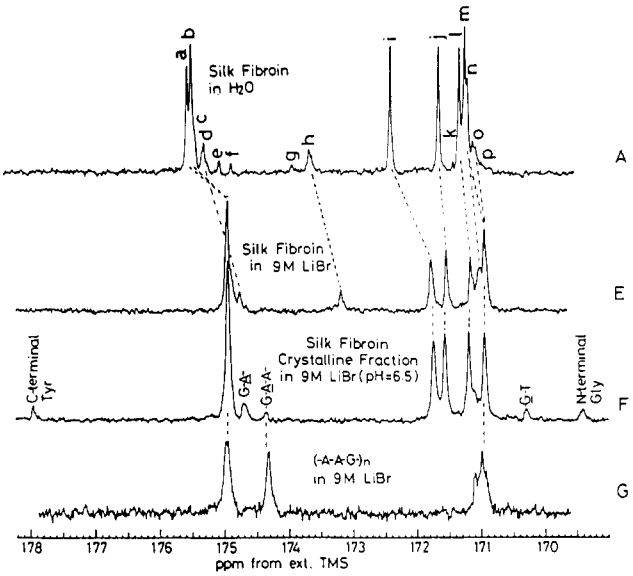


**Figure 2.** Comparison of $^{13}$C NMR spectra of the carbonyl region among (E) *B. mori* silk fibroin (3450 pulses), (F) the crystalline fraction of the silk fibroin at pH 6.5 (9250 pulses), and (G) (Ala-Ala-Gly) copolymer (3100 pulses) in 9 M LiBr at 40 °C. The spectrum of *B. mori* silk fibroin in aqueous solution, spectrum A (Figure 1), is also shown.

We then turned to the assignment of the small peaks. Since bond cleavage occurred mostly at the carbonyl of the Tyr residue in chymotrypsin hydrolysis of *B. mori* silk fibroin,[5,20,30] the assignments of the sequences involving the Tyr residue in the fibroin chain are also possible from a spectroscopic comparison of silk fibroin and the chymotrypsin hydrolysis product. First, in order to assign small peaks in the spectrum of the $C_p$ fraction, the chemical shifts of the carbonyl carbons were observed in 9 M LiBr as a function of pH (Figure 3). The peaks of C-terminal Tyr and N-terminal Gly carbonyl carbons were readily assigned from the titration curve. In addition, the peak at 174.50 ppm (pH 2) shifts downfield by 0.9 ppm when the N-terminal Gly peak shifts downfield by 4.5 ppm in the range of pH 4–8. Thus this peak is attributable to the Gly-*Ala*- carbonyl carbon. Similarly, since the peak at 170.36 ppm (pH 2) shifts upfield slightly (by 0.1 ppm) when the C-terminal Tyr peak shifts downfield by 0.9 ppm in the range of pH 2–3.5, this peak is attributable to the

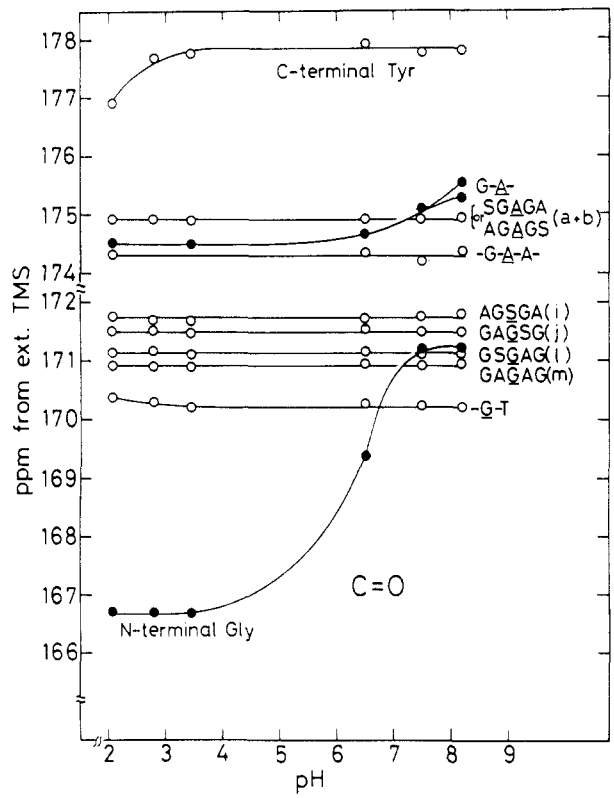


**Figure 3.** pH dependence of the carbonyl chemical shifts of the crystalline fraction of *B. mori* silk fibroin in 9 M LiBr at 40 °C.

-*Gly*-Tyr carbonyl carbon. In addition, the peak at 174.32 ppm was assigned to the Ala carbonyl carbon of the -Gly-*Ala*-Ala-sequence from a comparison of the spectra F and G and from the sequence of the $C_p$ fraction described above.

The soluble fraction ($C_s$ fraction) after chymotrypsin hydrolysis has been subdivided into four fractions, $S_I$, $S_{II}$, $S_{III}$, and $S_{IV}$, and only partial sequence analysis has been reported by Shimura et al.[5] The sum of the $S_I$ and $S_{III}$ fractions constituted ca. 80% of the $C_s$ fraction. The $S_I$ fraction (32%) has approximately the same sequence as the $C_p$ fraction, while the $S_{III}$ fraction (46%) contains significant amounts of Gly-Ala-Gly-Ala-Gly-Ala-Gly-Tyr and Gly-Ala-Gly-Val-Gly-Ala-Gly-Tyr sequences. Therefore, the N-terminal and C-terminal sequences are considered to be approximately Gly-Ala-Gly-Ala-... and ...-Gly-Ala-Gly-Tyr, respectively, for the $C_s$ fraction as well as the $C_p$ fraction. Since the chemical shift of peak d coincides approximately with that of the Gly-*Ala*- peak of the $C_p$ fraction in the range of pH 2–6.5 in 9 M LiBr (spectra E and F), peak d is attributable to the Tyr-Gly-*Ala*-Gly-Ala sequence. Actually, the relative intensities of peaks d and h are the same as listed in Table I, which supports the assignment of peak d. Another peak, c, which is also considered to be due to the sequence involving the Tyr residue (the intensity is also same as that of peak h) is attributable to the Ala-Gly-*Ala*-Gly-Tyr sequence.

New peaks, i.e., N-terminal Gly and -*Gly*-Tyr, were observed instead of the disappearance of peaks n and o (spectra E and F). Therefore peaks n and o may be attributable to either Gly-Ala-*Gly*-Tyr-Gly or Gly-Tyr-*Gly*-Ala-Gly sequences. If it is assumed that the chemical shift of the Gly-Ala-*Gly*-Ala-Gly sequence is close to that of the Gly-Tyr-*Gly*-Ala-Gly sequence rather than the Gly-Ala-*Gly*-Tyr-Gly sequence, peaks n and o are attributable to the Gly-Tyr-*Gly*-Ala-Gly and Gly-Ala-*Gly*-Tyr-Gly sequences, respectively. However, since the intensity of peak n or o is appreciably stronger than that of the Tyr residue,

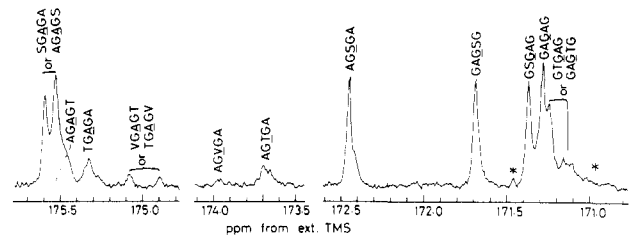


**Figure 4.** Assignment of the carbonyl region in *B. mori* silk fibroin spectrum. The peaks designated with asterisks were attributable to the Gly carbonyl carbons in the noncrystalline domain, but further assignments were not performed (see the text).

h, as listed in Table I, these peaks involve significant amounts of other sequences in the noncrystalline domain of the fibroin chain.

The intensities of peaks e and f in the Ala carbonyl resonance region are similar and close to the intensity of the Val carbonyl carbon, g (Table I). Since the presence of the Gly-Ala-Gly-Val-Gly-Ala-Gly-Tyr sequence has been reported in the $C_s$ fraction,[5,30] these peaks might be attributable to either Val-Gly-*Ala*-Gly-Tyr or Tyr-Gly-*Ala*-Gly-Val sequences. Although the appearance of the Ala carbonyl peak of the -Gly-*Ala*-Ala- sequence is also expected at the resonance position of peak f, the amount of this sequence becomes negligible if the $C_s$ fraction does not contain this sequence. Similarlities in the intensity and shape between peaks e and f would support this possibility. Peaks k and p of the Gly carbonyl carbon are attributable to the sequence in the noncrystalline domain, but further assignment is impossible at present. Thus most of the carbonyl peaks are assigned in the sequence of the *B. mori* silk fibroin chain as shown in Figure 4; the results are summarized in Table I.

**Conformational Analysis.** Splittings of the peaks in the carbonyl region of the *B. mori* silk fibroin spectrum were assigned to specific sequences in the previous section. The observation of the $^3J_{C'-N-C^\alpha-H}$ value in each carbonyl resonance is expected to yield further information concerning the solution conformation of silk fibroin averaged over the NMR time scale because the carbon–proton coupling constant, $^3J_{C'-N-C^\alpha-H}$, as well as the proton–proton coupling constant, $^3J_{H-N-C^\alpha-H}$, is well-known to provide information concerning the internal rotation angle $\phi$ around the N–$C^\alpha$ bond of the amino acid residue on the basis of the Karplus-type nature of its angular dependence.[17-19] However, the observation is not easy even for main peaks because of the decrease in the peak intensity of a given $^{13}C$ nucleus due to the spin coupling between $^1H$ and $^{13}C$ nuclei and because of the requirement of high spectral resolution (ca. 0.1 Hz). Isotope enrichment of the specified carbonyl carbon is very useful in overcoming these complications and we therefore used [1-$^{13}$C]Ala-labeled and [1-$^{13}$C]Gly-labeled silk fibroins. In addition, the NH protons of silk fibroin were completely exchanged with solvent deuteron before NMR observation to obtain a spectrum further simplified by the disappearance of the spin coupling between a given carbonyl carbon and the NH proton. Since a large portion of *B. mori* silk fibroin has the amino acid sequence -Gly-Ala-Gly-Ala-Gly-Ser-, the $^3J_{C'-N-C^\alpha-H}$ coupling constant observed at the carbonyl resonances of the Ala residue yields the angles $\phi$ of the Gly residue. Similar observations of the $^3J_{C'-N-C^\alpha-H}$ in the Gly carbonyl resonance yield the angle $\phi$ of the Ala residue (-*Gly*-Ala- sequence) and of the Ser residue (-*Gly*-Ser-sequence). The $^1H$-coupled and -decoupled spectra of the carbonyl carbon are shown for the Ala residue of [1-$^{13}$C]-Ala-labeled fibroin in Figure 5 and for the Gly residue of [1-$^{13}$C]Gly-labeled fibroin in Figure 6. Changes in the

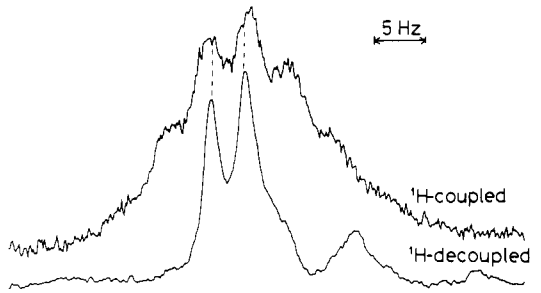


**Figure 5.** $^1$H-coupled (600 pulses) and -decoupled (90 pulses) $^{13}$C NMR spectra of the Ala carbonyl region of [1-$^{13}$C]Ala-labeled silk fibroin in $D_2O$ (concentration 4.7%) at 40 °C.

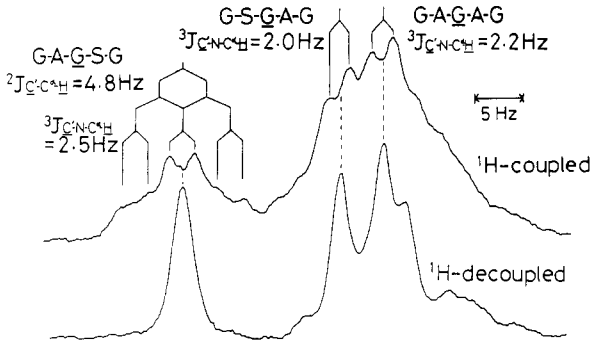


**Figure 6.** $^1$H-coupled (550 pulses) and -decoupled (150 pulses) $^{13}$C NMR spectra of the Gly carbonyl region of [1-$^{13}$C]Gly-labeled silk fibroin in $D_2O$ (concentration 2.2%) at 40 °C.

spectral pattern between $^1$H-coupled and -decoupled spectra were clearly observed for both residues. In the $^1$H-coupled [1-$^{13}$C]Ala spectrum, if the spin coupling constants $^2J_{C'-C^\alpha-H}$, $^3J_{C'-C^\alpha-C^\beta-H}$, and $^3J_{C'-N-C^\alpha-H}$ were assumed to be ca. 4 Hz,[17] appearance of roughly triplet patterns resulting from the long-range coupling between $^1$H nuclei and the Ala carbonyl carbon would be expected for each peak, Ala-Gly-*Ala*-Gly-Ser and Ser-Gly-*Ala*-Gly-Ala. This is the case in Figure 5, but determination of the accurate $^3J_{C'-N-C^\alpha-H}$ value was difficult. Therefore, the determination was performed only for the Gly peak. The coupling pattern is shown in Figure 6 by solid lines for the Gly-Ala-*Gly*-Ser-Gly carbonyl carbon. On the other hand, the coupling pattern was not clearly shown for the Gly-Ser-*Gly*-Ala-Gly and Gly-Ala-*Gly*-Ala-Gly peaks, but the value of the $^3J_{C'-N-C^\alpha-H}$ was readily obtained from the spacing of the doublet splitting. The $^3J_{C'-N-C^\alpha-H}$ values were observed as 2.5 Hz for the Gly-Ala-*Gly*-Ser-Gly sequence, 2.0 Hz for the Gly-Ser-*Gly*-Ala-Gly sequence, and 2.2 Hz for the Gly-Ala-*Gly*-Ala-Gly sequence. For a comparison, the $^3J_{C'-N-C^\alpha-H}$ values were determined for the acetyl carbonyl peaks of Ac-Ala-NDMe and Ac-Ser-NDMe in $D_2O$ as shown in Figure 7. The values were 2.5 Hz for Ac-Ala-NDMe and 2.9 Hz for Ac-Ser-NDMe. Bystrov[17] has calculated corrections due to component overlap of the doublet spacing for Lorentzian and Gaussian line shapes for obtaining the actual coupling constants. After the corrections by means of the peak-to-trough ratio of the doublet were applied to the determination of $^3J_{C'-N-C^\alpha-H}$ value of the silk fibroin, the values were determined as 2.8 Hz for Gly-Ala-*Gly*-Ser-Gly, 2.4 Hz for Gly-Ser-*Gly*-Ala-Gly, and 2.6 Hz for Gly-Ala-*Gly*-Ala-Gly sequence. Thus the values are in fair agreement with the corresponding values of Ac-Ala-NDMe and Ac-Ser-NDMe.

Using the Karplus-type relationship between the $^3J_{C'-N-C^\alpha-H}$ and the dihedral angle, $\theta$, reported by Bystrov[17]

$$^3J_{C'-N-C^\alpha-H} = 8.2\cos^2\theta - 4.4\cos\theta - 0.8\sin^2\theta$$

we obtained the angle $\theta$ and calculated the corresponding

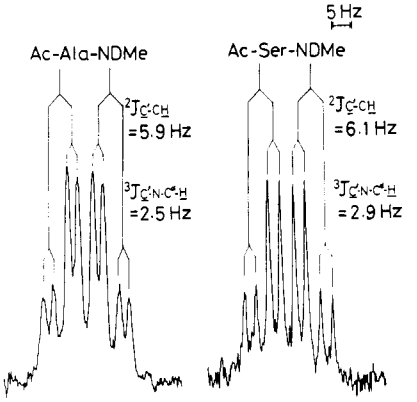


**Figure 7.** $^1$H-coupled $^{13}$C NMR spectra of the acetyl carbonyl resonances of Ac-Ala-NDMe (1000 pulses) and Ac-Ser-NDMe (1000 pulses) at 40 °C.

**Table II**
**Coupling Constants $^3J_{C'-N-C^\alpha-H}$[a] Dependent on the Dihedral Angle $\theta$ and the Internal Rotation Angle $\phi$ around the N–C$^\alpha$ Bonds of Ala and Ser Residues of [1-$^{13}$C]Gly-Labeled *B. mori* Silk Fibroin[b] and the Dipeptide Compounds[c] (Ac-Ala-NDMe and Ac-Ser-NDMe)**

| | $^3J_{C'-N-C^\alpha-H}$, Hz | $\theta$, deg | $\phi$, deg |
|---|---|---|---|
| GA*G*AG | 2.6[f] | 25, 115 | *-145*, *-95*, -5, 125 |
| GS*G*AG | 2.4 | 27, 114 | *-147*, *-93*, -6, 126 |
| Ac-Ala-NDMe[d] | 2.5 | 26, 114 | *-146*, *-94*, -6, 126 |
| GA*G*SG | 2.8 | 23, 116 | *-143*, *-97*, -4, 124 |
| Ac-Ser-NDMe[e] | 2.9 | 21, 116 | *-141*, *-99*, -4, 124 |

[a] A "resolution" was 0.12 Hz/point (1000-Hz spectral width; 8K data points supplemented with 8K zero points). [b] Determined from the Gly carbonyl carbons of GA*G*AG, GS*G*AG, and GA*G*SG sequences (concentration 2.2% (w/v)). [c] Determined from the acetyl carbonyl carbons. [d] Concentration 20% (w/v). [e] Concentration 9% (w/v). [f] The $^3J_{C'-N-C^\alpha-H}$ values were observed as 2.5 Hz for the GA*G*SG sequence, 2.0 Hz for the GS*G*AG sequence, and 2.2 Hz for the GA*G*AG sequence. After the corrections by means of the peak-to-trough ratio of the doublet[17] were applied to the determination of $^3J_{C'-N-C^\alpha-H}$ value of the silk fibroin, the values were determined as 2.8 Hz for the GA*G*SG sequence, 2.4 Hz for the GS*G*AG sequence and 2.6 Hz for the GA*G*AG sequence.

$\phi$ values (Table II). At 2.2% (w/v) solution of *B. mori* silk fibroin, the conformation is considered to be random coil.[2,31] Thus the $\phi$ values obtained for the silk fibroin should be regarded as an averaged value on the NMR time scale. On the basis of the conformational energy maps reported for Ac-Ala-NHMe and Ac-Ser-NHMe,[32-37] approximately -140° to -145° and 95° to -100° were adopted as the $\phi$ value (italicized values in Table II). We should especially emphasize that recent empirical conformational energy calculations including the effects of hydration explicitly for Ac-Ala-NHMe and Ac-Ser-NHMe indicate that the global minimum is in the D region ($-180° < \phi < -120°$)[37] for both dipeptide molecules and that the CD sign patterns observed for Ac-Ala-NHMe and Ac-Ser-NHMe in aqueous solutions could be predicted from the D region in the conformational map.[33] Thus we consider that -140° to -145° is more reasonable than -95° to -100° as the $\phi$ value in aqueous solution. These data indicate that the flexibility of the main backbone of the random coil fibroin chain in aqueous solution originates essentially from the possibilities of rotation about the N–C$^\alpha$ and C$^\alpha$–C′ single bonds adjacent to the $\alpha$-carbons.

## Concluding Remarks

It has been pointed out that the sequence effects on the $^{13}$C chemical shifts of peptides are very small except for proline residues.[16] However, peak splittings due to sequence are clearly observed in the carbonyl carbon reso-

nances of *B. mori* silk fibroin in the random coil state; i.e., the chemical shift dispersion was as large as 0.7 ppm for each Ala and Gly carbonyl carbon. Although it has been common practice to express conformation shifts in native proteins with reference to the chemical shifts of the appropriate amino acid in a simple peptide, the comparison must be done with care. The origin of the "sequence effects" in the random coil fibroin spectrum was attributed to differences in the local conformations defined by the internal rotation angles which were averaged on the NMR time scale.[38]

We are now proceeding with the conformational characterization of silk I type silk fibroin (a less stable conformation of *B. mori* silk fibroin in the solid state) by means of conformational energy calculation, CD, IR, X-ray diffraction, and NMR, including CP-MAS NMR. In conclusion, the effect of hydration on the silk I conformation is very important. The sample with silk I conformation in the solid state used for X-ray diffraction observation contains bound water.[39] The silk I conformation occurs when the liquid silk from *B. mori* is dried gradually at room temperature; silk I type fibroin converts easily to silk II type fibroin (antiparallel $\beta$-sheet structure and a more stable form in the solid state) with increasing temperature. Thus it is expected that the silk I type conformation should be closely related to the random coil conformation of silk fibroin in aqueous solution.

## References and Notes


(1) Fraser, R. D. B.; MacRae, T. P. "Conformation in Fibrous Proteins"; Academic Press: New York and London, 1973.
(2) Asakura, T.; Watanabe, Y.; Uchida, A.; Minagawa, H. *Macromolecules* **1984**, *17*, 1075.
(3) Asakura, T.; Ando, M. *Makromol. Chem., Rapid Commun.* **1982**, *3*, 723.
(4) Asakura, T.; Watanabe, Y.; Suzuki, H. *Macromolecules* **1983**, *16*, 1024.
(5) Shimura, K. "Zoku Kenshi no Kozo"; Hojo, N., Ed.; Shinshu University: Ueda, 1976.
(6) Torchia, D. A.; Lyerla, J. R., Jr.; Quattrone, A. J. *Biochemistry* **1975**, *14*, 887.
(7) Torchia, D. A.; VanderHart, D. L. "Topics in Carbon-13 NMR Spectroscopy"; Levy, G. C., Ed.; Wiley-Interscience: New York, 1976; Vol. 3, p 335.
(8) Jelinski, L. W.; Sullivan, C. E.; Torchia, D. A. *Nature* (London) **1980**, *284*, 531.
(9) Jelinski, L. W.; Torchia, D. A. *J. Mol. Biol.* **1980**, *138*, 255.
(10) Jelinski, L. W.; Sullivan, C. E.; Torchia, D. A. *J. Magn. Reson.* **1980**, *41*, 133.
(11) Lyerla, J. R., Jr.; Torchia, D. A. *Biochemistry* **1975**, *14*, 5175.
(12) Urry, D. W.; Mitchell, L. W. *Biochem. Biophys. Res. Commun.* **1976**, *68*, 1153.
(13) Saito, H.; Iwanaga, Y.; Tabeta, R.; Narita, M.; Asakura, T. *Chem. Lett.* **1983**, 427.
(14) Saito, H.; Tabeta, R.; Asakura, T.; Iwanaga, Y.; Shoji, A.; Ozaki, T.; Ando, I. *Macromolecules*, in press.
(15) Lotz, B.; Cesari, F. C. *Biochimie* **1979**, *61*, 205.
(16) Jardetzky, O.; Roberts, G. C. K. "NMR in Molecular Biology"; Academic Press: New York, 1981.
(17) Bystrov, V. F. *Prog. NMR Spectrosc.* **1976**, *10*, 41.
(18) Ando, I.; Asakura, T.; Watanabe, S. *J. Mol. Struct.* **1981**, *76*, 93.
(19) Hansen, P. E. *Prog. NMR Spectrosc.* **1981**, *14*, 175.
(20) Lucas, F.; Shaw, J. T. B.; Smith, S. G. *Biochem. J.* **1957**, *66*, 468.
(21) The $T_1$ values of the carbonyl carbons for the main peaks changed from 2.0 to 2.3 s and the NOE values also changed from 1.5 to 1.7 depending upon the amino acid residues and the sequence (Asakura, T.; Itoh, T., unpublished data). These data indicate that these carbonyl carbons were at internal positions of the silk fibroin. Thus, although the relative intensities were slightly changed if the NOE effect was taken into account, the conclusions in the text remain unchanged.
(22) An increase in intensity was observed for the Ser peak in the [1-$^{13}$C]Gly-fibroin spectrum (5.5 times) as shown in Figure 1, suggesting the partial conversion of [1-$^{13}$C]Gly to [1-$^{13}$C]Ser by a biosynthetic process through glyoxylic acid in vivo. Bricteux, S.; Fukuda, T.; Dewandre, A.; Florkin, M. *Arch. Int. Physiol. Biochem.* **1959**, *67*, 545. Fukuda, T. *J. Biochem.* **1960**, *47*, 581.
(23) Keim, P.; Vigna, R. A.; Marshall, R. C.; Gurd, R. N. *J. Biol. Chem.* **1973**, *248*, 6104.
(24) Keim, P.; Vigna, R. A.; Morrow, J. S.; Marshall, R. C. *J. Biol. Chem.* **1973**, *248*, 7811.
(25) Howarth, O. W.; Lilley, D. M. *Prog. NMR Spectrosc.* **1978**, *12*, 1.
(26) All peaks of *B. mori* silk fibroin are shifted upfield by ca. 1 ppm from the peak positions of the corresponding residue of the pentapeptides Gly-Gly-X-Gly-Gly, where X = Gly, Ala, Ser, and Val. This is considered mainly due to a difference in the chemical shift reference. Our data were calibrated from external $(CH_3)_4Si$ without correction for the volume magnetic susceptibility effect, but the reference data were calibrated from internal $CS_2$ and corrected to internal $(CH_3)_4Si$ reference.
(27) Kricheldorf, H. R. *Org. Magn. Reson.* **1980**, *14*, 455.
(28) Kainosho, M.; Tsuji, T. *Biochemistry* **1982**, *21*, 6273.
(29) Strydom, D. J.; Haylett, T.; Stead, R. H. *Biochem. Biophys. Res. Commun.* **1977**, *3*, 932.
(30) Lucas, F.; Shaw, J. T. B.; Smith, S. G. *Biochem. J.* **1962**, *83*, 164.
(31) Kobayashi, Y.; Fujiwara, T.; Kyogoku, Y.; Kataoka, K. "Abstracts of the 19th Meeting of Nuclear Magnetic Resonance in Japan", 1980, p 149.
(32) Pullman, B.; Pullman, A. *Adv. Protein. Chem.* **1974**, *16*, 347.
(33) Dungan, J. M., III; Hooker, T. M., Jr. *Macromolecules* **1981**, *14*, 1812.
(34) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. *Biopolymers* **1983**, *22*, 1885.
(35) Stern, P. S.; Chorev, M.; Goodman, M.; Hagler, A. T. *Biopolymers* **1983**, *22*, 1901.
(36) Vasquez, M.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 1043.
(37) Hodes, Z. I.; Némethy, G.; Scheraga, H. A. *Biopolymers* **1979**, *18*, 1565.
(38) Tonelli, A. *J. Am. Chem. Soc.* **1980**, *102*, 7635.
(39) Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 201.