Structural Analysis of Crystalline Fraction of *Bombyx mori* Silk Fibroin by using Advanced Solid State NMR

Doctoral Thesis By

Keiko OKUSHITA

March, 2015

Tokyo University of Agriculture and Technology

Contents

Chapter 1

Introduction

1.1. *Bombyx mori* silk fibroin with Silk II form as a heterogeneous polymer system

Silks are produced by a number of animals, predominantly spiders and moth larvae.[1-3] The best known moth silk comes from the pupa of the domesticated silk worm *Bombyx mori*, which it winds into a cocoon to protect the pupa as it develops into the mature moth. Its function thus requires it to be tough rather than strong[4,5]: it has to be able to resist attack from biting and pecking insects and birds.[6] This makes it a useful fiber for textiles, for which it is produced in quantities of around 100,000 tons pa. Its excellent biocompatibility also makes it useful for biomaterials.[7,8] In the last few years, many attempts have been made to mimic the natural process of producing robust silk filament under mild environmental conditions.[9-12] This has proved challenging, and the resultant fibers have been considerably weaker than native silk, but the potential is enormous. Thus, an understanding of the structural propeties of natural silk fibers is essential.

The *B. mori* cocoon is composed of a single fiber 700-1500 m long and 10-20 μm in diameter.[5] Each individual fiber is composed of twisted bunches of fibrils of diameter 20-25 nm[13] or 30 nm[14]: presumably this structure strengthens the fiber in the same way that a rope is strengthened by being composed of smaller twisted strands. It is a twin fiber and is composed of two proteins: fibroin and sericin. The fibroin molecule, which consists of a heavy (H) chain of 390 kDa and a light (L) chain of 26 kDa connected by a disulfide bond, as well as a glycoprotein named P25(30 kDa),[15-18] is secreted into the

posterior silk gland as an aqueous solution. Normally the sticky sericin coating is removed during processing, by a process known as degumming.

The amino acid composition of the H chain of silk fibroin (in mol %) is dominated by four amino acids: Gly (46 %), Ala (30 %), Ser (12 %) and Tyr (5.3 %).[19] The primary structure contains a large number of repeated sequences organized into 12 domains, each of which has related subdomains.[19-21] The silk fibroin is comprised 55% of repeated GAGAGS sequence,[19] which is thought to form the crystalline fraction (Cp) with averaged sequence of GAGAGSAAG$(SGAGAG)_8$Y.[22] This Cp-fraction can be prepared from *B. mori* silk fibroin by proteolytic digestion.

Two crystalline structures have been reported for the *B. mori* silk fibroin. One is Silk I and represents the structure of the silk fibroin stored in the silk gland of *B. mori* before spinning. The other is Silk II and is the structure of the silk fiber after spinning. We have previously proposed a structural model for Silk I as a "repeated β-turn type II structure" on the basis of several solid state NMR analyses.[23-25] The X-ray diffraction pattern of native silk fibroin fiber shows that Silk II is an anti-parallel β-sheet. A model was proposed by Marsh *et al.* in 1955[26] and remains the accepted structure. However, no high-resolution crystal structure has been reported, and experimental data in solid ^{13}C NMR spectra consistently indicate heterogeneity.[27] Then, as the first step for understanding whole structural heterogeneity of *B. mori* silk fiber, we have focused on the crystalline structure in the *B. mori* Cp-fraction and its model peptides of $(Ala\text{-}Gly)_n$ and $(Ala\text{-}Gly\text{-}Ser\text{-}Gly\text{-}Ala\text{-}Gly)_n$; hereafter they are called as $(AG)_n$ and $(AGSGAG)_n$.

We have recently proposed that there are two kinds of different local

packings at atomic level, which explain the chemical shift agreements including ^{1}H nucleus by using CASTEP program and GIPAW method, in the β-sheet structures of $(AG)_{15}$ peptide with Silk II form.[28] Other while, more previous papers from our group also have suggested lamella structure, in which there turns at each 8 residues, based on the intensity change of the highest field peak at Ala-Cβ region of selectively ^{13}C-labeled $(AG)_{15}$ peptides and molecular mechanics considering as single molecule.[29,30] In the latter model, the ratio tendency 2:1 which appears at two β-sheet peaks at lower field in the Ala-Cβ region of $(AG)_n$ peptides[30] may be explained by interpreting that Ala residue closest to the β-turn are located on β-sheet B environment. However, there have been no supports from the chemical shift calculation yet and many kinds of combination in considerable inter-sheet packing and lamella alignment. Moreover, in the model peptide $(AGSGAG)_5$, which has more similar sequence to the actual *B. mori* Cp-fraction, there has resulted in slight but significant perturbation of lamella periodicity seen in $(AG)_{15}$.[31] Thus, the detailed heterogeneity of β-sheet structure in the Silk II in the actual silk sample including Ser is still in controversial. For the construction of atomic model explaining the both previous models, we need to obtain more structural insight from experimental data.

1.2. Solid State NMR Approach for heterogeneous structure

The chemical shift calculation is very strong method to evaluate the structural agreement with the candidate structure.[32,33] There is an example showing that this method is sensitive especially to the local environment,

including torsion angle and even tertiary packing without directly bonded linkage, in the distance range of crystal lattice size.[34] In the heterogeneous structure of *B. mori* Cp-fraction with Silk II form, we need to evaluate structural agreement in more wide range of distances including both turn and β-sheets, that is, whole structural markers described below.

In the structural analysis for polymers, ^{13}C detected ^{1}H spin-lattice relaxation times for laboratory frame T_1^H and rotational frame $T_{1\rho}^H$ have been utilized for the evaluation of domain sizes included in the polymer about the scale of several ten nano meter and several nano meter, respectively.[35,36] This domain size evaluation is based on ^{1}H spin diffusion which can be imitated as diffusion equation.[37] Solid state ^{13}C NMR spectrum gives very sensitive signals for polymorphs,[38,39] so the detection by ^{13}C nucleus has the advantage for the separated detection of the polymorphs.

Moreover, in order to evaluate the longer distances around 1 nm range, polarization transfer matrix analysis[40,41] and master-equation approach[42] by using homonuclear ^{13}C-^{13}C spin diffusion in uniformly ^{13}C-labeled sample and molecular dynamics has been developed. Also about ^{1}H nucleus, NMR crystallography has been developed for low molecular weight compounds.[34] On the other hand, in the case of *B. mori* Cp-fraction, its primary sequence is highly repetitive with $(GAGAGS)_n$ and number of its correlation peaks is limited as displayed in later chapters. For this reason, it is difficult to construct the candidate structure based on structural information from many correlations. Furthermore, the two kinds of β-sheet peaks in ^{1}H solid NMR spectrum even at the ultra-fast spinning cannot be separated without

through ^{13}C nucleus interaction.[28] Therefore, for the evaluation of structural integrity in *B. mori* Cp-fraction with several structural components, utilization of ^{13}C detection with high resolution is essential.

In this thesis, the author constructs the method to evaluate the structural agreement of *B. mori* Cp-fraction via relaxation matrix approach, which is practical and conforming to experimental data more than other approach, and improve its quantitativity at first. Then, by combining the qualitative observation of 2D homonuclear ^{13}C-^{13}C correlations and domain size evaluation via ^{1}H spin-lattice relaxation times detected0 by ^{13}C nucleus, the author obtains insight into structural modeling for *B. mori* Cp-fraction.

1.3 Summary of Each Chapter

This thesis consists of the following chapters.

Chapter 2: Assignment of each ^{13}C NMR peak for each domain in Silk II structure of *Bombyx mori* (*B. mori*) silk fibroin Cp-fraction

As the first step for the structural elucidation of *B. mori* Cp-fraction via solid state NMR measurements, the author shows the results of assignment of Ser peaks in solid state ^{13}C CP/MAS NMR spectra which have heterogeneous line shapes in the spectra. These peaks show high sensitivity to the secondary structure of the *B. mori* Cp-fraction, which has primary sequence $(AGSGAG)_n$ mainly, because they also have the side-chain carbon as well as Ala residues. The author assigned the heterogeneous Ser-Cβ peak based on 2D ^{13}C-^{13}C dipolar-assisted rotational resonance (DARR) spectrum, which gives similar information to 2D NOESY spectrum in solution NMR, at

a short mixing time. Moreover, from ^{13}C spin-lattice relaxation time in laboratory frame T_1^C of the model peptide $(AGSGAG)_5$, the chemical environment difference of Ser-Cβ carbons possibly because of easiness for hydrogen bond formation in Ser-Cβ-OH is discussed.

Chapter 3: Quantitative method to evaluate the structural agreement of tertiary packing structure in a β-sheet domain –Investigation by antiparallel β-sheet alanine tetramer (AP-Ala_4) with already-known structure

In this chapter, the author discusses the method for evaluating structural agreement including the tertiary structure of AP-Ala_4 with already known structure. The author compares the matching of build-up curves, which mean the peak intensity change with mixing time in the 2D NMR measurements, obtained from both the 2D ^{13}C-^{13}C proton-driven spin diffusion (PDSD) as conventional method and dipolar-assisted rotational resonance (DARR) as new investigation in this thesis. The build-up curve depends on the inter-carbon distance. Therefore, from the agreement of calculated and simulated build-up curves, we can validate the candidate structure. For the application for *B. mori* Cp-fraction, the quantitativity of correlation peaks from the both experiments is discussed.

Chapter 4: An application of ^{13}C-^{13}C DARR build-up curve approach for the candidate models with chemical shift agreements and method development for the *B. mori* Cp-fraction with Silk II form

There are at least two kinds of β-sheet structure with many overlapping

peaks each other in the *B. mori* Cp-fraction with Silk II form. The build-up curve was simulated from each effective ^{13}C-^{13}C distance (r_A, r_B) estimated from the candidate atomic coordinate of the model peptide $(AG)_{15}$, which has been published most recently, and then added each other. For the explanation of the observed build-up curves, the author adds the consideration of ^{13}C-labeled percentage, inter/intra-molecular combination occurring by the insertion of Ser instead of Ala in the model, and detailed spin environment in their effective distances.

Chapter 5: Insight into structure modeling of β-sheet domain in Silk II structure of *B. mori* silk fibroin

In this chapter, the author discusses mainly the structure a little higher than tertiary packing of β-sheets by combining some kinds of spin diffusion experiments. The deep knowledge about its molecular structure and morphology of the silk materials will be a key in the material development. The primary structure of *B. mori* silk fibroin consists mainly of (GAGAGX)n repetitive region (X = S (main), Y or V). So far, we have focused on the local structure until about 1 nm at atomic level in the crystalline region, $(GAGAGS)_n$. Here, the author discusses the inter-domain structure of the crystalline region by correlating Ala and Ser peaks, which are markers of each domain with separated chemical shift, in solid state ^{13}C NMR spectra of *B. mori* Cp-fraction.

Chapter 6: Conclusion

Based on the discussions until the previous chapters, the author outlines the significances of the previous results and describes on the perspective of application of solid state NMR spectra for the *B. mori* silk fibroin with fiber or powder form.

REFERENCES

1. Fraser, B.; MacRae, T. P. *Conformation of Fibrous Proteins and Related Synthetic Polypeptides*; Academic Press: **1973**; pp293 - 343.
2. Asakura, T.; Kaplan, D. L., *Silk Production and Processing*, in *Encyclopedia of Agricultural Science*, Ed. by C. J. Arutzen, Academic Press, **1994**, *4*, 1-11.
3. Schoeser, M. *The Science of Silk* in *Silk*; Yale University Press, **2007**, 232-241.
4. Vollrath, F.; Porter, D. *Soft Matter* **2006**, *2*, 377-385.
5. Zhao, H-P.; Feng, X-Q.; Shi, H-J. *Mater. Sci. Eng. C.* **2007**, *27*, 675–683.
6. Zhao, H-P.; Fenga, X.-Q.; Yu, S.-W.; Cui, W.-Z.; Zou, F.-Z. *Polymer* **2005**, *46*, 9192-9201.
7. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L. *Biomaterials* **2003**, *24*, 401–416.
8. Gronau, G.; Krishnaji, S. T.; Kinahan, M. E.; Giesa, T.; Wong, J. Y.; Kaplan, D. L.; Buehler, M. J. *Biomaterials* **2012**, *33*, 8240-8255.
9. Porter, D.; Vollrath, F. *Adv. Mater.* **2009**, 487-492.
10. Yan, J.; Zhou, G.; Knight, D. P.; Shao Z.; Chen, X. *Biomacromolecules* **2010**, *11*, 1–5.
11. Jin, Y.; Hang, Y.; Luo, J.; Zhang, Y.; Shao, H.; Hu, X. *Int. J. Biol. Macromol.* **2013**, *62*, 162-166.
12. Mortimer, B.; Guan, J.; Holland, C.; Porter, D.; Vollrath, F. *Acta Biomaterialia* **2015**, *11*, 247-255.

13. Greving, I.; Cai, M.; Vollrath, F.; Schniepp, H. C. *Biomacromolecules*, **2012**, *13*, 676-682.
14. Xu, G; Gong, L.; Yang, Z.; Liu, X. Y. *Soft Matter* **2014**, *10*, 2116-2123.
15. Takei, F.; Kikuchi, Y.; Kikuchi, A.; Mizuno, S.; Shimura, K. *J. Cell Biol.* **1987**, *105*, 175-180.
16. Tanaka, K.; Mori, K.; Mizuno, S. *J. Biochem. (Tokyo)* **1993**, *114*, 1-4.
17. Tanaka, K.; Kajiyama, N.; Ishikura, K.; Waga, S.; Kikuchi, A.; Ohtomo, K.; Takagi, T.; Mizuno, S. *Biochim. Biophys. Acta.* **1999**, *1432*, 92-103.
18. Couble, P.; Chevillard, M.; Moine, A.; Ravel-Chapuis, P.; Prudhomme, J. C. *Nucleic Acids Res.* **1985**, *13*, 1801-1814.
19. Zhou, C. Z.; Confalonieri, F.; Medina, N.; Zivanovic, Y.; Esnault, C.; Yang, T.; Jacquet, M.; Janin, J.; Duguet, M.; Perasso, R.; Li, Z. G. *Nucleic Acids Res.* **2000**, *28*, 2413-2419.
20. Zhou C. Z.; Confalonieri, F.; Jacquet, M.; Perasso, R.; Li, Z. G.; Janin, J. *Proteins: Structure, Function, and Genetics* **2001**, *44*, 119-122.
21. Asakura, T.; Ohgo, K.; Ishida, T.; Taddei, P.; Monti, P.; Kishore, R. *Biomacromolecules* **2005**, *6,* 468-474.
22. Strydom, D. J.; Haylett, T.; Stead, R. H. *Biochem. Biopyhs. Res. Comm.* **1977**, *79*, 932-938.
23. Asakura, T.; Ohgo, K.; Komatsu, K.; Kanenari, M.; Okuyama, K. *Macromolecules* **2005**, *38*, 7397-7403.
24. Asakura, T.; Ashida, J.; Yamane, T.; Kameda, T.; Nakazawa, Y.; Ohgo, K.; Komatsu, K. *J. Mol. Biol.* **2001**, *306*, 7397-7403.
25. Asakura, T.; Suzuki, Y.; Yazawa, K.; Aoki, A.; Nishiyama, Y.; Nishimura, K.; Suzuki, F.; Kaji, H., *Macromolecules* **2013**, *46*, 8046-8050.

26. Marsh, R; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* 1955, *16*, 1-34.

27. Asakura, T.; Yao, J.; Yamane, T.; Umemura, K.; Ulrich, A. S. *J. Am. Chem. Soc.* **2002**, *124*, 8794-8795.

28. Asakura, T.; Ohata, T.; Kametani, S.; Okushita, K.; Yazawa, K.; Nishiyama, Y.; Nishimura, K.; Aoki, A.; Suzuki, F.; Kaji, H.: Ulrich, A. S.; Williamson, M. P. *Macromolecules* **2015**, *48*, 28-36.

29. Asakura, T.; Nakazawa, Y.; Ohnishi, E.; Moro, F. *Protein Science* **2005**, *14*, 2654–2657.

30. Asakura, T.; Sato, H.; Moro, F.; Nakazawa, Y.; Aoki, A. *J. Am. Chem. Soc.* **2007**, *129*, 5703-5709.

31. Suzuki, Y.; Aoki, A.; Nakazawa, Y.; Knight, D. P.; Asakura, T. *Macromolecules* **2010**, *43*, 9434–9440.

32. Cadars, S.; Layrac, G.; Gerardin, C.; Deschamps, M.; Yates, J. R.; Tichit, D.; Massiot, D. *Chem. Mater.* **2011**, *23*, 2821-2831.

33. Harris, R. K.; Hodgkinson, P.; Pickard, C. J.; Yates, J. R.; Zorin, V. *Magn. Reson. Chem.* **2007**, *45*, S174-S186.

34. Salager, E.; Day, G. M.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. *J. Am. Chem. Soc.* **2010**, *132*, 2564-2566.

35. Stejskal, E. O.; Schaefer, J.; Sefcik, M. D.; McKay, R. A. *Macromolecules* **1981**, *14*, 275-279.

36. Asano, A.; Takegoshi, K. Polymer Blends and Miscibility. Solid State NMR of Polymers (ed. by Ando, I. & Asakura, T.); Elsevier Science B. V.: Amsterdam, 1988; pp. 351-414.
37. Clauss, J.; Schmit-Rohr, K.; Spiess, H. W. *Acta Polymer* **1993**, *44*, 1-17.
38. Atalla, R. H.; VanderHart, D. L. Science **1984**, *223*, 283-285.
39. VanderHart, D. L.; Atalla, R. H. Macromolecules **1984**, *17*, 1465-1472.
40. Kato, T.; Endo, S.; Fujiwara, T.: Nagayama, K. *J. Biomol. NMR* **1993**, *3*, 653-673.
41. Kubo, A.; McDowell, C. A. *J. Chem. Soc. Faraday Trans* **1988**, *84*, 3713-3730.
42. Dumeza, J.-N.; Emsley, L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7363-7370.

Chapter 2

Assignment of each ^{13}C NMR peak for each domain in Silk II structure of *Bombyx mori* (*B. mori*) silk fibroin Cp-fraction

2-1. INTRODUCTION

At first, the author introduces the determination of secondary structure via solid state NMR spectrum of heterogeneous *B. mori* Cp-fraction in this chapter. Since Marsh model in 1955,[1] no high-resolution crystal structure for Silk II, which is the *B. mori* silk fibroin structure after spinning, has been reported, and experimental data consistently indicate heterogeneity. Lotz *et al.*[2] demonstrated structural disorder of the Silk II $(AG)_n$ sequence from their X-ray and electron diffraction analyses, with a Gly-Gly inter-sheet spacing greater than that seen in polyglycine, and an Ala-Ala spacing less than that seen in polyalanine. Fraser *et al.*[3] have also reported that the anti-parallel β-sheet has a slightly greater inter-sheet spacing for $(AGSGAG)_n$ as compared to that for $(AG)_n$. In the papers,[4,5] using ^{13}C CP/MAS NMR, we found that the Silk II structure is heterogeneous: the Cp fraction comprises18% distorted β-turns, 25% β-sheet (Domain A) and 13% β-sheet (Domain B), while the remaining 45% amorphous fraction is composed almost equally of distorted β-turn and β-sheet. This information was obtained from a detailed analysis of the Ala-Cβ carbon peaks in the ^{13}C CP/MAS NMR of the Cp fraction and appropriate model peptides based on its chemical shift contour plot, which has been made up of 40 proteins in PDB

database and Ramachandran plot.[6]

On the other hand, there exists also Ser residue in the Cp-fraction as a residue with its side chain. Serine is a key determinant of the Silk I, which is the silk structure before spinning, or Silk II structure through hydrogen bonding, involving the $O^{\gamma}H$ group of the Ser side chain as reported previously.[7-11] Namely, the Ser side chain is capable of forming intra-molecular hydrogen bonds involving the $O^{\gamma}H$ group and the backbone carbonyl group of the Gly residue, stabilizing the Silk I form.[10]

In this chapter, we clarify the chemical shift connectivity in the crystalline domain of *B. mori* silk fibroin in Silk II form using two-dimensional (2D) ^{13}C-^{13}C DARR (dipolar-assisted rotational resonance)[12] spectrum of the ^{13}C uniformly labeled Cp fraction of *B. mori* silk fibroin. Especially, a detailed assignment of the Ser-Cβ peaks, which give the asymmetric and broad peak, can be separated by the 2D DARR measurement. The ^{13}C chemical shifts of the Ser-Cβ obtained from the 2D DARR cross peak revealed that the Ser-Cβ has different local environments in the two β-sheet chains.

A large difference in the side chain mobility between hydrogen bonded and non hydrogen bonded formations through the Ser OH group has been reported from ^{13}C spin-lattice relaxation time (T_1^C) observation of the Cp fraction and model peptides.[7-9] On the other hand, the ^{2}H NMR powder pattern spectra of [3,3-^{2}H]Ser-*B. mori* silk fiber with Silk II form have been analyzed by line shape simulation.[11] Two types of motion could be characterized quantitatively: one component undergoing a rapid three-site jump and a second component representing a slow exchange between sites

with unequal occupancies and with a small amplitude of libration. In addition, ^{2}H solid state NMR studies of uniaxially aligned [3,3-^{2}H]Ser-*B. mori* silk fiber with Silk II form have been performed to determine the side-chain conformation and the orientational distribution of the Ser residues in the slow motional component: the dominant conformer of the Ser side chain is *gauche*$^{+}$ around the N-Cα-Cβ-O bond.[11] Therefore, we also obtain structural and dynamical information more selectively, by using already-prepared [3-^{13}C]Ser-$(AGSGAG)_5$[9] with different selective [3-^{13}C]Ser labeling positions as model peptides of the Cp fraction.

2-2. EXPERIMENTAL SECTION

2-2-1. Crystalline fraction of [U- ^{13}C] *B. mori* silk fibroin

B. mori larvae were reared in our laboratory. Uniformly ^{13}C labeled *B. mori* silk fibroin, [U-^{13}C] *B. mori* silk fibroin, was achieved biosynthetically by oral administration of an artificial diet with ^{13}C-uniformly labeled glucose to larvae of the fifth instar.[13,14] The supplement was mixed with 2 g of an artificial diet per day. The amount of [U-^{13}C] glucose was 10 mg each on the fourth and fifth day of the fifth larval stage. The total amount of [U-^{13}C] glucose administered was thus 20 mg per silkworm. The [U-^{13}C] Cp fraction of *B. mori* silk fibroin was prepared from regenerated [U- ^{13}C] *B. mori* silk fibroin.[15] Chymotrypsin (40 mg) dissolved in a few milliliters of water was added to an aqueous solution of about 4 g of fibroin buffered with $Na_2HPO_4 \cdot 12H_2O$ and $NaH_2PO_4 \cdot 2H_2O$ at pH 7.8. The solution (300 mL) was incubated at 37 °C for 24 h and the precipitate was separated by centrifuging at 10,000 rpm followed by washing with 0.03 N HCl to inactivate the enzyme reaction. The precipitate was further washed several times with distilled water, ethyl alcohol, and diethyl ether, and freeze-dried, yielding a crystalline fraction corresponding to 55% of the original fibroin. The obtained precipitate forms the Silk II structure and is referred to as the Cp-fraction.

2-2-2. Synthesis of peptides, [3-^{13}C]Ser^{9}-, [3-^{13}C]Ser^{15}-, [3-^{13}C]Ser^{21}-(AGSGAG)$_5$ and (AGSGAG)$_5$

H-[3-^{13}C]Ser OH purchased from Cambridge Isotope Lab., Inc., was converted to Fmoc-[3-^{13}C]Ser(t-Bu)-OH (yield: 60%).[16] The Fmoc--[3-^{13}C]Ser(t-Bu)-OH was synthesized from H--[3-^{13}C]Ser-OH according to the following process.

Step 1. The starting compound of [3-^{13}C]Ser was purchased from Cambridge Isotope Laboratories,Inc. The [3-^{13}C]Ser (5 mmol) was mixed in a 100 ml round-bottomed flask

with 1.2 equivalents of $NaHCO_3$ dissolved in a little amount of water. Then, 1.2 equivalents of *N*-(Benzyloxycarbonyloxy)succinimide (ZOSu) dissolved in the same amount of acetone was added into the aqueous solution and stirred overnight. Reaction completion was checked with TLC ($CHCl_3$: MeOH : AcOH = 85 : 25 : 2, colored by ninhydrin). The acetone in the reaction mixture was removed by evaporation. The remaining water was washed with CH_2Cl_2 three times. Into the separated aqueous layer, 1N HCl aq. was added until the pH of the solution became 1.0. Then, the solution was washed with EtOAc three times, and the organic layer was obtained as the target solution. This organic layer was washed with water and saturated NaCl aq. (Sat. NaCl aq.), and then Na_2SO_4 was added and left overnight. After filtration, the obtained organic layer was evaporated to remove the solvent, and the final yield of Z-Ser-OH was 87.86 % (4.393 mmol).

Step 2, 3. The obtained Z-Ser-OH (4.393 mmol) was dissolved into some water and mixed with 0.5 equivalents of $CsCO_3$ aq. (pH=7.0). After that, by evaporating the solution, water was removed. Next, a small amount of DMF and 1.2 equivalents of benzyl bromide (BzlBr) were added and stirred overnight. The reaction progress was checked with TLC ($CHCl_3$: MeOH : AcOH = 95 : 5 : 3). After adding Et_2O and 1N HCl aq. into the reaction mixture until the pH reached 1.0, the organic layer was extracted and washed again with water. Until the pH reached 9.0, $NaHCO_3$ aq. was added into the organic layer, which was separated from the aqueous layer. Then, the organic layer was washed with Sat. NaCl aq. The removal of water and evaporation of solvent were performed in the same way as for step 1. The final yield of Z-Ser-OBzl was 91.96 % (4.04 mmol).

Step 4. The obtained Z-Ser-OBzl (4.04 mmol) was reacted with isobutene, mixing with dry CH_2Cl_2 and conc.H_2SO_4 for three days at 0 ℃. The reaction progress was followed by TLC ($CHCl_3$:MeOH:AcOH=85:25:2). The reaction was quenched with a little amount

of saturated $NaHCO_3$ aq. (sat.$NaHCO_3$ aq.), and then the organic layer was extracted by adding Et_2O. The organic layer was washed with water. The removal of water and evaporation of solvent were performed in the same way as step 1. The crude yield of Z-Ser(*t*Bu)-OBzl was 99.75 % (4.03 mmol). For the purification, the product was recrystallized using Et_2O as a good solvent and petroleum ether as a poor solvent.

Step5. In a 200 ml round-bottomed flask, 0.25 equivalents of Palladium / activated charcoal (10% Pd) and dry MeOH were mixed and after that, 1.0 equivalents of AcOH was added. Then, Z-Ser(*t*Bu)-OBzl (4.03 mmol) dissolved into dry MeOH and THF was added and the air in the flask was exchanged into H_2 gas, stirred overnight at 0 ℃. The reaction progress was checked by TLC ($CHCl_3$: MeOH : AcOH = 95 : 5 : 3). After filtration to remove Pd/C, the solvent was removed with evaporation and the product H-Ser(*t*Bu)-OH·AcOH was obtained with a yield of 98.76 % (3.98 mmol).

Step 6. In a 300 ml round-bottomed flask, the obtained H-Ser(*t*Bu)-OH·AcOH (3.98 mmol) and Na_2CO_3 aq. were mixed. Then, 1.2 equivalents of *N*-(9-Fluorenylmethoxycarbonyloxy)succinimide (Fmoc-OSu) dissolved into the same amount of dioxane was added into the solution, and stirred overnight. The reaction progress was checked by TLC (BuOH : AcOH : H2O = 4 : 1 :1). The reaction mixture was washed with Et_2O. Next, into the aqueous layer, 5% citric acid was added until the pH reached 3.0, and then the organic layer was separated by EtOAc. After washing the organic layer with water and sat. NaCl aq., the removal of water and evaporation of solvent were performed as described above. Also in this step, the purification was performed in the same way. The yield of the purified product Fmoc-Ser(*t*Bu)-OH was 74.67 % (3.02 mmol), resulting in a total yield of the final product Fmoc-Ser(*t*Bu)-OH from the starting compound [3-^{13}C]Ser of 60.4 %.

The Fmoc solid-phase method using a fully automated Pioneer Peptide Synthesis System (Applied Biosystems Ltd. Japan) was used to prepare [3-^{13}C]Ser^{9}-, [3-^{13}C]Ser^{15}-, and [3-^{13}C]Ser^{21}- $(AGSGAG)_5$ peptides together with unlabeled peptide as reported previously.[17] These peptides were dissolved in 9 M LiBr and then dialyzed against distilled water. The obtained precipitate forms the Silk II structure.

2-2-3. ^{13}C CP/MAS NMR and T_1^C experiments

The ^{13}C CP/MAS NMR measurements were conducted on a Bruker AVANCE-400 spectrometer operating at 100 MHz for the ^{13}C nucleus. Cross polarization (CP) was employed for sensitivity enhancement with high-power ^{1}H decoupling during the signal acquisition. A ^{1}H 90° pulse width of 5 μs duration was used with 1 ms contact time and 3 s repetition time. The number of scans for uniformly ^{13}C-labeled *B. mori* silk fibroin samples was 64. Approximately 15K FIDs were added to generate the spectra of synthesized peptides. The ^{13}C chemical shifts were calibrated indirectly through the adamantane methylene peak observed at 28.8 ppm relative to TMS (tetramethylsilane) at 0 ppm. The ^{13}C spin-lattice relaxation times, T_1^C, were observed using a JEOL ECX 400 spectrometer operating at 100 MHz for the ^{13}C nucleus. We used a pulse sequence for T_1^C determination developed by Torchia.[18] The MAS spinning speed was 8 kHz and the ^{1}H 90° pulse, ^{13}C 90° pulse, contact and repetition times were 3.9 μs, 5 μs, 1.2 ms and 5 s, respectively. The T_1^C components were fitted to two exponentials according to the following equation:

$$M_{CP}(\tau) = M_{CP}(0)\left\{I \cdot exp\left(-\frac{\tau}{T_1^C{}_{short}}\right) + (1-I) \cdot exp\left(-\frac{\tau}{T_1^C{}_{long}}\right)\right\}$$

where $M_{cp}(\tau)$ and $M_{cp}(0)$ stand for the ^{13}C NMR peak intensities at times τ and 0, respectively. Spectra were accumulated 512 times with arrayed variable delay range of 0.1-8.0 s at room temperature.

2-2-4. ^{13}C-^{13}C dipolar assisted rotational resonance experiment.

The ^{13}C DARR measurements[12] were conducted on a Bruker DMX-500 spectrometer operating at 125 MHz for the ^{13}C nucleus. The experiment was performed at a magic-angle spinning (MAS) frequency of 11 kHz with a 4 mmϕ rotor. The CP contact time and the recycle delay were 1.5 ms and 3 s, respectively. The mixing time was varied 10 ms, and 64 scans were collected at each time increment of the indirect dimension. The ^{1}H power for the recoupling was adjusted to the same as that of the MAS frequency in Hz.

2-3. RESULTS AND DISCUSSION

2-3-1. Deconvolution of Ser-Cα and Cβ peaks and assignment to the β-sheets A and B.

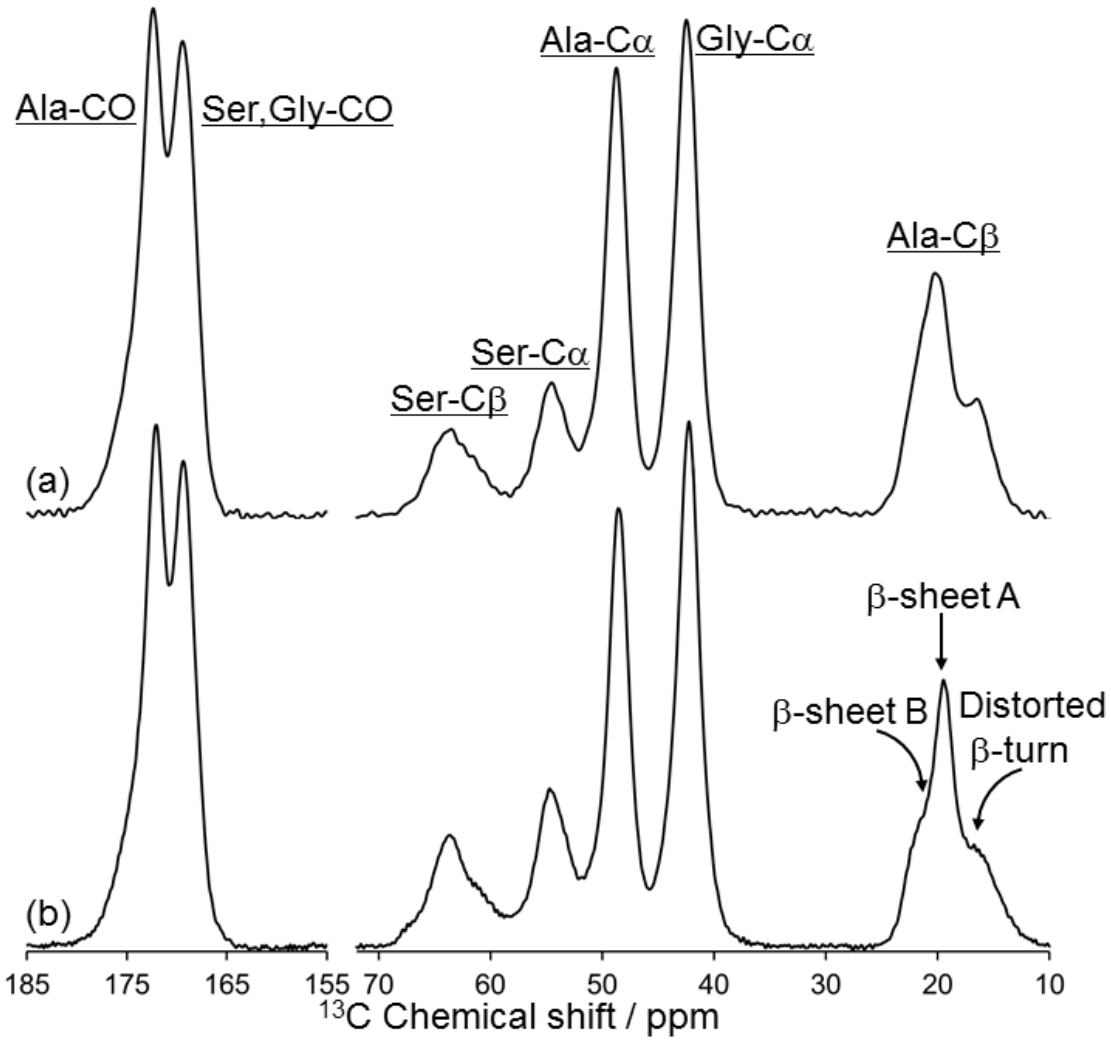

Figure 2-1. ^{13}C CP/MAS spectra of (a) [U-^{13}C] *B. mori* fiber and (b) [U-^{13}C] Cp-fraction (Silk II) observed using 400 MHz NMR at MAS = 9 kHz.

Fig 2-1 shows ^{13}C CP/MAS spectra of [U-^{13}C] *B. mori* silk fiber (a) and [U-^{13}C] Cp fraction in Silk II form (b). *B. mori* silk fiber mostly consists of two repeated sequences: AGSGAG and $(GX)_mGY$ (X=A or V).[19] The Cp fraction mainly consists of repeated AGSGAG sequences and occupies about 55 mol% of the silk fibroin. The AGSGAG sequence is thus the main part of the Silk II structure. The spectrum of the [U-^{13}C] Cp fraction shows similar line shapes and the same chemical shifts as those of the [U-^{13}C] *B. mori* silk fiber. This observation indicates that the main structure in *B. mori* silk fiber is the same as that of the Cp fraction in Silk II structure. However, the existence of other sequences like $(GX)_mGY$ [19] in the silk fiber makes analysis difficult and impedes clear investigation of the Silk II structure. Not surprisingly, the peak signals from Ser-Cβ and Ala-Cβ in *B. mori* fiber become relatively wider than those of the Cp fraction. This broadening is due to the existence of amorphous domains consisting of $(GX)_mGY$ sequence, and N-terminal and C-terminal regions.[19] In contrast, the Cp fraction has no such domains, so the observed ^{13}C NMR spectrum reflects the Silk II structure arising almost entirely from AGSGAG. Therefore, NMR investigation of the Cp fraction gives important information on the Silk II structure of *B. mori* silk fiber in detail.

In the Ala-Cβ region of the Cp fraction, three peaks are clearly seen at 16.2, 19.6 and 21.7 ppm. It has been reported that the characteristic ^{13}C NMR line shape in the Ala-Cβ region appears also in the Silk II structure. The three Ala-Cβ peaks were assigned to a distorted β-turn and two different packings of β-sheets, respectively.[4,5] The β-sheet peaks at 19.6 and 21.7 ppm have an intensity ratio close to 2:1 (Fig. 2-1b), suggesting the presence of two locally different β-sheets, as proposed previously by Takahashi et al,[20] which we designate as A and B, respectively. We have already proposed a candidate structure for the two β-sheets, that is, anti-parallel β-sheet structures with a

different inter-sheet arrangement.[4,5] The chemical shifts of both β-sheets A and B are consistent with the values predicted from the typical torsion angle (ϕ, ψ) of β-sheet structure on the Ala-Cβ chemical shift contour plot.[6] The typical torsion angle of a β-sheet structure is (-140°, 140°)[21] and the ^{13}C chemical shift is approximately 19.5 ppm (Fig. 2-2a, orange region). Thus, the β-sheet A at 19.6 ppm probably has the typical torsion angle of a β-sheet. In contrast, the β-sheet B is observed at the relatively lower field of 21.7 ppm. This observation implies the existence of an inter-sheet deshielding packing effect in the β-sheet B region. Therefore, both peaks at 19.6 and 21.7 ppm are attributed to β-sheets with different inter-sheet packing.

A broad peak was observed for the Ser-Cβ at around 63.7 ppm (Fig. 2-1b). This value is slightly larger than the 62.8 ppm that is predicted from the typical β-sheet geometry (the orange region of Fig. 2-2b).[6] Additionally, the peak shape indicates the existence of several Ser residues in different environments. To resolve individual peaks from the broad and complicated Ser-Cβ peak, we employed 2D ^{13}C-^{13}C DARR experiments. Fig. 2-3a shows a 2D ^{13}C-^{13}C DARR spectrum of the [U-^{13}C] Cp fraction at a mixing time of 10 ms. The cross peaks between Ser-Cα and Ser-Cβ shown in the expanded box in Fig. 2-3a must be intra-residue because spin transfer across long distance cannot occur with such a short mixing time. Four such cross peaks are clearly observed: hereafter we abbreviate them as (Cα 56.0, Cβ 61.4), (Cα 54.0, Cβ 63.5), (Cα 54.9, Cβ 64.2) and (Cα 55.6, Cβ 66.8). Three of these peaks have Cβ chemical shifts larger than that of the typical β-sheet chemical shift of 62.8 ppm. The large chemical shift value is ascribed to an inter-sheet packing effect as stated above for Ala-Cβ. In contrast, the cross peak (Cα 56.0, Cβ 61.4) is attributable to a random coil because both chemical shifts are in

agreement with the ^{13}C solution NMR chemical shifts of Ser-Cβ in proteins with random coil conformation in aqueous solution.[エラー! 参照元が見つかりません。]

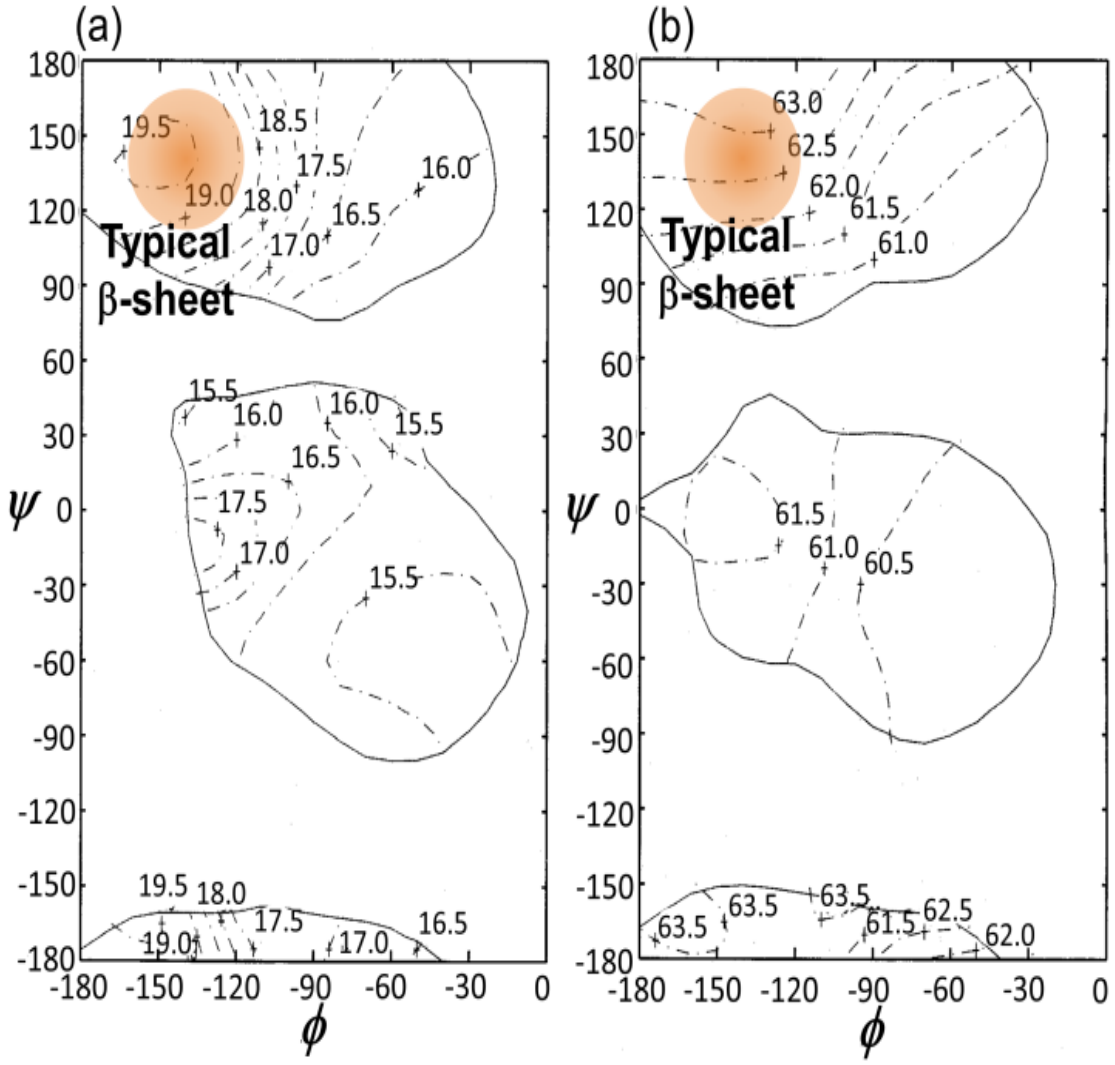

Figure 2-2. ^{13}C chemical shift contour plots of Cβ carbons of Ala (a) and Ser (b) residues in proteins published previously.[6] The (ϕ, ψ) angle regions colored in orange are typical β-sheet regions of (ϕ, ψ) = (-140°, 140°).

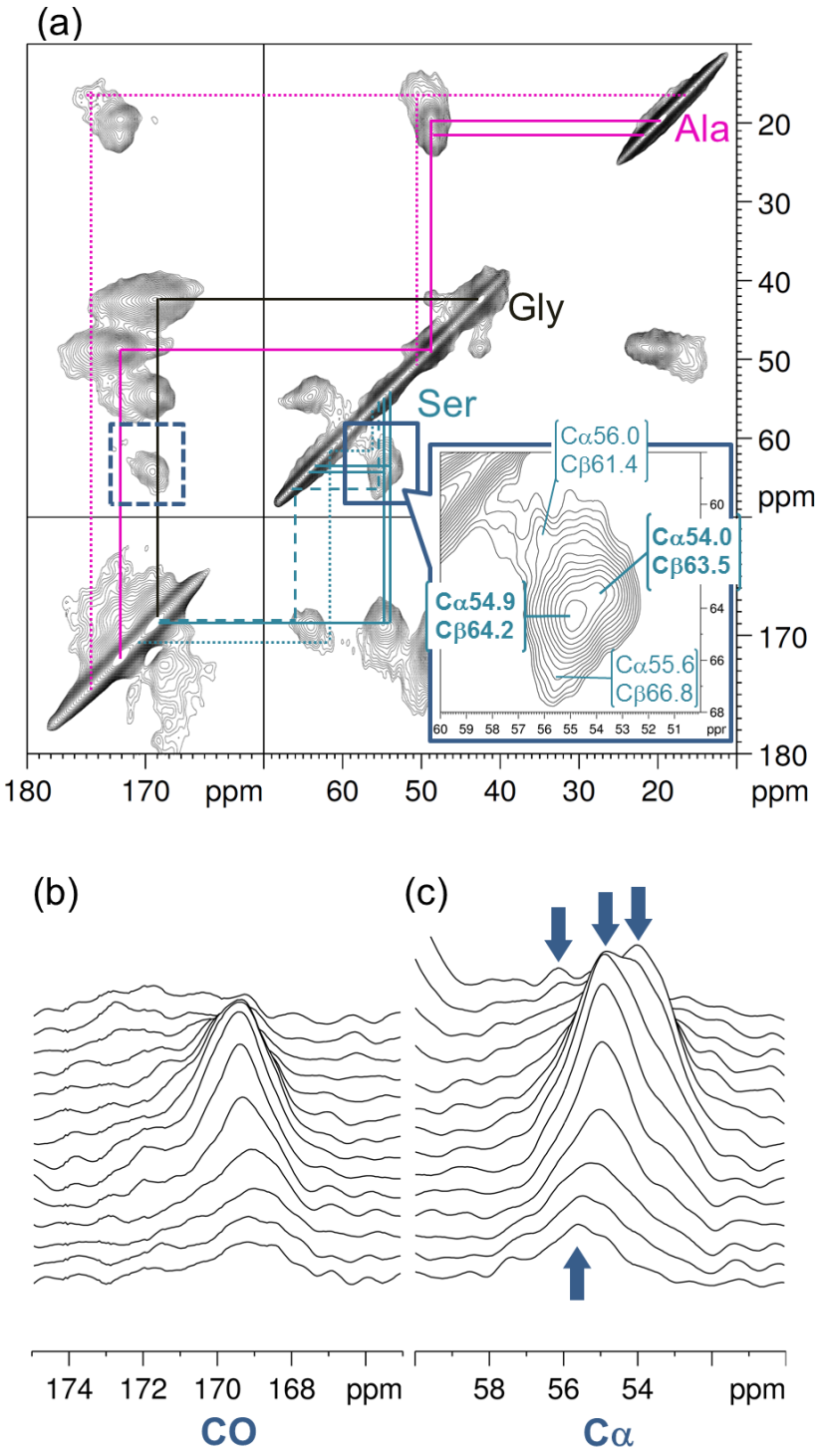

Figure 2-3. 2D ^{13}C-^{13}C DARR spectrum of [U-^{13}C] Cp-fraction (Silk II) observed with a mixing time of 10 ms. (a) Intra-residue assignment of the detected peaks and expansion of the Ser Cα-Cβ correlation. Stack plots of Ser-CO ((b), broken square region in (a)) and Cα ((c), expanded in (a)) regions.

To compare the amount of the three Ser-Cβ β-sheet peaks and to assign them to individual β-sheet structures, a peak deconvolution using four Gaussian functions was performed in the Ser-Cβ peak region for both [U-^{13}C] *B. mori* fiber and the [U-^{13}C] Cp fraction (Fig. 2-4). The four chemical shift values for the fitting were obtained from the 2D spectra in the expanded region of Figure 2-3a. Moreover, we limited the number of variables by using the same line widths to fit the 3 β-sheet peaks (peaks (i) to (iii) in Figure 2-4). Peak (iv) (random coil fraction) was fitted best using a larger peak width, which seems reasonable because the random coil or disordered region has a larger structural distribution than the ordered region in polymers. The Ser-Cβ peak from the *B. mori* fiber spectrum (Fig. 2-4a) includes signal from the soluble and/or low-molecular weight fraction[19] because the fiber is not cleaved at Tyr by α-chymotrypsin. Therefore, Ser-Cβ peaks in the fiber spectrum show a large structural distribution, resulting in a broad line width. In support of this argument, it has been shown that the relative amount of structural heterogeneity or randomness of the Ser-Cβ peak was increased on the replacement of Ser by Tyr, in the sequence AGYGAG.[22] The replacement induced a broadening of the Ser-Cβ peak around 67 ppm. The area of the peak at 64.2 ppm (ii in Fig. 2-4) is much larger than that at 63.5 ppm (iii). The relative ratio of β-sheets A and B has been determined from the Ala-Cβ peaks to be 2:1.[5] Considering the Ala-Cβ conclusion, the more intense Ser-Cβ peak at 64.2 ppm may thus be assigned to the β-sheet A structure from the intensity. But there is a discrepancy between the relative peak areas fitted in the Ser Cβ region (37:12) and the Ala Cβ region (2:1), which is presumably due to the difficulty of fitting the highly overlapping peaks in the Ser Cβ region. Peak assignments and their relative intensities are tabulated in Table 2-1. For

more detailed assignment, we analyze the correlation between Ala-Cβ and Ser- Cβ peaks in chapter 5.

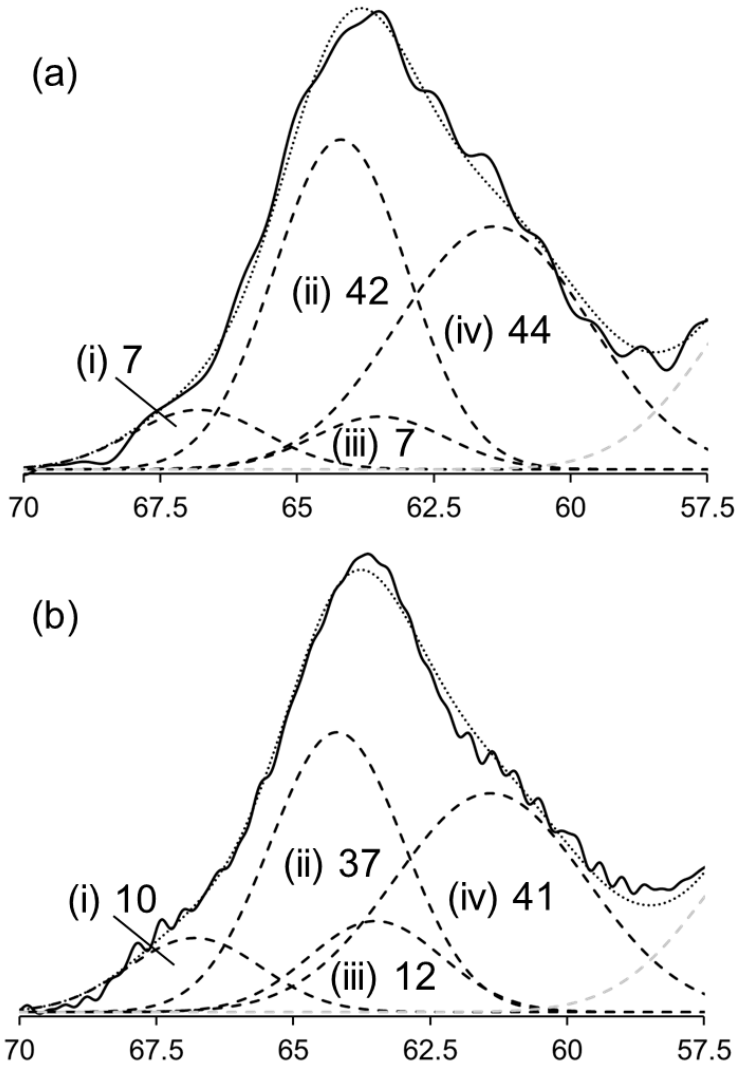

Figure 2-4. Gaussian line shape fitting result of Ser-Cβ peaks in ^{13}C CP/MAS spectra of (a) [U-^{13}C] *B. mori* silk fiber and (b) [U-^{13}C] Cp-fraction (Silk II). The solid lines show the observed spectra. The broken lines show individual Gaussian functions (– –) and the sum of all functions (·····). The fractions are assigned to (i) disordered A, (ii) A, (iii) B, (iv) random coil.

Table 2-1. ^{13}C chemical shift assignment of Cp-fraction (Silk II)

Residue	Domain	Cβ / ppm	Cα / ppm	CO / ppm	Fraction / % in each residue
Ala	Random coil / Distorted β-turn	16.2	50.5	173.4 – 175.1	32[*2]
	β-sheet A	19.6	48.5	172.0	45[*2]
	β-sheet B	21.7			23[*2]
Gly	-	-	42.4	169.0	-
Ser	Random coil	61.4	56.0	170.6	41
	β-sheet B	63.5	54.0	169.1	12
	β-sheet A	64.2	54.9	169.1	37
	β-sheet A[*1]	66.8	55.6	169.1	10

[*1] Distorted structure.

[*2] ref 10.

2-3-2. Dynamics of Ser residues and intermolecular hydrogen bonding formation through the Ser OH side chain

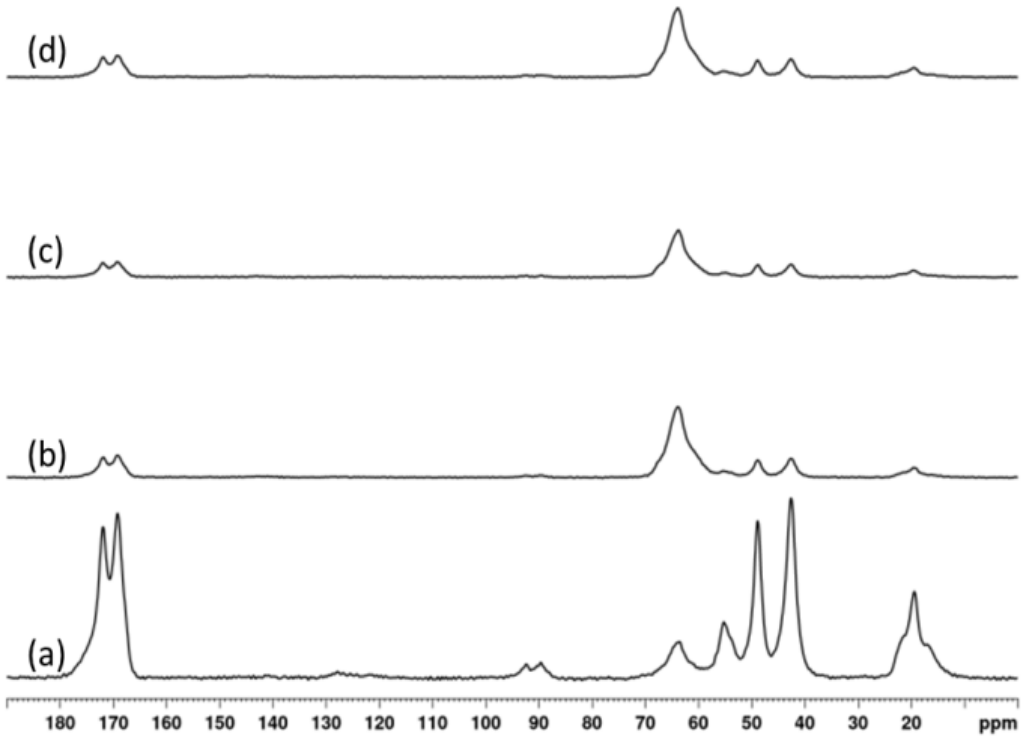

Figure 2-5. ^{13}C CP/MAS spectra of (a) non-labeled $(AGSGAG)_5$, (b) [3-^{13}C]Ser^{9}-$(AGSGAG)_5$, (c) [3-^{13}C]Ser^{15}-$(AGSGAG)_5$ and (d) [3-^{13}C]Ser^{21}-$(AGSGAG)_5$. The Ser-Cβ regions around 65 ppm in the spectra (b) – (d) are emphasized by the ^{13}C-labeling.

Expansions of the Ser Cβ peaks in the ^{13}C CP/MAS NMR spectra of [3-^{13}C]Ser^{9}-$(AGSGAG)_5$, [3-^{13}C]Ser^{15}- $(AGSGAG)_5$ and [3-^{13}C]Ser^{21}- $(AGSGAG)_5$ are shown in Fig. 2-6a, b and c, respectively. The peak intensities from natural abundance components were subtracted from the original intensities in the CP/MAS spectra. The whole regions of the original CP/MAS spectra are also shown in Figure 2-5. The number of components and the chemical shift values used for the fitting were again

obtained from Figure 2-3a. However, the fitted line widths were slightly narrower than those of the Cp-fraction (Fig. 2-4b), because the synthesized model peptides are chemically defined and therefore more uniform. Selective ^{13}C labeling of Ser-Cβ carbons is useful for analysis of the distribution of serines between the two domains, and of the dynamics of the domains using T_1^C measurements.

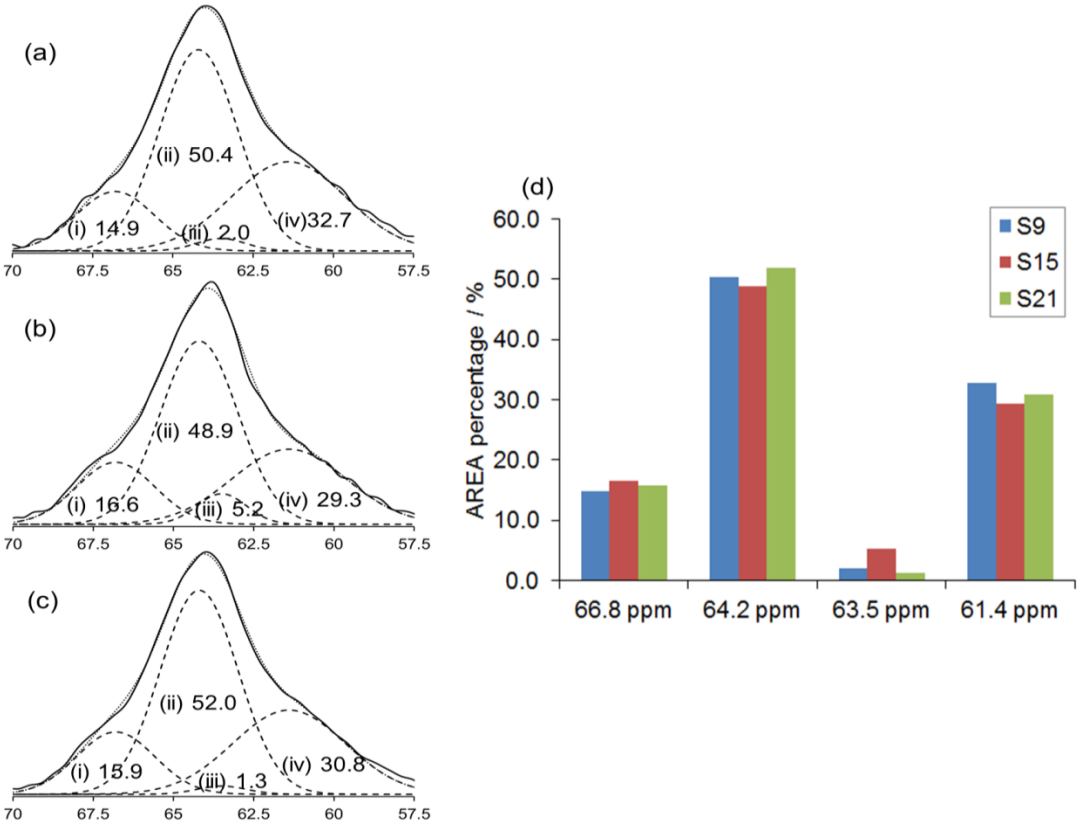

Figure 2-6. Gaussian line shape fitting results of Ser-Cβ peaks in 13C CP/MAS spectra of (a) [3-^{13}C] $Ser^{9}(AGSGAG)_5$, (b) [3-^{13}C] $Ser^{15}(AGSGAG)_5$ and (c) [3-^{13}C] $Ser^{21}(AGSGAG)_5$. The peak intensities from natural abundance components were subtracted from the original intensities in their CP/MAS spectra. The broken lines show individual Gaussian functions (– –) and the sum of all functions (·····). (d) A column chart of the percentage area at each Ser-Cβ peak of all 13C-labeled (AGSGAG)5 peptides. The fractions at 66.8, 64.2, 63.5 and 61.4 ppm are assigned respectively to (i) disordered A, (ii) A, (iii) B, (iv) random coil.

Table 2-2. T_1^C relaxation times of [3-^{13}C] $Ser_{15}(AGSGAG)_5$ measured at 400 MHz.

	66.8 ppm	64.2 ppm	63.5 ppm	61.4 ppm
I_{short}	0.59	0.68	-	0.44
I_{long}	0.41	0.32	1.00	0.56
$T_1^C{}_{short}$ / s	0.32	0.45	-	0.73
$T_1^C{}_{long}$ / s	4.92	4.43	10.75	8.43

First, comparing the distribution of peak area for Ser-Cβ peaks at each chemical shift, it can be seen that there is no significant difference within experimental error among all three ^{13}C labeled peptides. This implies that the chemical shift or the local conformation is independent of the position of the Ser-Cβ carbon in a chain. Therefore, it is suggested that the β-sheet A and B domains are formed independent of the location in the $(AGSGAG)_n$ sequence: in other words, the individual peaks of the Ser-Cβ carbons come from different domains, not from different positions within a chain.

To analyze the molecular mobility of each component in the Ser-Cβ carbon region, we measured T_1^C for each individual component, at 66.8 ppm, 64.2 ppm, 63.5 ppm and 61.4 ppm, in the Ser^{15}-Cβ peak of [3-^{13}C]Ser^{15}-$(AGSGAG)_5$, which was selected as a representative ^{13}C-labeled sample. The T_1^C plots are shown in Fig 2-7 and the T_1^C values are summarized in Table 2-2. From the log plots of peak intensity versus delay time (Fig. 2-7a), it is noted that three of the four peaks have multiple T_1^C components. The plots were fitted to two exponential decays as described above, and as a result, the ratio and T_1^C values of the shorter and longer relaxation time components ($T_1^C{}_{short}$ and

$T_1^C{}_{long}$) were obtained. The results are summarized in Table 2-2 and Fig. 2-7b. Only the peak at 63.5 ppm could be fitted by a single exponential function. The relative ratios of β-sheet with two components for the main peak at 64.2 ppm are very similar to the previous T_1^C results from non-labeled $(AGSGAG)_5$ peptide.[9]

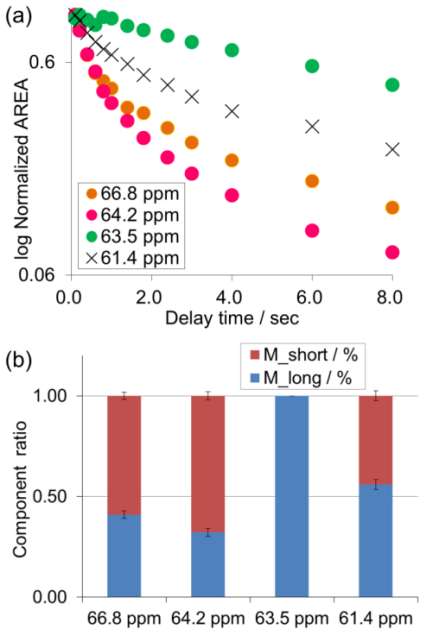

Figure 2-7. (a) Log plot of normalized area of each Ser-Cβ peak observed by T_1^C measurement of [3-^{13}C] Ser^{15}$(AGSGAG)_5$ via Torchia's sequence,[18] versus delay time. (b) Stacked column chart of components' ratios of shorter and longer T_1^C values for each Ser-Cβ peak.

From the previous results of T_1^C measurement with increasing temperature, it is known that the Ser-Cβ peaks have a mobility in the strong collision limit, that is, the components with longer T_1^C values show lower molecular mobility.[11] The peak at 63.5 ppm is a single component with the longest T_1^C value among the Ser-Cβ peaks (Table 2-2 and Fig. 2-7b), which means that this peak has a rigid conformation possibly arising from strong hydrogen bonds through the Ser side chain OH group. In our previous paper on the ^{2}H solid-state NMR analysis of [3,3-$^2H_2$] Ser-labeled *B. mori* silk fiber,[11] the

dominant side-chain χ1 dihedral angle was found to be the *gauche*+ conformation. Moreover, Fraser et al. have reported that the inter-sheet spacing became larger in the case of $(AGSGAG)_n$ than that in $(AG)_n$.[3] This is reasonable, since the side chain of Ser is bulkier than Ala, and so the replacement of Ala to Ser in the sequence is expected to distort the β-sheet packing. Moreover, in the Chapter 5, the Ser-Cβ peak at 63.5 ppm assigned to β-sheet B domain shows only a weak correlation with the Ala-Cβ peak of β-sheet B in the 2D DARR spectrum at the long mixing time (Chapter 5, Fig 5-1a), suggesting a relatively large distance between the two kinds of Cβ carbons in domain B. Taking these points together, we conclude that serine in domain B (the Ser-Cβ at 63.5 ppm) forms intermolecular hydrogen bonds most easily, thereby holding adjacent strands relatively rigidly and far apart.

On the other hand, in the main β-sheet peak at 64.2 ppm (the more highly populated β-sheet domain A), the longer T_1^C component was minor and the relaxation value was shortest (64.2 ppm in Table 2-2 and Fig. 2-7b). In the 2D DARR spectrum at 400 ms mixing time (Chapter 5, Fig. 5-1a), the Ser-Cβ peak at 64.2 ppm showed a strong correlation with the Ala-Cβ peak of β-sheet A. These results suggest that Ser residues in domain A are relatively stable but do not form stable hydrogen bonds with other chains. Interestingly, the random coil peak showed more percentage of the longer T_1^C component. This shows that in the random coil structure, the percentage of hydrogen-bonding conformation is larger than that of the main β-sheet A. It is worth noting that in a liquid NMR spectrum, random coil structure usually implies an isotropically mobile component, whereas in solid-state NMR spectra of materials with high hydrogen-bonding networks, random coil structure does not always mean a mobile component.

2-4. CONCLUSIONS

The author has assigned the heterogeneous Ser peaks observed in the ^{13}C solid-state CP/MAS NMR spectrum of [U-^{13}C] Cp-fraction in Silk II structure, which was obtained from [U-^{13}C] *B. mori* silk fiber. The Ser peaks comprise one random coil and three β-sheet components: A, B and a perturbed A structure at around 67 ppm. From T_1^C measurement, it was suggested that the preferred conformation of the Ser side-chain has comparatively little hydrogen bonding with other chains. The assignment in this chapter is important for the deconvolution of 2D peaks in chapter 4 and discussion of inter-domain structure in chapter 5.

REFERENCES

1. Marsh, R; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* 1955, *16*, 1-34.
2. Lotz, B.; Keith, H.D. *J. Mol. Biol.* **1971**, *61*, 201-215.
3. Fraser, B.; MacRae, T. P. *Conformation of Fibrous Proteins and Related Synthetic Polypeptides*; Academic Press: **1973**; pp293–343.
4. Asakura, T.; Yao, J.; Yamane, T.; Umemura, K.; Ulrich, A. *J. Am. Chem. Soc.* **2002**, *124*, 8794-8795.
5. Asakura, T.; Yao, J. *Protein Sci.* **2002**, *11*, 2706-2713.
6. Asakura, T.; Iwadate, M.; Demura, M.; Williamson, M. P. *Int. J. Biol. Macromol.* **1999**, *4*, 167-171.
7. Saito, H.; Tabeta, R.; Kuzuhara, A.; Asakura, T. *Bull. Chem. Soc. Jpn.* **1986**, *59*, 3383-3387.
8. Saito, H.; Ishida, M.; Yokoi, M.; Asakura, T. *Macromolecules* **1990**, *23*, 83-87.
9. Suzuki, Y.; Aoki, A.; Nakazawa, Y.; Knight, D. P.; Asakura, T. *Macromolecules* **2010**, *43*, 9434-9440.
10. Yamane, T.; Umemura, K.; Asakura, T.; *Macromolecules.* **2002**, *35*, 8831-8838.

11. Kameda, T.; Ohkawa, Y.; Yoshizawa, K.; Naito, J.; Ulrich, A. S.; Asakura, T. *Macromolecules* **1999**,*32*, 7166-7177.

12. Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344*, 631-637.

13. Zhao, C.; Asakura, T. *Prog. Nucl. Magn. Reson. Spectrosc.,* **2001**, *39*, 301-352.

14. Asakura, T.; Suzuki, Y. ; Nakazawa, Y.; Yazawa, K.; Holland G. P.; Yarger, J, L.; *Prog. Nucl. Magn. Reson. Spectrosc.,* **2013**, *69*, 23-68.

15. Asakura, T.; Kuzuhara, A.; Tabeta, R.; Saito, H. *Macromolecules* **1985**, *18*, 1841-1845.

16. Wang, J.; Okada, Y.; Li, W.; Yokoi, T.; Zhu, J. *J. Chem. Soc., Perkin Trans.* **1997**, *1*, 621-624.

17. Asakura, T.; Yao, J.; Ohgo, K.; Sugino, R.; Raghuvansh, K. *Biomacromolecules*, **2004**, *5*, 1763-1769.

18. Torchia, J. *J. Magn. Reson.* **1978**, *30*, 1969-1992.

19. Zhou, C.; Confalonieri, F.; Jacquet, M.; Perasso, R.; Li, Z.; Janin, J. *Proteins: Struct. Funct., Genetics* **2001**, *44*, 119-122.

20. Takahashi, Y.; Gehoh, M.; Yuzuriha, K. *Int. J. Biol. Macromol.* **1999**, *24,* 127-138.

21. Demura, M.; Minami, M.; Asakura, T.; Cross, T. A.; *J. Am. Chem. Soc.,* **1998**, *120*, 1300-1308.

22. Egawa, A.; Fujiwara, T.; Mizuguchi, T.; Kakitani, Y.; Koyama, Y.; Akutsu, H. *Proc. Natl. Acad. Sci.* **2007**,*104*, 790-795.

Chapter 3

Quantitative method to evaluate the structural agreement of tertiary packing structure in a β-sheet domain –Investigation by antiparallel β-sheet alanine tetramer (AP-Ala_4) with already-known structure

3-1. INTRODUCTION

In this chapter, the author uses the quantitative approach of solid state homonuclear correlation NMR spectrum for the evaluation of inter-carbon distance including tertiary structural packing geometry. Homonuclear correlation NMR is routinely implemented to obtain inter-nuclear distance information allowing for structural studies of a vast array of spin systems. In solution NMR, ^{1}H 2D homo-nuclear NOE methods are often used to measure distance constraints for structural elucidation.[1] On the other hand in solid materials, due to strong ^{1}H -^{1}H dipolar interactions, spectral resolution is usually poor which does not allow for separating individual peaks in a ^{1}H magic angle spinning spectrum. One robust method for obtaining internuclear distances in a solid sample involves ^{13}C homo-nuclear correlation NMR. These experiments often provide sufficient spectral resolution, and, combined with selective or uniform ^{13}C labeling methods, structural studies of solids are more readily approachable.

Among the currently available ^{13}C homonuclear correlation NMR methods, theoretical and simulation studies of proton-driven spin diffusion (PDSD) have been recently improved.[2] For example, theoretical study of PDSD via two-dimensional (2D) spectroscopy in solids has been presented by Suter and Ernst in 1985[3], combining with experimental data of a single crystal. Additional groups have explored the energy conservation process via

spatial separated spins[4] or spins with different frequency spins[5,6], based on experimental results from single molecules. One merit of the PDSD experiment is that any given peak intensity is not influenced by dipolar truncation. As a result, information through all inter-carbon distances of the sample can be obtained simultaneously.[7] Combined with molecular dynamics simulations, the PDSD experiment has been recently utilized to obtain the distance parameter for elucidating structure.[8] ^{13}C - ^{1}H dipolar assisted rotational resonance (DARR)[9-11] experiment is a second solid state ^{13}C homonuclear NMR experiment that makes use of proton-driven spin diffusion. With rotational resonance, however, the line width of C-H coupling spectrum becomes broadened. As a result of peak broadening, the sensitivity of correlation peaks may be improved Therefore, DARR experiments are very useful for detection of weak signals, and to our knowledge, this method has been used mainly for distance determination. In this chapter, I discuss the applicability for distance measurements by DARR, by comparing its results with the conventional PDSD results.

In previous PDSD studies, the build-up curves have been simulated and fit to the observed spectra where peaks are seperable[2,12]. But in many samples with iterative sequences such as silks, there are structural heterogeneities and many overlapping peaks which make analysis of the data challenging. As a solution to this problem, it is thought to be relevant to compare the experimental build-up curves with those calculated from candidate structures. In this chapter we introduce an example of how to evaluate the error by comparing experimental build-up curves of anti-parallel β-sheet alanine tetramer (AP-β-Ala_4) with that of an already-known β-sheet structure.[13]

In general AP-β-Ala_n peptides which are found in iterative sequences of many silks, serve as useful models for understanding the structure and dynamics of silk. The ^{13}C NMR spectra of an alanine monomer to alanine

trimer often show narrow NMR line widths and it is relatively easy to obtain the single crystals for X-ray diffraction analysis.[14] In the alanine tetramer, however, there are several peaks that overlap in the ^{13}C NMR data and it becomes difficult to prepare large crystals required for X-ray diffraction analysis. Alanine peptides longer than a hexamer show broaden ^{13}C spectral patterns reminiscent of a large polymer[14]. However, in our previous work, the heavy atom coordinates of the AP-β-Ala_4 structure, including the intermolecular structure, have been determined by X-ray diffraction with a small single crystal ($100 \times 200 \times 30$ μm^3).[14] Moreover, the ^1H coordinates of the AP-β-Ala_4 peptide have been determined from ^1H chemical shift assignments in a ^1H MAS spectrum by ultra-fast MAS and chemical shift calculations with the use of CASTEP.[13] Thus, AP-β-Ala_4 is the largest in the polyalanine peptide family with an already-known structure, and is well suited for the evaluation of errors of the build-up curves observed from polymer-like compounds with iterative sequences.

The central aim of this chapter is to report on the applicability of distance measurements by DARR and PDSD on a uniformly ^{13}C-labeled ([U-^{13}C]) AP-β-Ala_4, whose structural coordinates have been completely determined. In the build-up curve calculations we make use of the C-H coupling spectra for the zero-quantum line shape function and spin-lattice relaxation times, for obtaining relaxation matrix. In the comparison of DARR and PDSD, we characterize the error between experimental and simulated build-up curves (a term we refer to as the 'evaluation value' EV, defined below).

3-2. EXPERIMENTAL SECTION

3-2-1. Sample Preparation

Uniformly ^{13}C labeled ([U-^{13}C]) anti-parallel (AP) alanine tetramer (Ala_4) was synthesized from an [U-$^{13}C_3$] alanine monomer in the solid-phase. The terminal compound was treated with H_2O / EtOH = 1:1 to form an AP β-sheet fine crystal[13], and is denoted as the AP-β-Ala_4 peptide.

3-2-2. Solid-state ^{13}C NMR experiments

2D ^{13}C-^{13}C correlation NMR measurements.

The ^{13}C DARR (Figure 3-1a) and PDSD (Figure 3-1b) measurements of [U-^{13}C] AP-Ala_4 were performed on a Bruker DMX-500 spectrometer. The experiments were performed at Larmor frequency of 125 MHz for the ^{13}C nucleus, under magic-angle spinning (MAS) at 11 kHz with a 4 mm rotor at room temperature. Cross polarization (CP) was employed for sensitivity enhancement with high-power ^{1}H decoupling during signal acquisition. The ^{1}H powers for TPPM decoupling was 53.1 kHz. The CP contact time was 1.5 ms, and the repetition times for were 4.5 s. The number of scans for ^{13}C CP/MAS NMR measurement was 64. The number of scans for the direct dimension and the number of slices for the indirect dimension for ^{13}C DARR measurement was 8 and 512, respectively. The transmitter frequency was set to 100 ppm in both experiments. The mixing times were varied from 0.1 ms to 300 ms, and the ^{1}H power for DARR was set to that of the MAS rotor frequency. The C-H coupling spectra of DARR and PDSD for the zero-quantum line shape function were obtained from the experiments shown in Figure 3-1c and 3-1d, by modifying the ^{1}H decoupling power during t_1 and fixing the mixing time to 10 μs in the DARR and PDSD sequences.

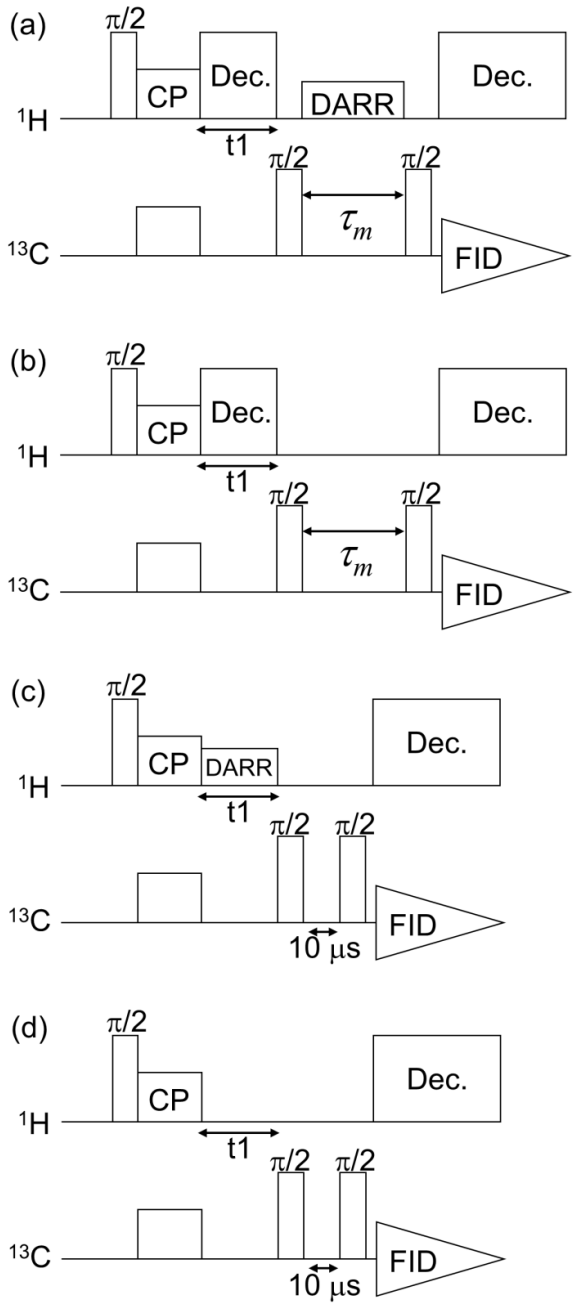

Figure 3-1. Pulse sequences implemented in this study. (a) DARR, (b) PDSD, 2D chemical-shift resolved CH dipolar experiments under (c) DARR and (d) PDSD condition. Experiments (c) and (d) were configured by modifying ^{1}H decoupling power during t_1 and fixing the mixing time to 10 μs in the 2D sequences (a) and (b).

T_1^C masurements.

The ^{13}C T_1 measurements of [U-^{13}C] and non-labeled AP-βAla$_4$ were conducted on a Bruker DMX-500 spectrometer operating at 125 MHz, under a MAS frequency of 11 kHz with a 4 mm rotor at room temperature, using a sequence reported by Torcha.[15] The delay times were varied from 0.1 ms to 10 s and the initial decays less than 1 s were used for the ^{13}C T_1 values. The inverse values of the ^{13}C T_1 for the build-up curve simulation are listed in Table 3-1. The^{13}C chemical shifts were calibrated indirectly through the adamantane methylene peak observed at 28.8 ppm relative to TMS (Tetramethylsilane) at referenced to 0 ppm.

3-2-3. Analysis

A. Deconvolution of peaks in 2D DARR spectra

To obtain the observed peak volume, the line widths and the chemical shifts were determined in each dimension (ω_1, ω_2) of the 2D DARR or PDSD spectra from the average of 14 build-up curve slices at each peak, and they were fixed and the intensities were fit. The parameters of up and down side in a 2D spectrum were not the same, so we needed to fit the (n, m) and (m, n) peaks individually. The fitting was performed using a Fortran program that assumed a two-dimensional Gaussian line shape. The volume obtained was normalized by the sum of diagonal peaks in the observed peak intensity matrix $\boldsymbol{M_{obs}}(0)$ obtained by the extrapolation of peak intensities from the DARR or the PDSD spectra.

B. Calculation of build-up curves though relaxation matrix

The underlying recoupling mechanism for ^{13}C-^{1}H dipolar-assisted rotational resonance (DARR)[9-11] can be classified as a combined mechanism of ^{1}H-driven and rotor-driven recoupling. The recoupling for the requisite residual energy for the ^{13}C-^{13}C flip-flop transition is provided from the ^{1}H reservoir as well as macroscopic sample rotation (MAS).[10] From the DARR

spectra, a build-up curve which correlates with polarization transfer rate for a ^{13}C-^{13}C pair can be obtained. Moreover, the polarization-transfer rate for a ^{13}C-^{13}C correlates well with the inter-nuclear distance of the ^{13}C-^{13}C pair.[10] Therefore, for the determination of intermolecular distances in a uniformly ^{13}C-labeled sample by using a 2D DARR spectrum, we can apply a polarization transfer matrix analysis[8] or in the case of a 2D PDSD spectrum the master-equation approach[12].

In the following, we outline the salient physics of the experiment and consider the ^{13}C-^{13}C polarization transfer. The polarization-transfer rate matrix $\boldsymbol{K}$ and the peak intensity matrix $\boldsymbol{M}(\tau_{\mathrm{m}})$ may be written as[8]

$$\boldsymbol{M}(\tau_m) = \exp(-\boldsymbol{K}\tau_m)\boldsymbol{M}(0) \qquad [1]$$

, where the $\boldsymbol{M}(0)$ is a matrix with the peak ratio in its diagonal elements. In my analysis I set the value of $\boldsymbol{M}(0)$ to that observed $\boldsymbol{M_{obs}}(0)$ in the experiment.

According to Dumez, J. N. and Emsley, L,[12] the factor k'_{pq} in the matrix $\boldsymbol{K}$ may be written as

$$k'_{pq} = \sum_{j=1}^{N_q} k_{a_1 b_j} \qquad [2]$$

, where the subscripts p or q donate each group of N_p or N_q equivalent spins, and the sum of the polarization constants between an a_1 spin in group p and a spin b_j in group q. In this study, we used the following diagonal elements k_{pp}^{diag} considering the decay in the experimental build-up curves at long mixing times.

$$k'_{pp} = \sum_{q=1}^{n} -k_{pq} \qquad (k'_{pp} \neq k_{pp}) \quad [3]$$

$$k_{pp}^{diag} = k'_{pp} - \frac{1}{T_{1,p}^{\mathrm{C}}} \qquad [4]$$

This equation was shown as proper by the comparison of build-up curve matching described in the section 3-3-2. The experimentally measured values of the carbon relaxation rates, $\frac{1}{T_{1,p}^{\mathrm{C}}}$, are listed in Table 3-1 both about

[U-^{13}C] and non-labeled AP-β-Ala$_4$. Finally, non-labeled data was used in the calculation as tested in the section 3-3-2.

Table 3-1. Carbon relaxation rates ($1/T_1^C$) used for the simulated build-up curves in this study for (a) [U-^{13}C] and (b) non-labeled AP-β-Ala$_4$.

(a)

Obs No.	$(1/T_1^C)$ /s^{-1}
1	0.995
2	1.619
3	1.617
4	1.652
5	1.970
6	2.680
7	3.214
8	1.577

(b)

Obs No.	$(1/T_1^C)$ /s^{-1}
1	0.033
2	0.062
3	0.032
4	0.031
5	0.165
6	7.137
7	4.207
8	1.175

From the above expressions any given element in the $\boldsymbol{K}$ matrix may be written as[12]

$$k'_{pq} = \frac{1}{15}\omega_{pq,\mathrm{eff}}{}^2\left(G_{a_1b_1}(\nu_r) + \frac{1}{2}G_{a_1b_1}(2\nu_r)\right) \qquad [5]$$

$$\omega_{pq,\mathrm{eff}}{}^2 = \sum_{j=1}^{N_q}\omega_{a_1b_j}{}^2 \qquad [6]$$

, where $\omega_{pq,\mathrm{eff}}$ is "effective dipolar coupling" resulting from the summation of polarization constants of single spin pairs and $G_{a_1b_1}(n\nu_r)$ is the zero-quantum lineshape function which is calculated as[8]

$$G_{a_1b_1}(n\nu_r) = \frac{1}{2\pi}\int_{-\infty}^{\infty}F_{a_1}(n\nu_r - \nu)F_{b_1}(\nu)d\nu \qquad [7]$$

The term ν_r in the above expressions is the rotor spinning frequency and $F_{a_1}(\nu)$ is the single-quantum dipolar lineshape function of spin a_1 under CH

coupling. In this study, the lineshapes for the convolution were determined from a 2D chemical-shift resolved CH dipolar spectrum[5] obtained by the experiment shown in Figure 3-1c (3-1d for PDSD).

From Eq [5] and [6] the distance $r_{pq,\mathrm{eff}}^{NMR}$, what we term the "effective distance" [12], may be written

$$r_{pq,\mathrm{eff}}^{NMR} = A_{pq}(k'_{pq})^{-1/6} \quad [8]$$

where

$$A_{pq} = \left(\frac{\mu_0 {\gamma_\mathrm{C}}^2 \hbar}{4\pi}\right)^{1/3} \left(\frac{1}{15}\left(G_{a_1 b_1}(\nu_r) + \frac{1}{2} G_{a_1 b_1}(2\nu_r)\right)\right)^{-1/6} \quad [9]$$

In the case of DARR, it has not yet been reported, to my knowledge whether [9] can be applicable for the distance measurements. In my algorithm, I set the distance limit value as 0.8 nm by testing the simulated build-up curves through the steps described in the section 3-3-2.

C. Comparison with the observed build-up curves

In this chapter, we determine the error range of the build-up curve analysis of DARR spectra of AP-β-Ala_4 whose structure is already known, and then, the error of *B. mori* Cp-fraction will be compared with it to evaluate structural agreement in the next chapter. In this evaluation, we used a metric which we term the evaluation value (EV) for each element which is defined by the expression

$$\mathrm{RMSD} = \sqrt{\sum_{i=1}^{N} \left\{ \boldsymbol{M}_{\mathrm{cal},i}(\tau_\mathrm{m}) - \boldsymbol{M}_{\mathrm{obs},i}(\tau_\mathrm{m}) \right\}^2 / N} \quad [14]$$

$$\mathrm{EV} = \mathrm{RMSD} / \boldsymbol{M}_{\mathrm{obs,MAX}}(\tau_\mathrm{m}) \qquad N = \text{number of mixing times } \tau_\mathrm{m} \quad [15]$$

, where $\boldsymbol{M}_{\mathrm{obs,MAX}}(\tau_\mathrm{m})$ is the maximum value of each observed element. In my algorithm, the RMSD is normalized to remove any differences in peak intensities.

3-3. Results and discussion

3-3-1. Comparison of PDSD and DARR experiments of AP-β-Ala_4

To investigate the build-up curve analysis, or 'polarization transfer matrix approach'[8], we compared the observed build-up curves and the errors between the experimental and simulated build-up curves of the DARR experiment with those of PDSD experiment by using the AP-β-Ala_4 whose atomic coordinates are known(Figure 3-2a).

The assignment of each peak in ^{13}C CP/MAS spectrum of the [U-^{13}C] AP-β-Ala_4 is shown in Figure 3-2b. In these spectra there were several cases of overlapping peaks (for example peak numbers 4 with peaks 2, 3 and 4-Cα spins). The build-up curves with overlapping components were analyzed by summing the all components included in a given peak. The numbering in Figure 3-2b spectrum corresponds to that of data of AP-β-Ala_4 shown in later.

Figure 3-3 shows the DARR spectra of [U-^{13}C] AP-β-Ala_4 and the comparison of the slice spectra of DARR and PDSD spectra at the 4-CO chemical shift. As expected, the higher enhancement of the cross peak intensities was detected in the slice spectra of DARR (Figure 3-3b, red), especially in the cross peak with large chemical shift difference such as (4-CO, 4-Cβ) peak. The enhancement was remarkable even at longer mixing times (Figure 3-3a, black, 50ms).

The observed build-up curves of each peak in the DARR and PDSD spectra are shown in Figure 3-4 in red and blue, respectively. The experimental data are plotted on a normalized scale with the maximum intensities in individual observed build-ups curves. The build-up curves from peak 5 (assigned as 1-Cα) were removed because the effect from the tail of the neighboring peak 4 was so large that peak 5 elements show decreasing tendency like the diagonal peaks. In Figure 3-4, the diagonal elements of the DARR matrix decreased more rapidly than those of PDSD, while the off-diagonal elements of DARR matrix increased more rapidly than those of PDSD.

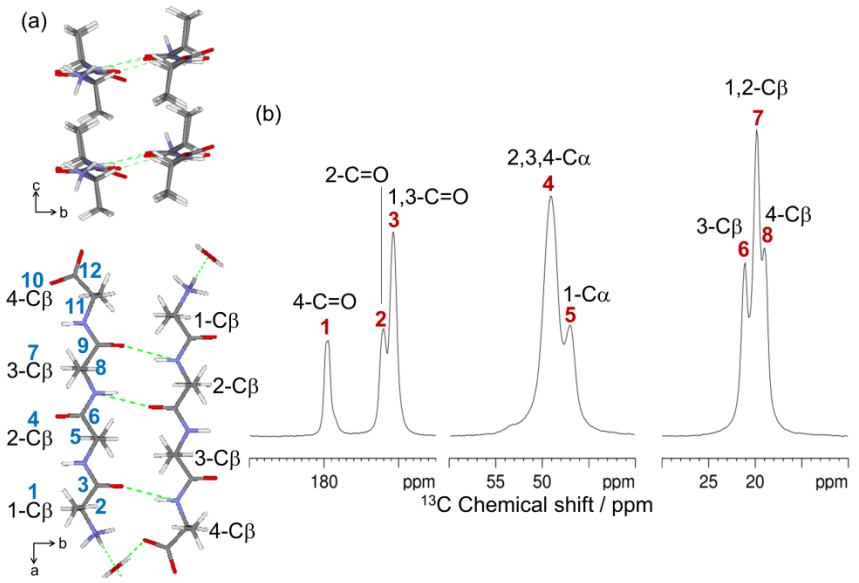

Figure 3-2. (a) X-ray structure and (b) the ^{13}C CP/MAS spectrum of [U-^{13}C] AP-β-Ala_4. The bold numbers with blues and red colors in (a) and (b) correspond to the numbers of element IDs in the following relaxation / effective distance and build-up curve / EV matrx.

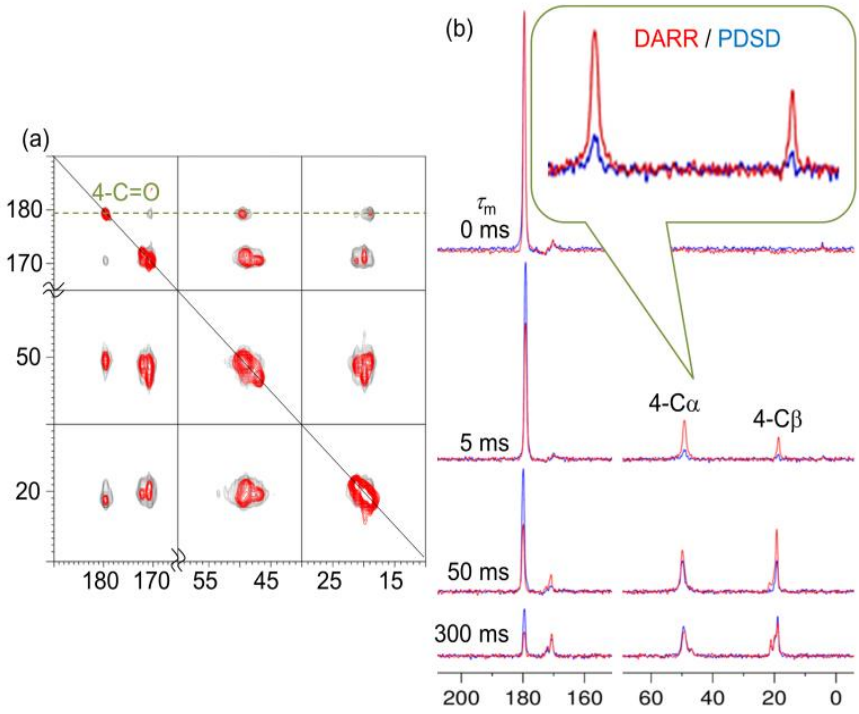

Figure 3-3. (a) 2D ^{13}C-^{13}C DARR spectra of [U-^{13}C] AP-β-Ala_4 observed at the mixing times (τ_m) of 5 ms (red) and 50 ms (black). (b) Comparison of the cross peak intensities between the DARR (red) and PDSD (blue) spectra at the cross-section of fourth Ala-CO peak in the [U-^{13}C] AP-β-Ala_4.

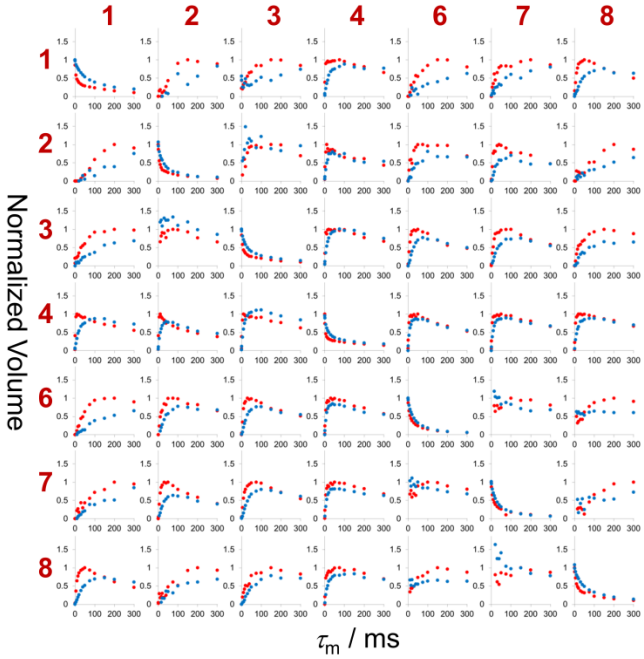

Figure 3-4. Comparison of build-up curves of AP-β-Ala_4 observed with DARR (red) and PDSD (blue). The numbers shown correspond to those of Figure 3b. The curves of number 5 were removed due to overlapping intensity with peak 4.

In 2D homonuclear experiments, there are two pathways for spin diffusion; spatial spin diffusion and spectral spin diffusion.[6] In the case of spatial spin diffusion, the nuclei are identical, the energy level differences are equivalent and the nuclear spin system is isolated from other lattice interactions (e.g. phonons). On the other hand in the case of spectral diffusion, the energy level differences for dipolar coupled spins are not equivalent and an additional reservoir is required for coherence transport to occur. In these experiments, the contribution from the atomic coordinates (spatial spin diffusion) are the same, so that the difference in the build-up curve rate in Figure 3-4 comes from spectral spin diffusion. For example, in the cross peaks corresponding to the most spectrally separated spins along the ^{13}C chemical shift, such as the (row, col) = (1, 8) build-up curve in Figure 3-4, the build-up rate difference between PDSD and DARR is the most remarkable among the other build-up curves. That is, the large chemical shift difference

leads to inefficient spectral spin diffusion in PDSD condition (Figure 3-4 blue). In general, to overcome the inefficient spectral spin diffusion between peaks with large chemical shift differences, the DARR experiments (Figure 3-1a and c) have been performed where experiments with a wide-range of line broadening is realized by its ^{1}H-^{13}C recoupling pulse[9,10] as shown in Figure 3-5a and c. Then, as expected, such an effective build-up rate by DARR was easily observed (see Figure 3-4).

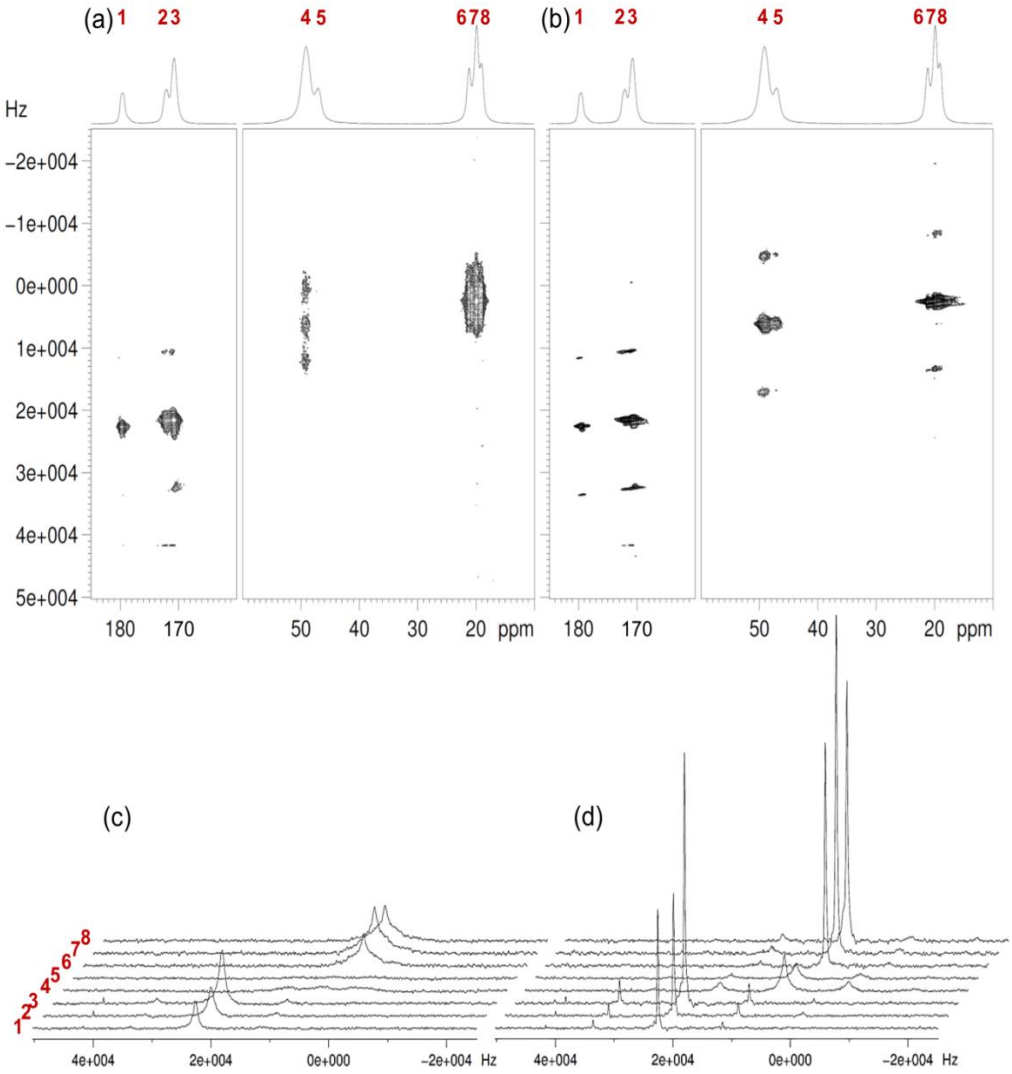

Figure 3-5. ^{13}C chemical shift resolved C-H coupling spectra under (a) DARR and (b) PDSD conditions. The cross-sections at each chemical shift for DARR and PDSD condition are displayed in (c) and (d), respectively.

In Figure 3-5a and b, ^{13}C chemical shift resolved C-H coupling spectra in DARR and PDSD conditions are shown, respectively. In addition, the cross-section along each ^{13}C chemical shift in the direct ω_2 dimension is shown in Figure 3-5c and d for both the DARR and PDSD conditions, respectively. The ^{1}H recoupling power for DARR (Figure 3-5a) is the same with that of 2D ^{13}C-^{13}C DARR experiments for obtaining the build-up intensities in Figure 3-4. In the t_1 dimension of the PDSD condition (Figure 3-1d), there is no ^{1}H decoupling pulse, so the spectrum affected just from MAS appears in the indirect ω_1 dimension. The detected spinning side bands in the indirect dimension of PDSD condition are also shown in Figure 3-5b and d. On the other hand, the detected spectral region in the DARR condition (Figure 3-5a) is wider than that of PDSD (Figure 3-5b), and the peak intensity is smaller in Figure 3-4c. This is because the C-H dipolar coupling was recoupled by DARR irradiation and then the C-H coupling spectrum was broadened.

On the basis of the cross-section spectra in Figure 3-5c and d, the zero-quantum lineshape functions (equation [7]) for the individual peaks were calculated. The simulated and experimental build-up curves of DARR and PDSD experiments are displayed in Fig 3-8a and b, with blue lines and red points. The parameters for the build-up curve calculation was optimized as will be described in the section 3-3-2. In the DARR build-up curves, the rapid build-up was well simulated even if the correlation arise from peaks with large chemical shift differences such as the (1, 8) build-up curve shown in Figure 3-4a. On the other hand, the (1, 8) build-up curve of the PDSD experiment in Figure 3-8b showed large differences between the observed and calculated build-up curve. For the CH_3-CO correlations that have a larger chemical shift difference, the discrepancy between the observed and calculated build-up curves had a higher discrepancy than those of DARR. In the CH_3-CO build-up curves, the intensities of experimental build-up curves became smaller than that of our simulated curves. In these simulations, we

used the experimental C-H coupling spectra, not calculated ones, for the zero-quantum line shape function. In the condition of the same spatial spin diffusion, the build-up error would be the same between DARR and PDSD if the other effect such as spectral spin diffusion contributed to the experimental build-up curves equally. But there were larger errors in the CH_3-CO build-up curves of the PDSD matrix than those in DARR. Therefore, by increasing the spectral overlap in C-H coupling spectrum with DARR irradiation, the reproducibility of the simulated build-up just from spatial information would also be affected and allow for more reliable quantitative distance determination.

The evaluation value (EV) for quantifying the error between the simulated and observed buildup rates for DARR and PDSD experiments are shown in Figure 3-6c and d, respectively. In the EV matrix, a smaller error is displayed with a darker color, and the coloring limit was set as EV = 0.32, as this was the sum of the average (0.20) and standard deviation (0.12) of EV matrix for DARR (Figure 3-6c) while those of PDSD (Figure 3-8d) were 0.37 and 0.23. In the EV matrix for PDSD, the EV showed larger values (corresponding to higher discrepancy) than that for DARR as a whole, especially in the lower elements. This higher discrepancy in PDSD matrix was partially due to more experimental build-up curve elements with unreliability in PDSD matrix than those in DARR matrix which had very close chemical shifts such as the (6, 7), (6, 8), (7, 6), (7, 8), (8, 6) and (8, 7) couplings shown in Figure 3-6b. This finding indicates that one may obtain more reliable peak intensities in DARR spectra because of its better sensitivity. Even if the unreliable peaks were removed, the averaged EV and standard deviation became 0.18 and 0.10 in the DARR matrix and then 0.34 and 0.25 in the PDSD matrix, respectively. As a result of the comparison for DARR (Figure 3-6c) and PDSD (Figure 3-6d), it appears that the error in the build-up curve analysis is less in DARR data than the conventional analysis in PDSD. Thus, the DARR experiment appears adequate and applicable for quantitative structural

evaluation via the build-up curve analysis.

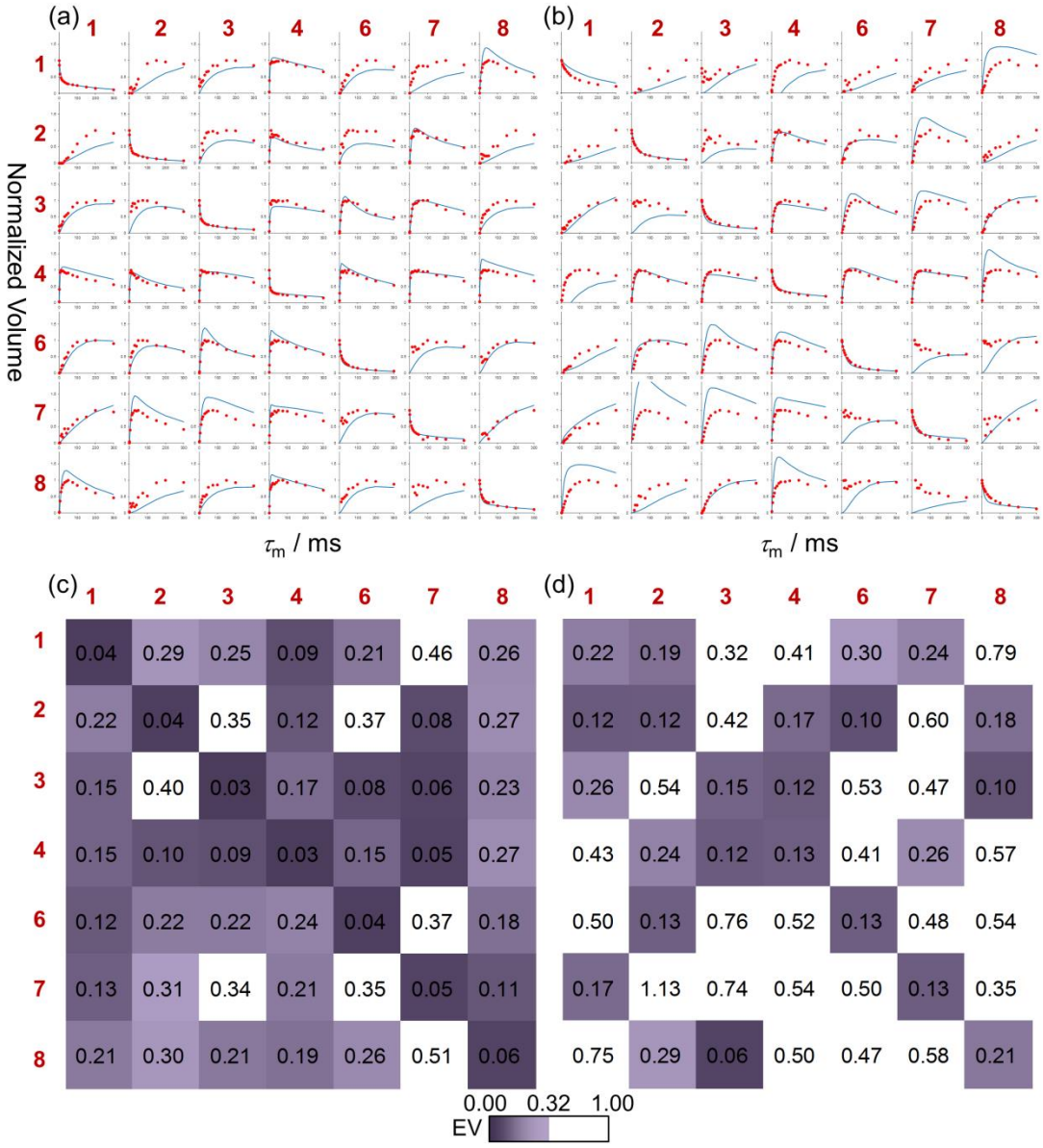

Figure 3-6. Build-up curve comparison until the mixing time of 300 ms. In this figure the color scheme shown corresponds to the observed (red point) and calculated (blue line) build-up curves of AP-β-Ala_4 for (a) DARR and (b) PDSD conditions. In (c) and (d) we show the EV matrices for DARR and PDSD. The color range was determined from the DARR results such as (the average 0.20) + (the standard deviation 0.12) = 0.32. The average EVs for DARR and PDSD were 0.20 and 0.37, respectively.

3-3-2. Optimization of parameters for build-up curves

Before the getting the final results for comparing DARR and PDSD data, we optimized parameters for calculation of the build-up curves. At first, we investigated the matrix elements should be included. This is because we could not obtain so good matching about both the data of DARR and PDSD calculated just from polarization transfer elements expressed with Eq. [3]. Figures 3-7 a and b display the result of build-up curve calculation by only polarization transfer matrix (Eq. [3]) without any diagonal T_1^C consideration. For example, the build-up curve elements (Figure 3-7a) with large EV (Figure 3-7b) in DARR data showed the large divergence in the late mixing times. This tendency was the same in PDSD data. This discrepancy of calculated build-up curve in the late mixing times makes it difficult to judge whether the error depends on the difference of molecular mobility or carbon atomic coordinate. In Figure 3-7, the build-up curves were calculated from the effective distance with the distance limit of 0.80 nm. Therefore, the comparison with the results in Figure 3-6 directly shows the remarkable effect by the addition of T_1^C factors in the diagonal elements. In fact, the averaged EVs for DARR and PDSD data were 0.24 and 0.36, respectively. Especially in the DARR data, the averaged EV became larger. Some previous papers didn't describe on the necessity of T_1^C factor in diagonal elements clearly. Therefore, we emphasis on the necessity of T_1^C consideration for the longer distance evaluation reflected in the late mixing times here, through the matching with experimental data.

Furthermore, we tested T_1^C data of which sample shows better results, non-labeled sample or uniformly ^{13}C-labeled sample. The build-up curve results calculated from the [U-^{13}C] labeled sample are shown in Figure 3-8. In the both of DARR (Figure 3-8a, c) and PDSD (Figure 3-8b, d) data, the matching of build-up curves was improved from the data in Figure 3-7. Interestingly, the EV results of PDSD data showed better matching in the calculation from the T_1^C values from [U-^{13}C] sample on the contrary to those

of DARR data. In DARR spectrum, the spatial ^{13}C-^{13}C spin diffusion occurs as a result of an additional energy reservation by ^{1}H-^{13}C recoupling.[10] On the other hand, the PDSD spectrum did not show large overlapping in the observed C-H coupling spectra (Figure 3-5d), then the C-C distance quantitativity of PDSD was a little lower than that of DARR (Figure 3-6). Therefore, better matching with using the [U-^{13}C] sample's T_1^C data (Figure 3-8) suggests that the lack of coherent transfer in PDSD data is complemented partially with contribution from molecular mobility.

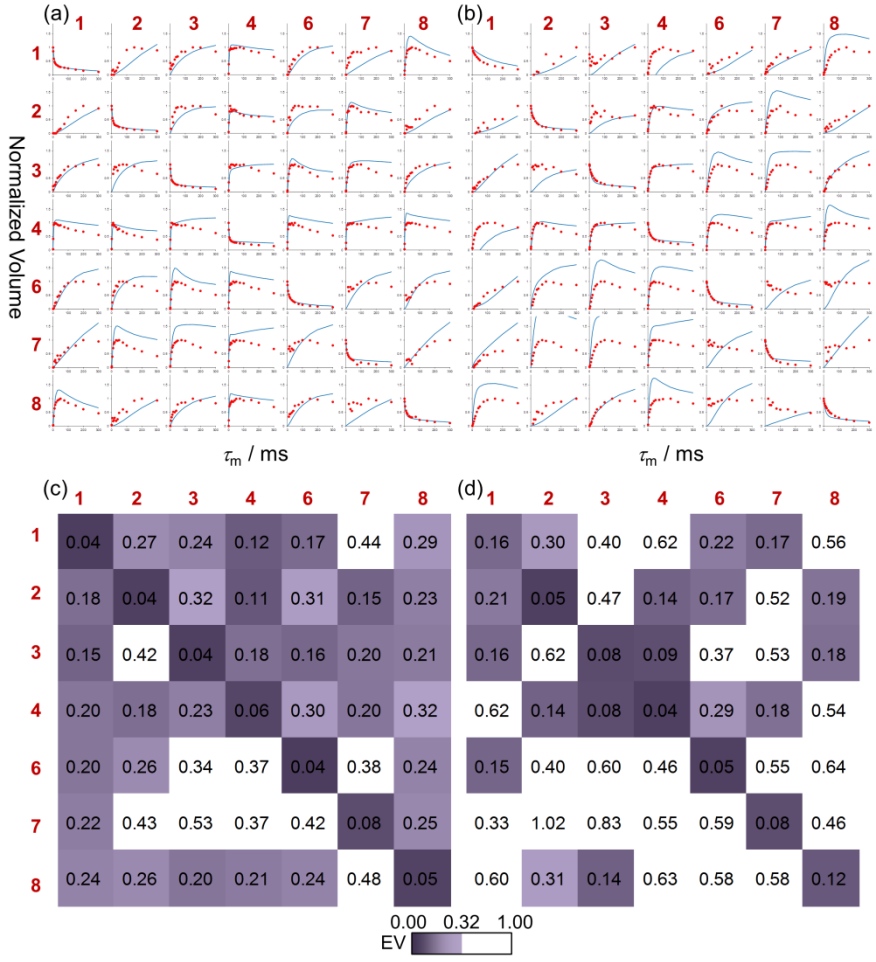

Figure 3-7. Build-up curve comparison calculated without T_1^C factors of diagonal elements until the mixing time of 300 ms. In this figure the color scheme shown corresponds to the observed (red point) and calculated (blue line) build-up curves of AP-β-Ala_4 for (a) DARR and (b) PDSD conditions. In (c) and (d) we show the EV matrices for DARR and PDSD. The average EVs for DARR and PDSD were 0.24 and 0.36, respectively.

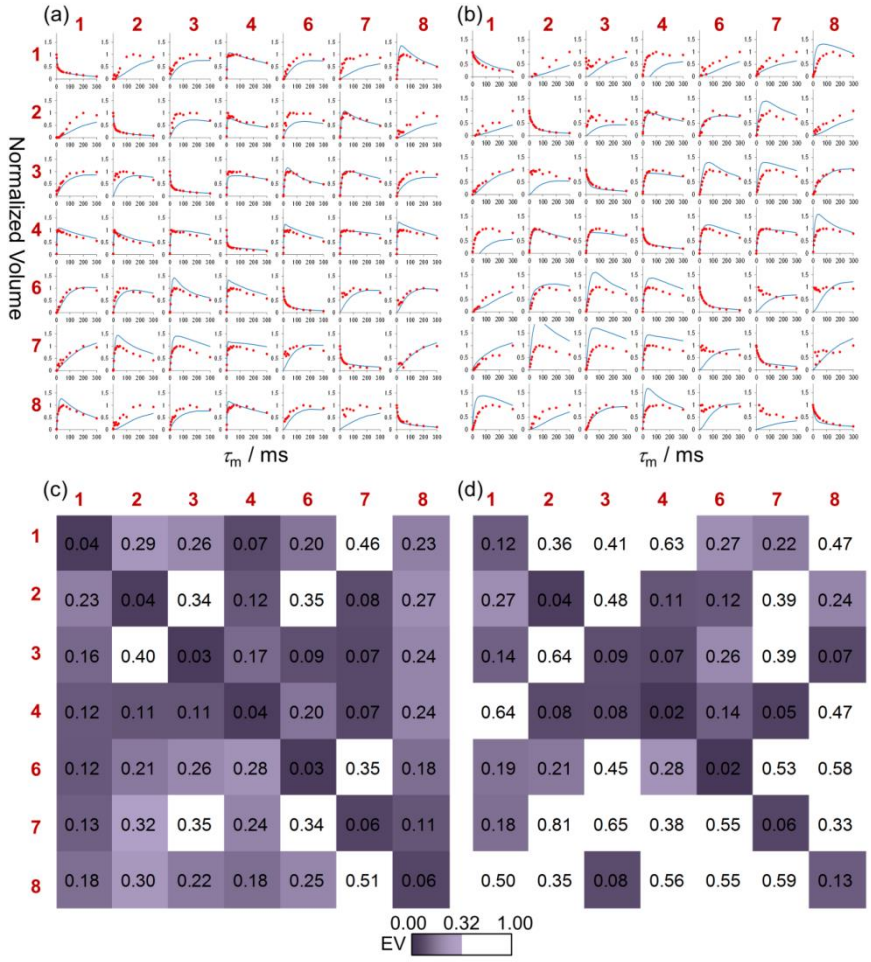

Figure 3-8. Build-up curve comparison until the mixing time of 300 ms calculated using T_1^C values measured with [U-^{13}C] AP-β-Ala$_4$. In this figure the color scheme shown corresponds to the observed (red point) and calculated (blue line) build-up curves of AP-β-Ala$_4$ for (a) DARR and (b) PDSD conditions. In (c) and (d) we show the EV matrices for DARR and PDSD. The average EVs for DARR and PDSD were 0.21 and 0.32, respectively.

In the next, we needed to know C-C distance which should be considered in the build-up calculation. We investigated the limit for our experimental condition thorough the following steps: (1) the distance limits for the effective distance calculation were varied from 0.35 nm to 1.00 nm for AP-β-Ala$_4$ coordinates as displayed in Fig 3-9. (2) From these effective distances of Fig 3-9, all the build-up curves were simulated and compared with the experimental data (Fig 3-10). And then, the distance limit was set to

a value where the error between the experimental and simulated data saturates (Fig 3-11).

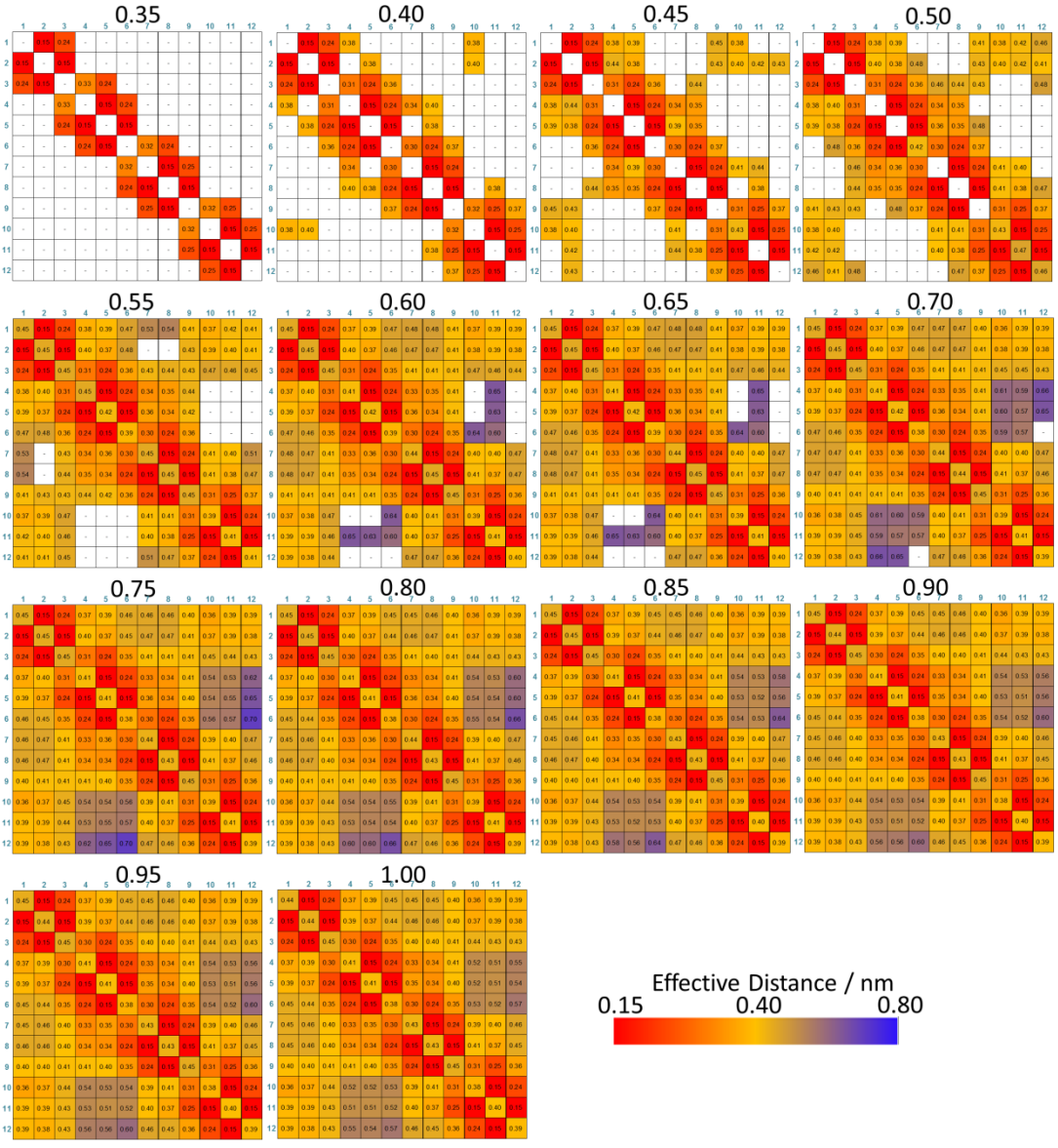

Figure 3-9. Series of effective distance matrix used in the test for distance limit to be considered. The number on the top of each matrix shows C-C distance limit (nm) considered in the calculation of the effective distances. Index of each matrix element shows the index number of spins displayed in Fig 3-2.

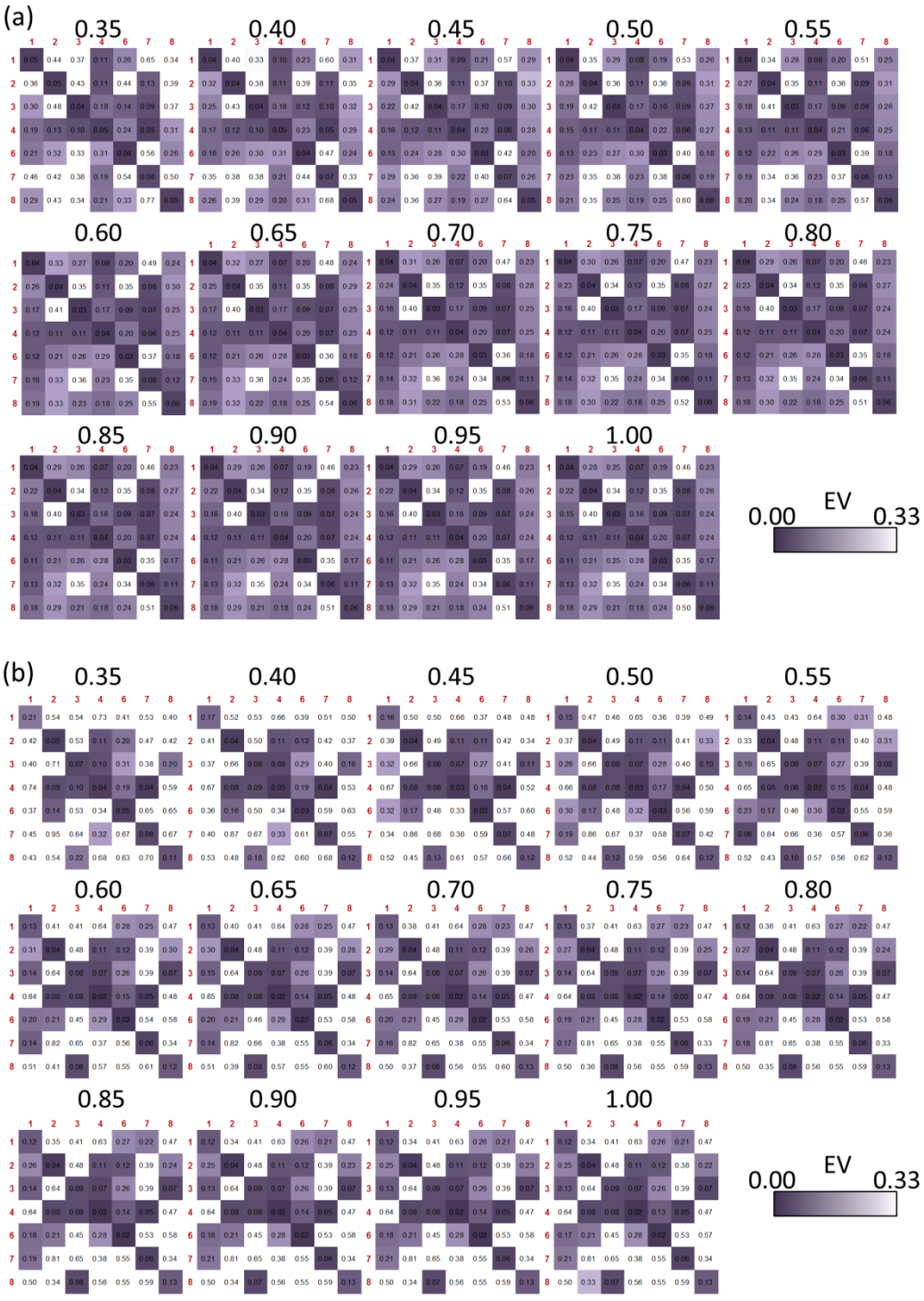

Figure 3-10. EV matrix of (a) DARR and (b) PDSD obtained by using each effective distance shown in Fig 3-9. The number on the top of each matrix shows C-C distance limit (nm) considered in the calculation of the effective distances.

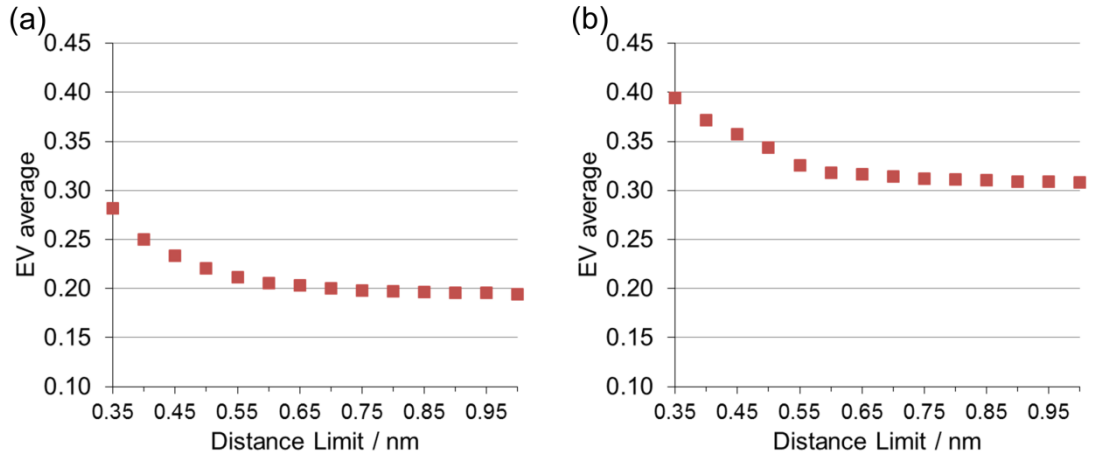

Figure 3-11. The plots of averaged EV of (a) DARR and (b) PDSD data versus C-C distance limit which was considered in the build-up curve calculation. Both the plots are saturated at 0.8 nm.

Also in Figure 3-10, EV became smaller in the DARR data (Figure 3-10a). Focused on the change of the averaged EV in the both data (Figure 3-11), it converges until the mixing time of 0.80 nm. About the converged EV, the value for DARR (Figure 3-11a) showed smaller value 0.20 than 0.32 for PDSD.

3-4. Conclusion

We have simulated build-up curves of DARR and PDSD spectra of [U-^{13}C] AP-β-Ala$_4$ with already-known structure, and discussed the applicability of DARR experiments for quantitative structural evaluation. The results showed smaller errors in DARR data, thus, DARR experiments appear better suited for structural evaluation rather than PDSD experiments.

References

1. Wuthrich, K.; Wider, G.; Wagner, G.; Braun, W. *J. Mol. Biol.* **1982**, *3*, 311-319.
2. Dumez, J. N.; Halse, M. E.; Butlerz, M. C.; Emsley, L. *Phys. Chem. Chem. Phys.* **2012**, *14*, 86–89.
3. Suter, D.; Ernst, R. R. *Phys. Rev. B* **1985**, *32*, 5608–5627.
4. Henrichs, M. P.; Linder, M.; Hewitt, M. J. *J. Chem. Phys.* **1986**, *85*, 7077－7086.
5. Kubo, A.; McDowell, C. A. *J. Chem. SOC., Faraday Trans. I* **1988**, *84*, 3713－3730
6. Kubo, A.; McDowell, C. A. *J. Chem. Phys.* **1988**, *89*, 63－70.
7. Grommek, A.; Meier, B. H.; Ernst, M. *Chem. Phys. Let.* **2006**, *427,* 404–409.
8. Egawa, A.; Fujiwara, T.; Mizuguchi, T.; Kakitani, Y.; Koyama, Y.; Akutsu, H. *PNAS* **2007**, *104*, 790－795.
9. Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, 344, 631－637.
10. Takegoshi, K.; Nakamura, S.; Terao, T. *J. Chem. Phys.* **2003**, *118*, 2325－2341.
11. Ohashi, R.; Takegoshi, K. *J. Chem. Phys.* **2006**, *125*, 214503.
12. Dumez, J. N.; Emsley, L. *Phys. Chem. Chem. Phys.,* **2011***, 13,* 7363–7370.
13. Asakura, T.; Yazawa, K.; Horiguchi, K.; Suzuki, F.; Nishiyama, Y.; Nishimura, K.; Kaji, H. *Biopolymers* **2013**, *101*, 13-19.
14. Asakura,T.; Okonogi, M.; Horiguchi, K.; Aoki, A.; Saito, H.;Knight, D. P.; Williamson, M. P. *Angew Chem Int Ed Engl* **2012**,*51*, 1212-1215.
15. Torchia, J. *J. Magn. Reson.* **1978**, *30*, 1969－1992.

Chapter 4

An application of ^{13}C-^{13}C DARR build-up curve approach for the candidate models with chemical shift agreements and method development for the *B. mori* Cp-fraction with Silk II form

4-1. INTRODUCTION

In the previous chapter, the permissible error rang for the build-up curve analysis by DARR spectrum has been determined. In this chapter, the author applies it for the real system of *B. mori* Cp-fraction that contains many overlapping peaks in the ^{13}C NMR spectra, and then clarifies and overcomes the problem which should be considered in the real system. We have discussed on the Silk II structure in detail so far and even the structure in β-sheet crystallite is still not conclusive as described in chapter 1.[1,2] However, the build-up curve approach constructed in the chapter 3 needs the carbon atomic coordinate whose chemical shift tentatively assigned for some explicable results such as the combination of selectively ^{13}C-labeling and X-ray diffraction analysis in AP-β-Ala_4 and chemical shift calculation of a certain model. Therefore, the author uses the $(AG)_{15}$ model reported, whose chemical shift has been shown to agree well with the observed chemical shift including ^{1}H nucleus most recently, as a candidate model here.

4-2. EXPERIMENTAL SECTION

4-2-1. Sample Preparation

The [U-^{13}C] *B. mori* Cp-fraction (Silk II), which is the crystalline fraction of *B. mori* fibroin with Silk II structure, was prepared as in the previous chapter 2.[3]

4-2-2. Solid-state ^{13}C NMR experiments

2D ^{13}C-^{13}C correlation NMR measurements.

The ^{13}C DARR measurements[4-6] of [U-^{13}C] *B. mori* Cp-fraction were performed on a JEOL ECA 500 spectrometer. The experiments were performed at Larmor frequency of 125 MHz for the ^{13}C nucleus, under magic-angle spinning (MAS) at 11 kHz with a 4 mm rotor at room temperature. Cross polarization (CP) was employed for sensitivity enhancement with high-power ^{1}H decoupling during signal acquisition. The ^{1}H power for TPPM decoupling was 86.2 kHz. The CP contact time was 1.5 ms, and the repetition time was 4 s. The number of scans for the direct dimension and the number of slices for the indirect dimension for ^{13}C DARR measurement was 8 and 512, respectively. The transmitter frequency was set to 100 ppm. The mixing times were varied from 0.1 ms to 300 ms, and the ^{1}H power for DARR was set to that of the MAS rotor frequency. The C-H coupling spectrum of DARR for the zero-quantum line shape function was obtained from the experiments by modifying the ^{1}H decoupling power during t_1 and fixing the mixing time to 10 μs in the DARR sequence.

T_1^C masurements.

The ^{13}C T_1 measurement of non-labeled *B. mori* Cp-fraction was conducted on a Bruker DMX-500 spectrometer operating at 125 MHz, under a MAS frequency of 11 kHz with a 4 mm rotor at room temperature, using a sequence reported by Torcha.[7] The delay times were varied from 0.1 ms to 10 s and the initial decays less than 1 s were used for the ^{13}C T_1 values. The

inverse values of the ^{13}C T_1 for the build-up curve simulation are listed in Table 4-1. The^{13}C chemical shifts were calibrated indirectly through the adamantane methylene peak observed at 28.8 ppm relative to TMS (Tetramethylsilane) at referenced to 0 ppm.

4-2-3. Analysis

The deconvolution, build-up curve calculation and error evaluation processes are described in the previous chapter. However, there are some modifications in the build-up curve analysis as mentioned below.

Table 4-1. Carbon relaxation rates ($1/T_1^C$) used for the simulated build-up curves in this study for [U-^{13}C] *B. mori* Cp-fraction (Silk II) measured by non-labeled Cp-fraction.

Obs No.	$(1/T_1^C)$ /s^{-1}
1	0.079
2	0.044
3	1.164
4	0.056
5	0.056
6	0.053
7	1.284
8	2.859

For the explanation of modified process, Eqs [8] ("effective distance") and [9] are rewited.

$r_{pq,\mathrm{eff}}^{NMR} = A_{pq}(k'_{pq})^{-1/6}$ [8]

where

$$A_{pq} = \left(\frac{\mu_0 \gamma_C{}^2 \hbar}{4\pi}\right)^{1/3} \left(\frac{1}{15}\left(G_{a_1 b_1}(\nu_r) + \frac{1}{2} G_{a_1 b_1}(2\nu_r)\right)\right)^{-1/6} \qquad [9]$$

The build-up curves for the DARR spectra of [U-^{13}C] *B. mori* Cp-fraction appear to indicate at least two β-sheet components, which is very similar to the lineshapes in Ala-Cβ region of solid ^{13}C CP/MAS spectra of $(AG)_{15}$ with Silk II form.[3,8,9] Thus their basic structures appeared to be quite similar. One of our proposed models (Figure 4-1) has shown to have a good result from its latest chemical shift calculation,[1] so the calculation of the build-up curve based on the carbon coordinate of the model (Figure 4-1) was performed by the following summation.

$$\boldsymbol{M}_{\mathrm{cal}}(\tau_{\mathrm{m}}) = \boldsymbol{M}_{\mathrm{cal,A}}(\tau_{\mathrm{m}}) + \boldsymbol{M}_{\mathrm{cal,B}}(\tau_{\mathrm{m}}) \qquad [10]$$

$$\boldsymbol{M}_{\mathrm{cal,A}}(\tau_{\mathrm{m}}) = \exp(\boldsymbol{K}_{\mathrm{A}}\tau_{\mathrm{m}})\boldsymbol{M}_{\mathrm{A}}(0) \qquad [11]$$

$$\boldsymbol{M}_{\mathrm{cal,B}}(\tau_{\mathrm{m}}) = \exp(\boldsymbol{K}_{\mathrm{B}}\tau_{\mathrm{m}})\boldsymbol{M}_{\mathrm{B}}(0) \qquad [12]$$

The ratio of the two β-sheet structures denoted A and B were determined from the ratio of diagonal elements in $\boldsymbol{M}_{\boldsymbol{obs}}(0)$. Here, we considered the models A and B as individual domains with large size and no contact each other. By applying the error evaluation after the summation, we can discuss and construct the method to evaluate structural agreement as a whole matrix or spectrum. Even though the detailed and candidate β-sheet models for Silk II structure have been proposed and discussed as the simplest model compound $(AG)_n$ generally,[1,2,10] in this work I focus on the method to verify and complement these models in the real *B. mori* Cp-fraction system including also Ser residues.

In the investigation for constructing the build-up curve method for *B. mori* Cp-fraction, we attempted 1) the application of the same effective distance with Ala residues in the $(AG)_n$ model (5 spin system) to Ser residues (8 spin system), improvement of the method by considering 2) stoichiometry aspects and difference of dipolar coupling probability among intra- and inter-chain occurring by Ser insertion into the AG model, 3) effect to apparent dipolar

coupling by the difference of ^{13}C-labeled percentage, and 4) effect to relaxation matrix by the difference of detailed and restricted spin system in one chain sequence in the presence of Ser. For the test of step (2), we used the following equation [13] for calculating the effective distance in the *B. mori* Cp-fractoin system from the proposed $(AG)_{15}$ model.

$$r_{pq,\mathrm{eff}} = \left(\sum_{j=1}^{x_j} \left(r_{a_1 b_j} \right)^{-6} + \sum_{j=1+x_j}^{Nq} \left(P_{a_1} \cdot P_{b_j} \right)^2 \left(r_{a_1 b_j} \right)^{-6} \right)^{-1/6} \qquad [13]$$

, where p, q: spin which is not separable in the observed spectra (8spin system in Figure 4-3a), $\{a_1\}$,$\{b_j\}$: the most center spin p and each spin of q which exists within 0.80 nm radius from $\{a_1\}$, N_q: number of spin q, L_p, P_{a_1}, P_{b_j}: probability to dipole-couple with spin p, q in a given intermolecular chain. This probability is 1/3, because the patterns of the interacting unit in the other chain consist of three variations as shown in later (Figure 4-3b and 4-5b), the x_j: number of spin $\{b_j\}$ which dipole–couples with spin $\{a_1\}$ on the intra-chain.

For the investigation of step (3), the equation [13] was improved to [14], considering ^{13}C-labeled percentage of each residue in the real *B. mori* Cp-fraction system.

$$r_{pq,\mathrm{eff}} = \left(\left(L_p \cdot L_q \right)^2 \left(\sum_{j=1}^{x_j} \left(r_{a_1 b_j} \right)^{-6} + \sum_{j=1+x_j}^{Nq} \left(P_{a_1} \cdot P_{b_j} \right)^2 \left(r_{a_1 b_j} \right)^{-6} \right) \right)^{-1/6} \qquad [14]$$

, where L_q: ^{13}C-labeled percentage of spin p, q, The ^{13}C-labeled percentage was determined from the J-coupling peak percentage in the solution NMR spectra of the [U-^{13}C] *B.mori* silk solution (Table 4-2).

For testing possibility by the effect in (4), we distinguished Ala spins at back and forth of Ser residue in a chain and considered as 15 spin system (p, q: magnetically equivalent spin). The details of spin numbering are described later (Figure 4-5a).

Table 4-2. ^{13}C-labelled percentage in each residue obtained from the peak analysis of the solution NMR spectrum measured with regenerated [U-^{13}C] *B. mori* silk solution (D_2O).

^{13}C labeled %	Ala	Gly	Ser
Cβ	84.7	-	81.9
Cα*	85.6~87.7	86.3	83.0~88.7
CO	88.6	88.7	87.2

* For the Cα, the average value of the percentages determined from Cβ-Cα and CO-Cα couplings were used.

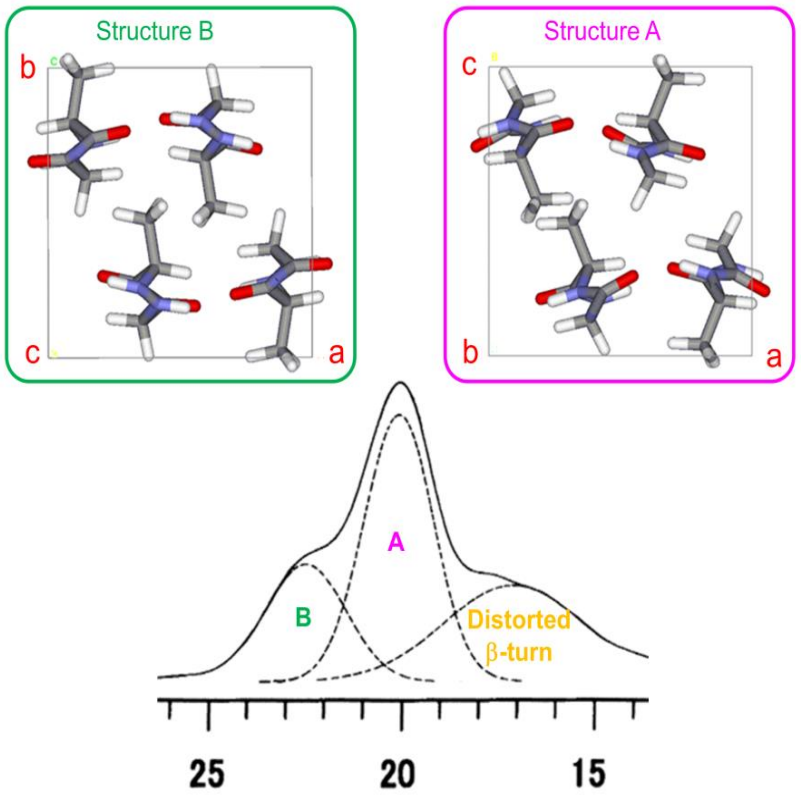

Figure 4-1. Silk II model structure of poly AG peptide (upper side), which is the model peptide for the *B. mori* Cp-fraction and its Ala-Cβ region in ^{13}C CP/MAS spectrum. These two models have been proposed in our laboratory based on the reported X-ray lattice parameter and energy minimization by CASTEP program. The calculated chemical shifts have explained the observed chemical shifts of ^{1}H, ^{13}C and ^{15}N very well. The detail of these models has been described in the literature.[1]

4-3. Results and discussion

4-3-1. Spectral overlap in the 2D ^{13}C-^{13}C DARR spectrum of *B. mori* Cp-fraction.

The ^{13}C CP/MAS and DARR (mixing time 10 ms) spectrum of [U-^{13}C] Cp-fraction are shown in Figure 4-2a and b, respectively[3]. The Cp-fraction contains two kinds of β-sheet structures we denote A , B and distorted β-turn structure, as reflected in the ^{13}C chemical shift 20.3 ppm, 19.6 ppm and 17.5 ppm of the Ala-Cβ region.[3] On the other hand, the other ^{13}C peaks of β-sheet structures in the spectra cannot be separable with their chemical shifts even in the 2D experiment. For example, Ala-Ca chemical shifts of β-sheet A and B showed the same values, which could be detected from (Ala-Cβ, Ala-Cα) cross peaks displayed as intersections in Figure 4-2b. Then, the β-sheet A and B correlations with such as Gly peaks could not be detected as separated peaks even in the 2D spectrum. Therefore, one needs to consider the overlap between the two β-sheet structures in the build-up curve analysis with all the other carbons. Also, the Ser-peaks with Cβ side-chain carbon, there was not enough spectral resolution in the 2D DARR spectra for obtaining the experimental intensities in this study, so the summed values were used as the experimental data for β-sheet peaks of Ser although the Ser β-sheet peaks may also require a more detailed signal decomposition as shown in the previous chapter.[3]

To evaluate the structural agreement with the candidate Silk II models in the local intermolecular geometry for all carbons, we measured the DARR spectra and obtained the experimental build-up curves (Figure 4-4 and 4-6, red dots denote experimental data). The line width for each correlation was determined from peak deconvolution of sliced 1D spectrum at each peak top in some of mixing time's spectrum. The averaged peak width for each peak was used to obtain the volume intensity of each cross peaks on a 2D spectrum. About a diagonal peak, the volume intensity was obtained from fitting of 2D Gaussian with projection length to the diagonal line as the

longer width and averaged width for all the peaks as the shorter width.

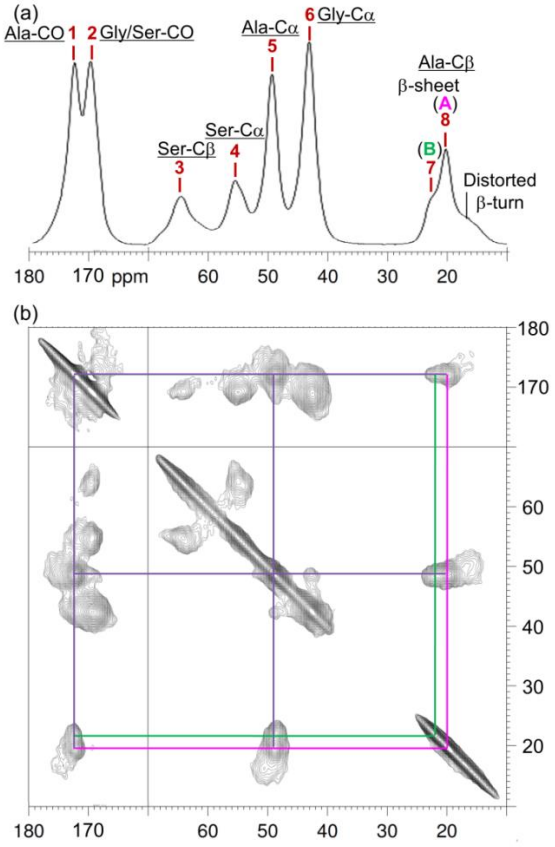

Figure 4-2. (a) ^{13}C CP/MAS and (b) 2D ^{13}C-^{13}C DARR spectra of [U-^{13}C] *B. mori* CP-fraction observed at τ_m 10 ms.[3] Except for Ala-Cβ peaks, the β-sheet peaks are completely overlapping even in the 2D spectrum (b). The bold numbers in (b) correspond to the observed intensity's matrix numbers.

4-3-2. Modification of effective distances calculated from the $(AG)_n$ model in to $(AGSGAG)_n$ sequence more proper for the real *B. mori* Cp-fraction system.

We applied the effective distances from Ala residue calculated from the carbon atomic coordinate in the two kinds of $(AG)_n$ structures to effective distances of Ser in the effective distance matrix for *B. mori* Cp-fraction. The $(AG)_n$ structure has 5 carbon spins in the system, so by dealing with one of Ala in the sequence as Ser, the number of spins to be considered increases to 8 spins (Figure 4-3a). At first, we used the same effective distances with Ala

in the $(AG)_n$ structures A and B, and then calculated the build-up curves from the equation [10] (Figure 4-4a, blue lines in the left side matrix). The alternation of the value is the simplest way. There is no elements for (7, 8) and (8, 7) because we supposed here that the β-sheet structures A and B has individual and large domains, respectively, and that there is no contact each other. The shapes of calculated build-up curves did not match at all (Figure 4-4a, left) and few of EV elements showed the permissible value (Figure 4-4a, left). This is because the build-up curve behavior is related to number of spins or dipolar coupling pairs expressed in the Eqs. [5] and [6]. Namely, in the case of just using the Ala effective distances as those of Ser, the considered dipolar coupling pairs were too much each other.

Then, we considered stoichiometry in the case of $(AGSGAG)_n$ sequence more carefully and distinguished the difference of dipolar coupling probability among intra- and inter-chain (Figure 4-3b). When one of Ala residues in a $(AG)_3$ sequences is altered with Ser, the countable numbers as Ala spins changes in a sequence. About the dipolar coupling in a chain, there is only one combination of dipolar coupling pair along the sequence, so the probability of the pair should be 1 (the first term in Eq. [13]). On the other hand, about the one pair of inter-chains (or intra-chains with very long term for loop), the pattern to close each other increase up to 9 because there are possibilities for the order of sequence to start even if the structural coordinate in the $(AG)_n$ model is determined as only one pattern (Figure 4-3b and the second term of Eq. [13]).

The calculated build-up curves from the effective distance matrix considering such a stoichiometry are shown in Figure 4-3b. The matching with the experimental data was improved but more than half of all elements did not reach the permissible EVs. In this calculation, we have supposed that all the carbon spins in the system are ^{13}C-labeled. In actual, it has been confirmed from solution ^{13}C NMR measurement of *B. mori* silk solution used really in the DARR experiments that the ^{13}C-labeled percentage was high

but not 100 % for all the carbon spins (Table 4-2). Therefore, it is expected that the non-fully labeled ^{13}C percentage in the real system would influence the relaxation behavior based on ^{13}C-^{13}C dipolar interaction. The strength of ^{13}C-^{13}C dipolar interaction depends on the square of gyromagnetic ratio γ_C which is an intrinsic value for one ^{13}C spin. Hence, the ^{13}C-labeled percentage becomes the probability that the spin in a certain pair has γ_C detectable with NMR. Namely, the square of effective dipolar coupling strength (inverse of distance cubed) shown in the blanket of Eq. [13] is to be multiplied by the square of product of ^{13}C-labeled percentages in the focused spin pair (Eq. [14], the first term).

The build-up curves calculated by the effective distance matrix considering ^{13}C-labeled percentage for all the spins are shown in Figure 4-4c (Eq. [14], 8 spin system). The matching with the experimental data was further improved and almost all the elements of upper side in the EV matrix (Figure 4-4c, right) showed the permissible values. However, there still have been large EVs in the lower elements.

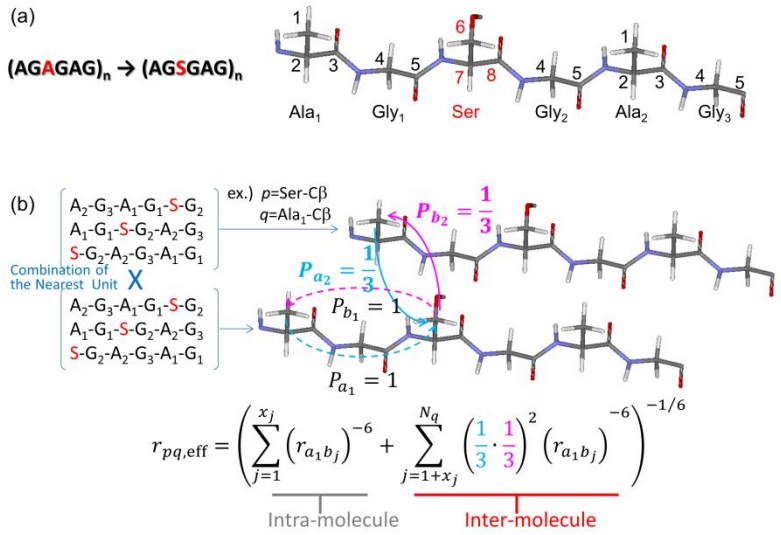

Figure 4-3. Schematic of the effective distance calculation, based on the probability of dipole-dipole couplings in *B. mori* Cp-fraction. (a) Spin numbering considered as 8 spin system. (b) Different combination and probability in the calculation of effective distance between intra- and inter-chain distances based on the $(AGSGAG)_n$ system. The equation in (b) corresponds to Eq. [13].

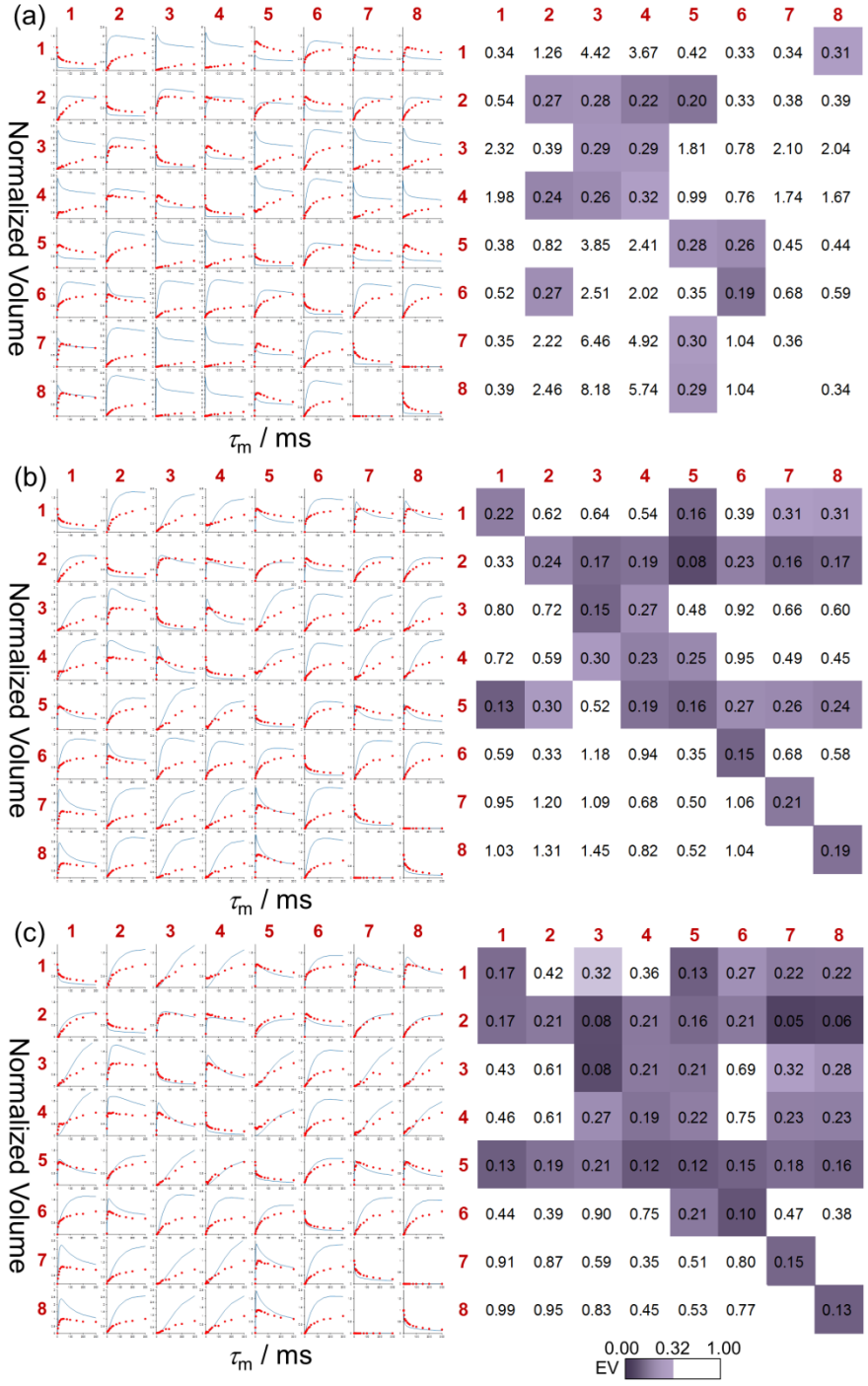

Figure 4-4. The build-up curves (left) and EV (right) matrices for DARR data of [U-^{13}C] *B. mori* Cp-fraction (considered as 8 spin system). The results of (a) just the alteration of effective distances for Ser by the same effective distances with Ala residue, (b) consideration of stoichiometry by insertion of Ser into the previous $(AG)_{15}$ model, and (c) consideration of ^{13}C-labeled percentage in each residue. The color scale in the EV matrix is the same with that of AP-β-Ala_4 in Figure 3-6c and d.

As the remained possibilities of such mismatching in the build-up, there were the two possibilities; the mistakes of ^{13}C-labeling consideration in Eq. [14] which has been newly introduced in this work, and the lack of consideration of spin environment after Ser alternation into the $(AG)_n$ model. In the previous solution NMR study of *B. mori* silk solution, it has been reported that the carbonyl region, which has been detected as on signal in the solid ^{13}C spectrum (Figure 4-2), showed different chemical shifts peaks for how to order on the sequence.[11] Then, when following the definition of chemical environment strictly, we need to consider the spin number of different environment in a chain as 15 spins, even without any packing difference from up and down (Figure 4-5a).

Then, we compared the build-up curve results of 15 spin system without and with the consideration of ^{13}C-labeling percentage (Figure 4-6a and b). In the both case, the EVs of lower elements from the diagonal became smaller than so far. This means the restrict spin numbering in the system is essential for the proper simulation of the model. The average and standard deviation for without and with the term of ^{13}C-labeling percentage are 0.44 and 0.37 for Figure 4-6a, and 0.33 and 0.29 for Figure 4-6b. The smaller EV results in Figure 4-6b (with ^{13}C-label term) proves that this ^{13}C-labelling consideration is better for the *B. mori* Cp-fraction system. Especially, relative long C-C distance correlations between Ser and Ala residues such as (1, 3), (1, 4), (3, 8) and (4, 8) showed better EVs by ^{13}C-labeling consideration. Therefore, this also implies that the ^{13}C-labeling consideration is important for simulation of longer distance build-up curve behavior.

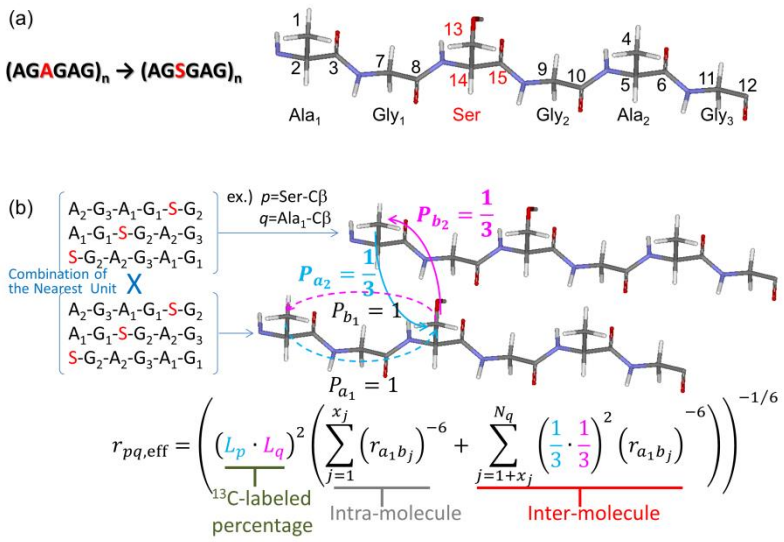

Figure 4-5. Schematic of the effective distance calculation, based on the probability of dipole-dipole couplings in *B. mori* Cp-fraction. (a) Spin numbering considered as 15 spin system. (b) Different combination and probability in the calculation of effective distance between intra- and inter-chain distances based on the $(AGSGAG)_n$ system with the consideration of ^{13}C-labeled percentage. The equation in (b) corresponds to Eq. [14].

4-3-3. Evaluation of the $(AG)_n$ structure with Silk II form proposed most recently.

Until the section 4-3-2, we have constructed the method for structural agreement of a candidate structure with the experimental data of *B. mori* Cp-fraction measured by 2D DARR sequence. Based on this method, we evaluate the integrity of the $(AG)_n$ model proposed most recently. In Figure 4-6a, the observed and calculated build-up curves showed reasonable agreement, although the contact between the two β-sheet structures was not considered in the simulation.[3] In all the 64 elements in the EV matrix of Figure 4-6b, as many as 41 elements showed the EV values within the permissible value 0.32 which was discussed above. The average of EV matrix elements in Figure 4-6b was 0.33 and comparable to the permissible value 0.32.

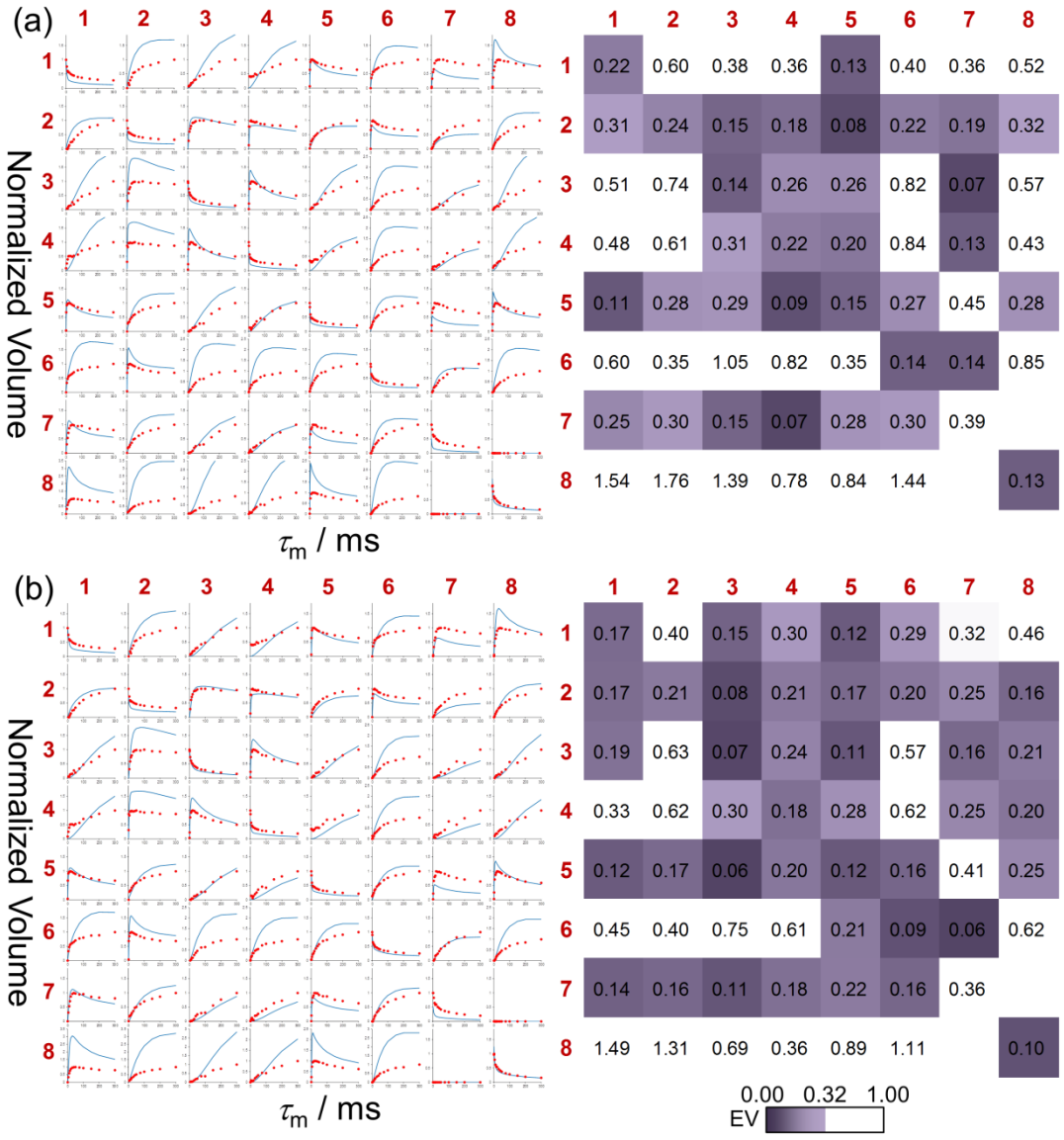

Figure 4-6. The build-up curves (left) and EV (right) matrices for DARR data of [U-^{13}C] *B. mori* Cp-fraction (considered as 8 spin system). The results of (a) just the alteration of effective distances for Ser by the same effective distances with Ala residue, (b) consideration of stoichiometry by insertion of Ser into the previous $(AG)_{15}$ model, and (c) consideration of ^{13}C-labeled percentage in each residue. The color scale in the EV matrix is the same with that of AP-β-Ala_4 in Figure 3-6c and d.

In EV matrix of the Cp-fraction data (Figure 4-6b), the correlation peaks with Gly-Cα showed relatively large discrepancy with the experimental data in both (upper and downer elements) dimensions. Figure 4-7 shows the

chemical shift dependency of Gly-Ca on Ramachandran plot[12] and the typical β-sheet conformation region is shown as orange region. The chemical shift dependency of the Gly-Cα for the β-region showed very small differences corresponding to a range of approximately 1.5 ppm. The Gly residue has no carbon side chain, so there is little restriction about the non-hydrogen bonding Gly-Cα position. Therefore, it is suggested that the methylene Gly-Cα in the main chain includes relatively large distribution of structure even if in a β-sheet structure. For this reason, the single-conformation models of the two β-sheets in Figure 4-1, which does not assume the structural distribution in Gly-Cα, may not agree well in the build-up curves of relatively long distances, for example 0.39 nm, which is the nearest one between Gly-Cα and Ala/Ser-Cβ for structure A : (6,3), (6,8), (3,6) and (8,6) in Figure 4-6b.

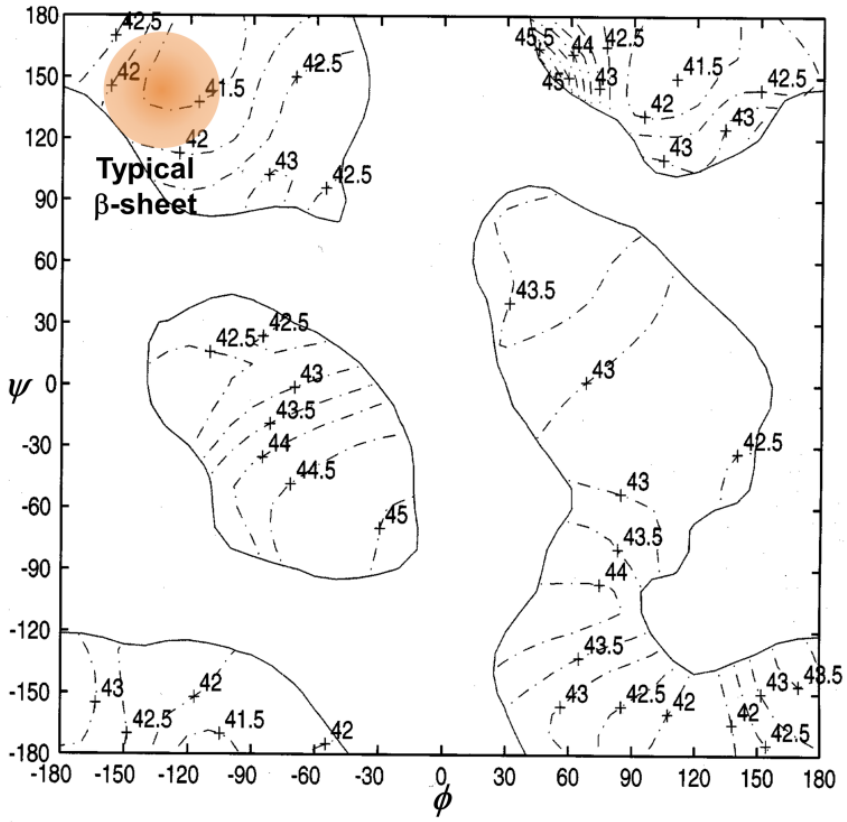

Figure 4-7. Chemical shift contour plot of Gly-Cα.[12]

In addition, elements of Ala-Cβ side-chain peaks especially from β-sheet A had relatively large EVs in Figure 4-6b. As we will discuss in the next chapter, DARR spectrum of the *B. mori* Cp-fraction with longer mixing time includes the correlation of an inter-domain contact between β-sheets A and B in its Ala-Cβ correlation region and it is suggested that the β-sheets form separated small domains less than 5 nm in size from the $T_{1\rho}^{H}$ measurement.[3] These results mean that consideration of the crystallite domain structure contacting between β-sheet A and B is necessary for the greater matching between the observed and the calculated build-up curves. Furthermore, the contribution of turn elements may have large effects on the build-up curve behavior supposed the previously proposed lamella model.[2]

4-4. Conclusion

The quantitative build-up curve analysis described in the chapter 4 has been applied and constructed for the structural evaluation of the [U-^{13}C] *B. mori* Cp-fraction with heterogeneous Silk II structure. We have made several improvements for the method inserting Ser into the previous $(AG)_{15}$ model, such as consideration of different effect from intra/inter-chain interaction, restricted spin environments occurring as a result of Ser insertion to the model and the ^{13}C labeled percentage in each residue. In the data after the all improvement in the calculation, the errors in many correlations appeared reasonable and within a comparable error range determined from the DARR results of AP-β-Ala_4. However, it may be necessary to consider the intrinsic effect of structural contact between the two β-sheet structures and further distorted β-turn structure to really reach the permissible range of EV. This method will be necessary for verifying the possible inter-domain/packing relation as will be discussed in the next section because the results described later are based on the same relaxation mechanism.

References

1. Asakura, T.; Ohata, T.; Kametani, S.; Okushita, K.; Yazawa, K.; Nishiyama, Y.; Nishimura, K.; Aoki, A.; Suzuki, F.; Kaji, H.; Ulrich, A. S.; Williamson, M. P. *Macromolecules* **2015**, 48, 28-36
2. Asakura, T.; Nakazawa, Y.; Ohnishi, E.; Moro, F. *Protein Sci.* **2005**, *14*, 2654-2657.
3. Okushita, K.; Asano, A.; Williamson, M. P.; Asakura, T. *Macromolecules* **2014**, *47*, 4308–4316.
4. Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, 344, 631－637.
5. Takegoshi, K.; Nakamura, S.; Terao, T. *J. Chem. Phys.* **2003**, *118*, 2325－2341.
6. Ohashi, R.; Takegoshi, K. *J. Chem. Phys.* **2006**, *125*, 214503.
7. Torchia, J. *J. Magn. Reson.* **1978**, *30*, 1969－1992.
8. Asakura, T.; Yao, J.; Yamane, T.; Umemura, K.; Ulrich, A. *J. Am. Chem. Soc*. **2002,** *124*, 8794－8795.
9. Asakura, T.; Yao, *J. Protein Sci*. **2002**, *11*, 2706-2713.
10. Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 42-49.
11. Asakura, T. *Makromol. Chem., Rapid Commun.* **1986**, *7*, 755-759.
12. Asakura, T.; Iwadate, M.; Demura, M.; Williamson, M. P. *J. Mol. Biol.* **1999**, *24*, 167－171.

Chapter 5

Insight into structure modeling of β-sheet domain in Silk II structure of *B. mori* silk fibroin

5-1. INTRODUCTION

The Silk II structure has heterogeneous structure as introduced in the Chapter 2 and the author has already assigned the domain of each Ser-Cβ peaks detected in ^{13}C CP/MAS or ^{13}C-^{13}C DARR spectrum at the short mixing time of 10 ms, based on the secondary structure and the fraction ratio in the Chapter 2. The peak assignments of *B. mori* Cp-fraction are summarized in Table 2-1. In this chapter, the author investigates the higher structure such as inter domain relation in Silk II structure of *B. mori* Cp-fraction by using some kinds of solid state NMR approaches.

5-2. EXPERIMENTAL SECTION

5-2-1. Crystalline fraction of [U- ^{13}C] *B. mori* silk fibroin

The preparation process for Cp-fraction is described in Chapter 2. The obtained Cp-fraction with Silk II form and H_2O were mixed with boltex mixer and stayed a day. Before the NMR measurement, the extra water was removed, and then only precipitate was used for NMR measurements.

5-2-2. ^{13}C-^{13}C DARR experiment.

The ^{13}C DARR measurements[1] were conducted on the same equipment as in chapter 2. The CP contact time and the recycle delay were 1.5 ms and 3 s, respectively. The wet Cp-fraction was inserted into 4mmϕ HR-MAS rotor purchased from Bruker.co.Ltd. The MAS rates were set at 11000 Hz for dry sample and 8750 Hz for wet sample. About, dry sample, the mixing time and number of scan were 400 ms and 72 scans. On the other hand, about wet sample, the mixing times were 10 ms and 400 ms, and 128 and 144 scans were collected at each time increment of the indirect dimension. The ^{1}H power for the recoupling was adjusted to the same as that of the MAS frequency in Hz.

5-2-3. T_1^H and $T_1\rho^H$ experiments.

The T_1^H and $T_1\rho^H$ experiments[2,3] were conducted on a Varian 400WB NMR system operating at 100 MHz for the ^{13}C nucleus. The experiments were performed at a MAS frequency of 9 kHz with a 3.2 mm ϕ rotor. The CP contact time and the recycle delay were 200 μs and 5 s, respectively. Both T_1^H and $T_{1\rho}^H$ measurements were performed at room temperature. A ^{1}H spin-locking frequency of 64 kHz was used for $T_{1\rho}^H$ measurement.

5-3. RESULTS AND DISCUSSION

5-3-1. Local intermolecular structure around Ser in each β-sheet structure

Fig. 5-1a shows the correlations between Ala-Cβ and Ser in the DARR spectrum at the long mixing time of 400 ms. Several correlations between Ser-Cβ around 64 ppm and Ala-Cβ at 19.6/21.7 ppm were detected. In standard anti-parallel β-sheet models of Silk II such as the Marsh model (Fig. 5-1b),[4] the intermolecular inter-carbon distance between Ser-Cα/Cβ and Ala-Cβ in the adjacent sheet is 0.48/0.41 nm while the intra-chain distance is much longer at 0.71/0.70 nm. This tendency is the same also in our proposed β-sheet models.[5] Moreover, in our solid-state NMR approaches such as ^{1}H DQ MAS combined with ^{2}H-labeling, it has been shown that the main inter-chain hydrogen bonding consists of Ala and Gly in $(AG)_{15}$ model peptide with Silk II form.[5] Therefore, also about Ser-Cα/Cβ Ala-Cβ distance of intra-sheet, it becomes further than those of Marsh model with A-A or G-G hydrogen bonding. In a DARR spectrum, shorter internuclear distances contribute more effectively to cross peak intensity. Therefore, one expects that the cross peak between Ser- Cα/Cβ and Ala-Cβ will be dominated by the inter-sheet rather than the intramolecular contribution. In Fig 5-1a, the Ser-Cα peaks at 54.9 and 54.0 ppm (corresponding to geometries A and B, respectively) show correlations with the Ala-Cβ peaks at 19.6 and 21.7 ppm, also corresponding to geometries A and B respectively. This observation supports the presence of separate A and B domains, because any model that has alternating or statistically random A and B strands would have A-B interactions as the shortest inter-sheet distance in the case of (Ser-Cα) – (Ala-Cβ) correlations. The Ser-Cβ peak at 64.2 ppm (geometry A) also has a strong cross peak with Ala-Cβ peak A, while the Ser-Cβ B peak has correlations that are too weak to interpret. However, correlations are also seen between geometries A and B,

for Ser-Cβ A to Ala-Cβ B and between Ser-Cα and Ala-Cβ. If the two geometries A and B formed large separate domains, inter-domain cross peaks could not be observed in the 2D DARR spectrum. Hence, it is concluded that the β-sheets A and B form small domains with extensive inter-domain contact.

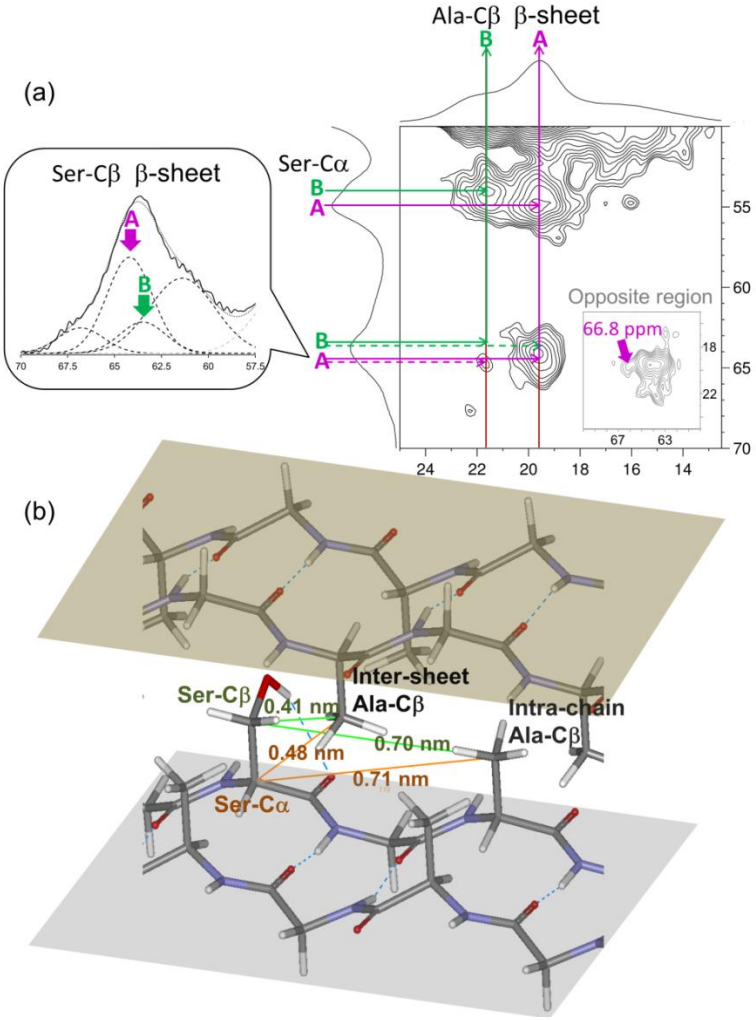

Figure 5-1. (a) (Ala-Cβ) − (Ser-Cα/Cβ) correlations in the 2D ^{13}C-^{13}C DARR spectrum of [U-^{13}C] Cp-fraction (Silk II) at a mixing time of 400 ms. Solid lines show the correlations within each β-sheet structure A and B, which have been assigned in Figs 2-3 and 2-4. The spectrum on the left is the ^{13}C CP/MAS spectrum of the same sample observed at 500 MHz. (b) Inter-carbon distances in the Marsh model, which is the generally accepted antiparallel Silk II model. Even in this model, the shortest inter-carbon distances between Ser-Cα/Cβ and Ala-Cβ are intermolecular rather than intramolecular.

In the 2D DARR spectrum with a 400 ms mixing time, a cross peak can be observed between Ala-Cβ A and Ala-Cβ B (Fig. 5-2), indicating that the two geometries A and B must be very close together. In a preliminary DARR measurement, a very weak correlation was detected from an inter-carbon distance of 0.65 nm in a > 99% ^{13}C triple-labeled model compound. A previous paper[6] has also estimated 0.70 nm to be the upper limit for DARR-detectable distances. Also in the chapter 3, the error in the quantitative build-up curve analysis converged until the distance limit of 0.8 nm. These results therefore imply that Ala-Cβ A and Ala-Cβ B must be within 0.7 nm or 0.8 nm, which is short compared to the inter-sheet increment of *B. mori* silk fibroin which has been shown to be 0.94 nm by the previous X-ray study.[4]

For the Ser-Cβ peak at 66.8 ppm (Fig. 2-4, component (i) in Chapter 2), a correlation with Ala-Cβ geometry A was found on the other side of the diagonal (Fig. 5-1a, "Opposite region"), demonstrating that the Ser-Cβ at 66.8 ppm is close to the β-sheet A structure. This peak tends to increase in intensity when the β-sheet structure is perturbed by the replacement of Ser to Tyr in the sequence.[7] I therefore assign the peak at 66.8 ppm to a distorted β-sheet A structure in the Cp fraction. This peak has a chemical shift 2.6 ppm higher than the normal domain A resonance, and the shift difference is larger for Ser Cβ than Ser Cα (Table 2-1 in Chapter 2). A change of this magnitude is not compatible with a change in χ_1 angle or backbone hydrogen bonding,[8,9] and we therefore suggest that the distortion may involve hydrogen bonding to the Ser OH.

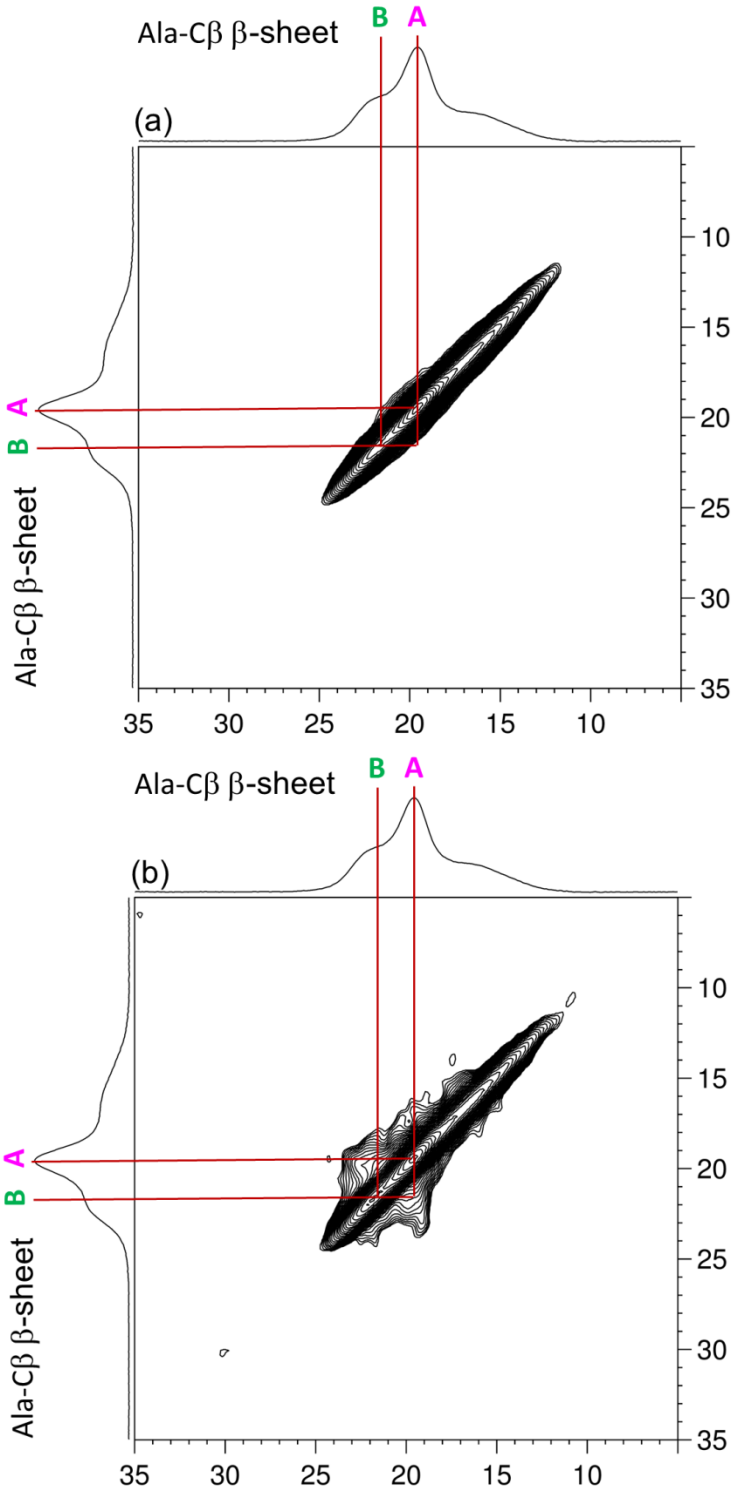

Figure 5-2. Correlation between the two β-sheet structures A and B in the 2D ^{13}C-^{13}C DARR spectra of [U-^{13}C] Cp-fraction (Silk II) at mixing times of 10 ms (a) and 400 ms (b).

5-3-2. Domain size of A and B β-sheet geometries.

To place limits on the domain size of the β-sheets A and B, we measured ^{1}H spin-lattice relaxation times in the rotating frame ($T_{1\rho}^{H}$) and in the laboratory frame (T_1^{H}).[2,3,10,11] Spin diffusion is an efficient mechanism for cross-relaxation in the solid over distances of several nm. If domains are within a few nm of each other, spin diffusion will cause spin-lattice relaxation times to be averaged, whereas if domains are larger than this, different relaxation times can be expected for different domains. Each observed relaxation curve was observed as a single-exponential decay although the highest field peak of 16.2 ppm had some distribution in the $T_{1\rho}^{H}$ relaxation curve. The obtained $T_{1\rho}^{H}$ and T_1^{H} values are tabulated in Table 5-1. All T_1^{H} values were 0.9 ± 0.1 s; there is no clear difference among the respective peaks. Similarly, the $T_{1\rho}^{H}$ of the [U-^{13}C] Cp-fraction was 19 ± 1 ms for the β-sheet peaks, whereas the random coil signal has a relatively short $T_{1\rho}^{H}$ of 12 ms. This result indicates that the random coil region is spatially distinct from the crystalline region, but suggests that the domain size of A and B geometries must be no more than 5 nm. About the random coil signal, bi-exponential fitting of the $T_{1\rho}^{H}$ relaxation decay with some distribution showed the $T_{1\rho}^{H}$ values of 16 ms and 3 ms. The longer $T_{1\rho}^{H}$ showed close value with those of β-sheets, so portion of random coil / distorted β-turn peak may be contained in the β-sheet domains. The ^{13}C CP/MAS spectrum of a short model peptide G(AG)$_3$ with ca. 2.1 nm length in a complete β-sheet shows only β-sheet peaks in the Ala-Cβ region.[12] However, the longer peptide (AG)$_9$ with ca. 6.0 nm length shows a distorted β-turn/random coil peak at the higher field side of the main Ala-Cβ peak.[12] These observations suggest that the estimated domain size of 5 nm is reasonable because the Cp fraction consists of both complete and distorted β-sheets. We have conducted

detailed calculations of cross peak intensities in DARR spectra for different domain sizes. Calculated intensities turn out to be relatively insensitive to the domain size, because relayed transfer of magnetization is efficient. Therefore an upper limit of 5 nm is the most precise limit imposed by the data.

Table 5-1. T_1^H relaxation times measured at 400 MHz and $T_{1\rho}{}^H$ relaxation times measured with ^{1}H spin-locking frequency of 64 kHz.

			T_1^H r.t.	$T_{1\rho}{}^H$ r.t.
Ala-Cβ	16.2 ppm	Random coil	0.89 s	12.0 ms
	19.6 ppm	β-sheet A	0.92 s	18.9 ms
	21.7 ppm	β-sheet B	0.93 s	18.1 ms
Ser-Cβ	61.4 ppm	Random coil	0.92 s	11.7 ms
	63.5 ppm	β-sheet B	0.95 s	18.5 ms
	64.2 ppm	β-sheet A	0.91 s	19.5 ms
	66.8 ppm	Distorted β-sheet	0.95 s	18.8 ms

5-3-3. Domain structure in a crystallite.

From the results in the previous section, the nano-scale morphology should be as summarized in Figure 5-3. The *B. mori* Cp-fraction, which consists of the crystalline region sequence in the whole *B. mori* silk fibroin and occupies about 55 %, has three kind of structural domains such as β-sheets A, B and distorted β-turn/random coil region as reflected in Ala-Cβ region of Figure 2-1. These three domains are mixed at the range of about 50 nm from the T_1^H results, but the random coil domain is spatially distinct from the crystalline within the range from the T_1^H and $T_{1\rho}^H$ results (Figure 5-3a, Table 5-1). On the other hand, within the crystalline domain, the two β-sheets A and B, which have different inter-molecular arrangements as reflected in the difference of their Ala-Cβ chemical shifts (Figure 2-1), makes separate small domains rather than alternative β-strands as suggested from the observed (Ala-Cβ)- (Ser-Cα) correlations in Figure 5-1a, and contact each other at the range as short as their correlation peaks can be observed in 2D DARR spectrum (Figure 5-2, Figure 5-3c). Moreover, the domain sizes of the two β-sheets are assumed as no more than 5 nm from their $T_{1\rho}^H$ results. Therefore, if the distribution of the two β-sheets in a crystallite is considered as random, domain structure may be displayed as Figure 5-3b. However, the detail of the domain structure in a crystallite has not been clarified yet because the two β-sheet existence can be detectable only with solid state ^{13}C NMR spectrum clearly.

Moreover, the 2D ^{13}C-^{13}C DARR build-up curve results of AP-Ala_4 with already known have shown that until at least 0.8 nm distances should be considered in the effective distance calculation procedure for the better agreement of the observed and calculated build-up curves. In the previous chapter, the matching of build-up curves in the *B. mori* Cp-fraction data, in which the calculation of build-up curves just assumed

the two separate β-sheet domains considering their ratio, was comparable to the permissible range but not enough. From the fact that correlation between β-sheet domain A and B is detectable, it may be necessary to be considered that the contact region between A and B is large enough for the interaction between the domains to affect the build-up curve behavior. Therefore, we need to discuss the domain structure in the β-sheet crystallite next.

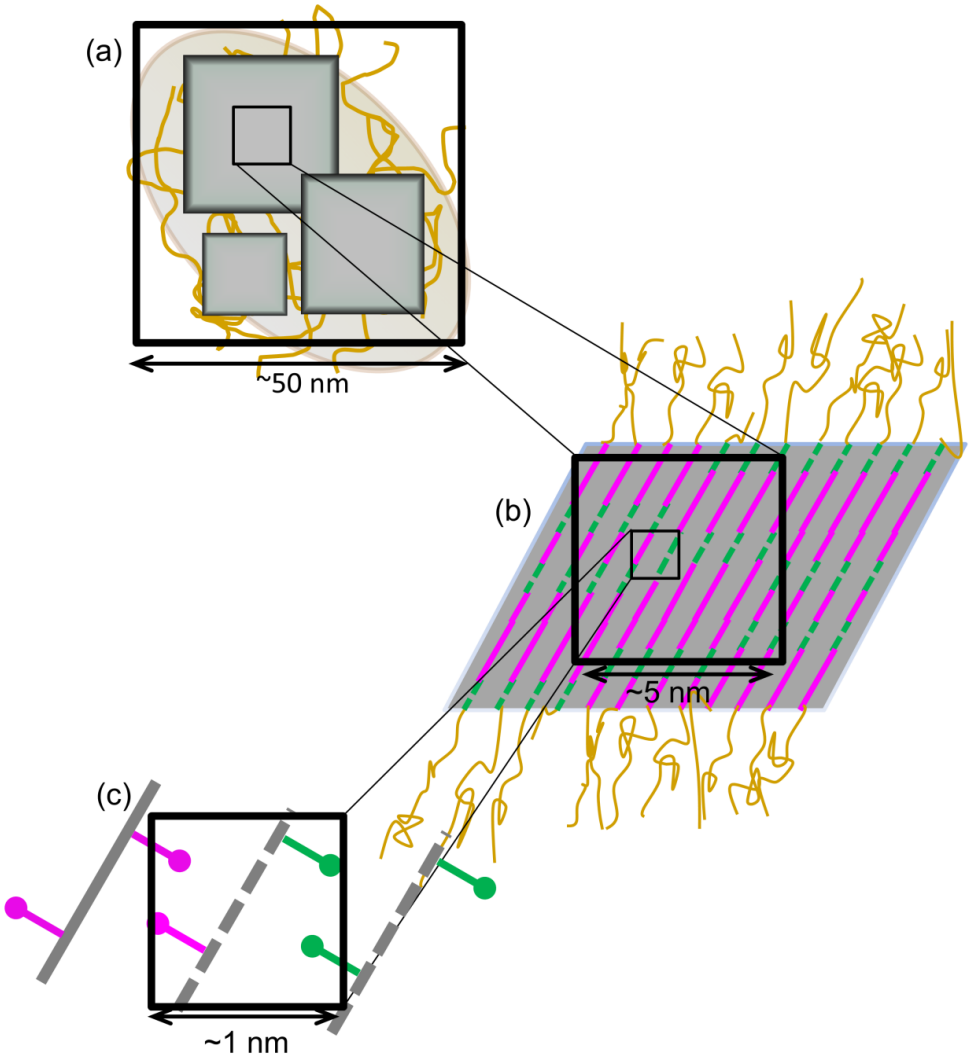

Figure 5-3. Nano-scale morphology of *B. mori* Cp-fraction proposed in this study, assuming the distribution of the β-sheet domains in a crystallite is random.

At first, in order to consider how the two β-sheet domains A and B distribute in a crystallite, we need to judge whether each domain is localized in the crystallite or not. In the *B. mori* fibroin case, it is known that β-sheet domains are formed by stripping off of the water surrounding silk fibroin chains.[13,14] Then, it becomes possible to detect the localization of the β-sheets in the surface of the crystallites by adding water in the observed system. It is took into granted that adding water to the *B. mori* Cp-fraction should influence the whole molecular mobility, but if there is any localization on the crystalline surface, specific change would be observed in the ^{13}C CP/MAS or 2D ^{13}C-^{13}C DARR spectrum of the *B. mori* Cp-fraction in water.

Table 5-2. The detected^{13}C chemical shift of Ala in the 2D ^{13}C-^{13}C DARR spectrum of Cp-fraction (Silk II) in water observed at the mixing time of 10 ms

Residue	Domain	Cβ / ppm	Cα / ppm	CO / ppm
Ala	Distorted β-turn	16.9	50.0	174.1 – 174.6
	β-sheet A*	19.1	48.8	171.8
		19.9		172.1
	β-sheet B*	21.7	48.5	171.9
		22.5		172.2

* These chemical shifts were detected at their Cα-CO correlations. They may be resolved because of restriction of ^{1}H spin diffusion even in a chain by adding water to the sample.

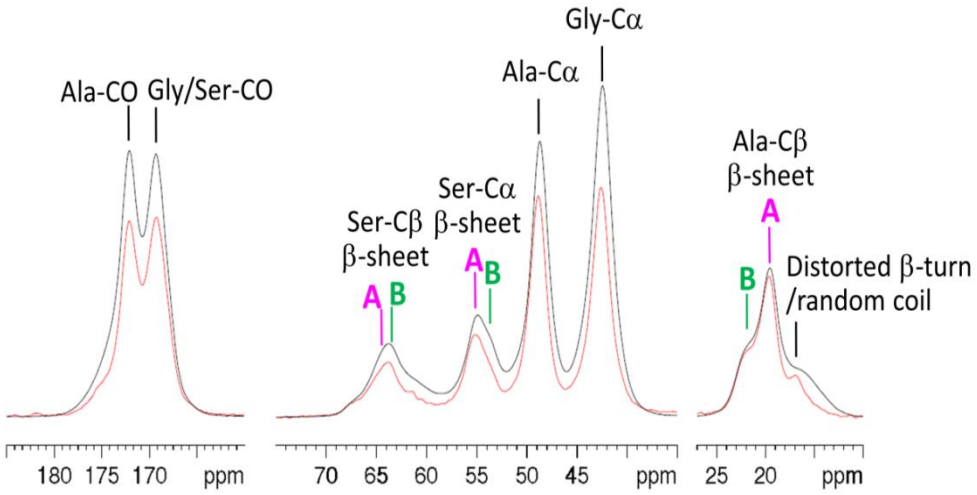

Figure 5-4. ^{13}C CP/MAS spectra of *B. mori* Cp-fraction with dry (black) and wet (red) state observed at MAS rete of 11 kHz and 8.75 kHz, respectively. These spectra were normalized at the peak top of Ala-Cβ β-sheet A peak. In the wet condition spectrum (red), peak intensities around disordered peak region of side chain carbons such as Ala-Cβ and Ser-Cβ decreased remarkably.

Figure 5-4 and Figure 5-5 show the comparisons of ^{13}C CP/MAS and 2D ^{13}C-^{13}C DARR spectra of dry (black) and wet (red) state, respectively. In the Ala-Cβ region of ^{13}C CP/MAS spectrum (Figure 5-4), the clear difference between dry and wet samples was observed in the distorted β-turn or random coil region around 16.2 ppm. In a ^{13}C CP/MAS NMR spectrum, signal enhancements in the mobile region are suppressed. In the Ala-Cβ region spectrum (Figure 5-4), the intensity of highest field peak at 16.2 ppm decreased and the line width also decreased. This intensity decreasing can be considered as that the chain mobility was influenced by the mobile water penetrating into the domain pore and increased much enough to suppress its CP enhancements. The previous paper also discussed that there is relatively large space in the disordered domain by correlating the pore size measured by included water in *B. mori* silk fibers with tendency of the Ala-Cβ peak intensity in ^{13}C CP/MAS NMR spectrum of them.[15]

On the other hand, the narrower line width at the 16.2 ppm means rather the decrease of the structural distribution than the increase of chain mobility of the components, because higher mobility components cannot be detect by CP. Focused on the chemical shift observed on the 2D DARR spectrum at the mixing time of 10 ms (Table 5-2), the Ala-Cβ peak at 16.9 ppm peak correlated with Ala-Cα peaks at 50.0 ppm. By comparing the observed chemical shifts with the chemical shift contour plots (Figure 5-6) similarly to chapter 2, the possible torsion angle region was similar to that of Silk I structure, type II β-turn structure previously reported.[16-18] Moreover, the Ala-CO chemical shift showed the 174.1 ppm, which is higher field chemical shift than that of Silk I structure[16-18] or typical α-helix structures.[19] Therefore, it was shown that the Silk II structure partially contains distorted β-turn structure also in wet state.

On the contrary, the β-sheet line shapes did not show remarkable difference between dry and wet spectrum in their line width or peak ratio in the 1D ^{13}C CP/MAS spectrum (Figure 5-4). In the Figure 5-4, the peak intensity was normalized by the peak top of β-sheet A peak. This means that the β-sheets are stable in certain extent and the mobility of the β-sheet chains is restricted enough to detect the peaks by CP even if in water. Then, it may seem to be impossible to detect the influence to the two β-sheet domains by adding water at a glance. However, about the observation between Ser-Cα and Ala-Cβ in 2D ^{13}C-^{13}C DARR spectrum of wet sample at the mixing time of 400ms, the difference between the β-sheet A and B was recognized (Figure 5-5). Namely, the correlation between Ser-Cα and Ala-Cβ was detected about A domain but not about B domain in Figure 5-5 red spectrum. This may be partially because of decreasing CP enhancement at the Ser-Cα peak of β-sheet B domain (Figure 5-5 left). Moreover, in a DARR spectrum, the stronger peak intensity means generally the shorter inter-carbon

distance. The (Ser-Cα)-(Ala-Cβ) correlation is contributed from the inter-carbon distance in inter-sheet relation as discussed in the previous section. Therefore, the disappear of the (Ser-Cα)-(Ala-Cβ) peak in the wet sample spectrum (Figure 5-5, red) is also considered as the evidence of widened inter-sheet distance around Ser in the β-sheet B domain by water. In the both interpretation of decreased CP enhancement and widened β-sheet increment around Ser, the result in Figure 5-5 means that there is localized the β-sheet B domain on the surface of the crystallite and the Ser residues in the β-sheet B domain are affected by water specifically.

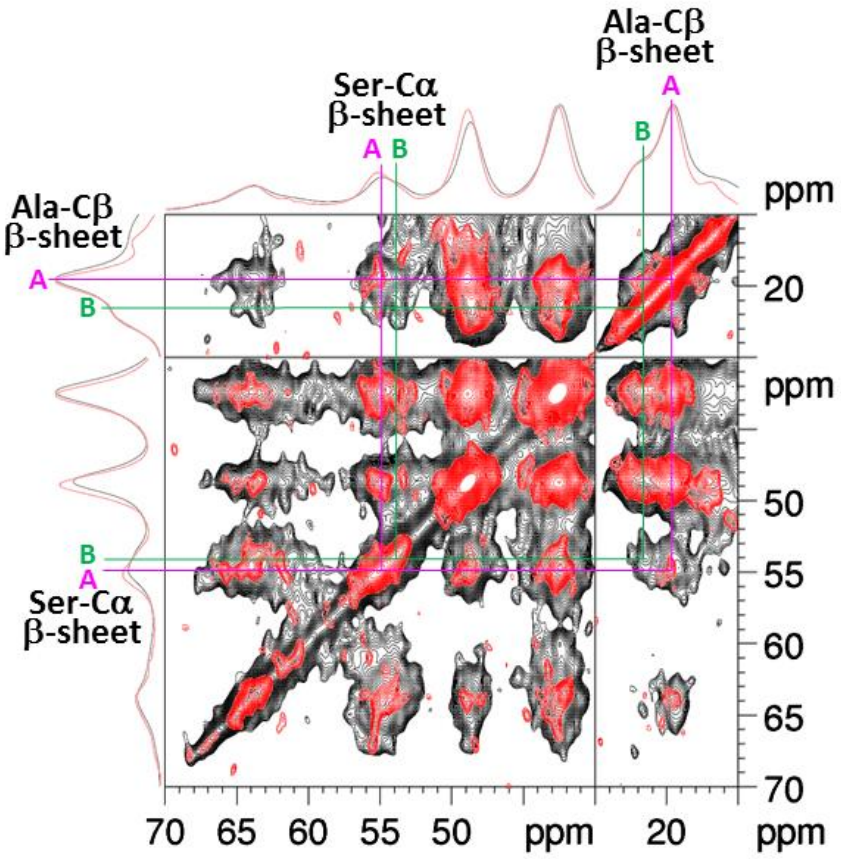

Figure 5-5. 2D ^{13}C-^{13}C DARR spectra of *B. mori* Cp-fraction with dry (black) and wet (red) state observed at MAS rete of 11 kHz and 8.75 kHz, respectively, and the mixing times were 400 ms in the both spectra. In the wet condition spectrum (red), correlation between Ser-Cα and Ala-Cβ in β-sheet B domain, which is intra-domain correlation, disappeared selectively.

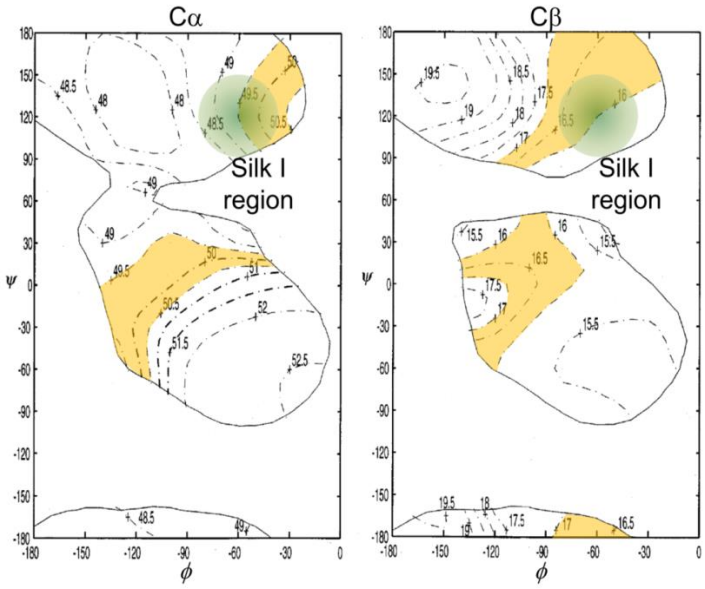

Figure 5-6. Chemical shift contour plot of (a) Ala-Cα and (b) Cβ in proteins published previously.[20] The (ϕ, ψ) angle regions colored in blue are typical Silk I, type II β-turn regions of (ϕ, ψ) = (-60°, 120°)[16-18]. The orange regions show the observed chemical shift region in this study.

In the previous NMR study, it has been shown that the fiber axis corresponds to the molecular chain axis in the *B. mori* fiber.[21] On the other hand, it has been shown that the crystals observed with Low Voltage Transmission Electron Microscopy (LVTEM) had a large distribution of sizes ranging from only a few nanometers to a few tens of nanometers in length (fiber axis), implying the correlation with the size distribution of crystallizable segments of amino acid in the *B. mori* sequence, and then broad and diffuse nature of reflections in WAX pattern also indicated the small size of crystallite.[22] The apparent sizes of β-sheet stacking and hydrogen bonded direction of β-sheet crystallite were ranging from 1.8 nm to 3.2 nm and from 2.6 nm to 5.2 among several papers, respectively,[22-29] which also imply the small size of crystallite constructs of the whole β-sheet domain.

As a possible structure as a unit crystallite in whole β-sheet domain, there are lamella structures reported previously.[34,31] The lamella structure proposed in our group[34] has β-turn at each eight residues and by considering Ala residue closest to the turn in a chain as in structure B geometry, the stoichiometry of A:B as 2:1[12,30] seems to be explained. This model has hydrogen bond from Ala-Gy pair as shown experimentally and possibly explains the selective disappear of β-sheet B peak in 2D DARR spectrum of wet Cp-fraction in this study because the Ala residue in structure B geometry is at the closest location to the flexible turn edge in the lamella model. In addition, the result from bi-exponential fitting of T_{1r}^{H} in this study may be explainable for the 16.2 ppm peak at Ala-Cβ region to partially include turn-like structure in the β-sheet domain with the size of around 5 to 10 nm at the maximum axis. However, there has been no investigation on the inter-sheet packing and alignment among the lamella so far and no support from chemical shift calculation by the latest methods used in the other model of Silk II.[5] About the inter-molecular arrangements, the candidate structure of whole β-sheet domain should satisfy the A/ B contact and non-alternating layer structure as shown experimentally in this study at least. The proposed lamella has been construct by molecular mechanics of single $(AG)_{15}$ chain,[34] so the model for the Cp-fraction with inter-molecular chain and averaged amino sequence including Ser $[GAGAGSAAG(SGAGAG)_8Y]$[32] may be possible to have other type of edge structure.

Furthermore, to obtain structural insight between β-sheet domain size and size of our proposed models,[5,34] the author considers the localization of β-sheet geometry most simply. The average size of the β-sheet crystallite of *B. mori* fiber has been measured as around 2.5 nm × 5 nm × 13 nm from X-ray study (Figure 5-7a).[22-29] In our proposed Silk II models for the β-sheets A and B,[5] the lattice parameters of a = 0.938 nm (inter-sheet/

inter- chain increment), b = 0.949 nm (inter- chain / inter-sheet increment), c =0.698 nm (fiber axis) with space group $P2_1$ are used on the basis of X-ray study[33], respectively. Based on these parameters, the number of included AGSGAG units along the chain axis should be about six, that of included chains in a β-sheet plane should be about ten, and then that of stacking β-sheet layers should be about six. From the β-sheet ratio previously reported, the ratio of included β-sheet A and B chains should be about 2:1.[12,30] Hence, the possible domain structures in a crystallite when considering the β-sheet B domain localization on the surface become as shown in Figures 5-7 b to d. Among these domain structure models, it is expected that the model in Figure 5-7 d is difficult for surrounding water to access into its β-sheet domain B because of penetration by disordered chains on the crystallite side although it is possible in the case of small lamella[34] aggregation.

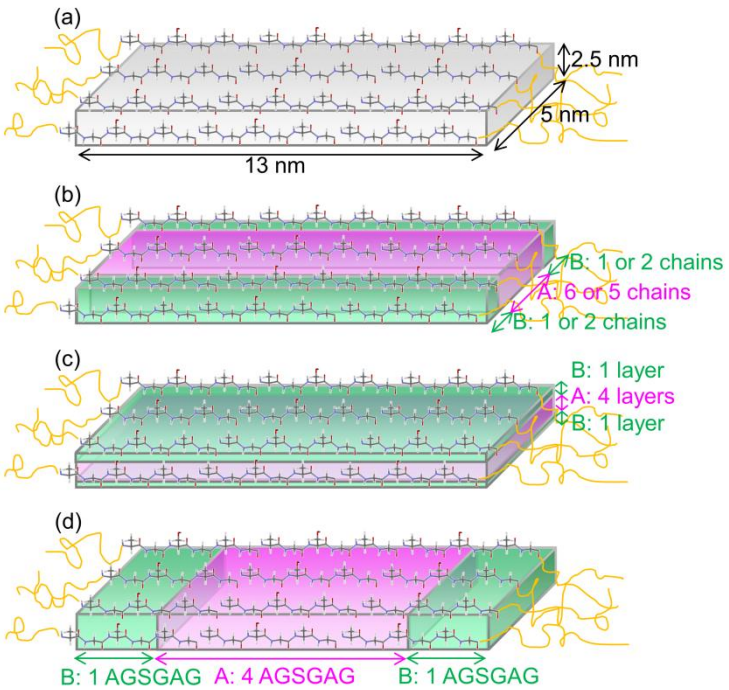

Figure 5-7. The simplest morphology in a β-sheet crystallite in the *B. mori* Cp-fraction supported in this study.

4-4. CONCLUSIONS

The author has uncovered the heterogeneous Silk II domain structure in *B. mori* Cp-fraction by using ^{13}C-^{13}C DARR measurement at the longer mixing time of 400 ms and ^{1}H spin-lattice relaxation measurements. As a result of the 2D DARR measurement, it was suggested the close proximity of the two β-sheet domains within a few nm. This result was also supported from the ^{1}H spin-lattice relaxation measurements. Also from Chapter 2 consequently, it is necessary to revise the standard model of *Bombyx mori* silk fibroin from the uniform antiparallel β-sheet proposed more than half a century ago. Moreover, by combining the water addition to the Cp-fraction and 2D DARR measurement at the longer mixing time, it was shown that the β-sheet B domain localizes on the surface of the crystallite. Thus, through this chapter, it was shown that the combination of solid state NMR experiments can successfully clarify the inter-domain structure in heterogeneous *B. mori* Cp-fraction with Silk II form.

REFERENCES

1. Takegoshi, K.; Nakamura, S.; Terao, T. *Chem. Phys. Lett.* **2001**, *344*, 631-637.

2. Stejskal, E. O., Schaefer, J., Sefcik, M. D.; McKay, R. A. *Macromolecules* **1981**, *14*, 275-279.

3. Asano, A.; Takegoshi, K. Polymer Blends and Miscibility. *Solid State NMR of Polymers* (ed. by Ando, I. & Asakura, T.); Elsevier Science B. V.: Amsterdam, 1988; pp. 351-414.

4. Marsh, R; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* 1955, *16*, 1-34.

5. Asakura, T.; Ohata, T.; Kametani, S.; Okushita, K.; Yazawa, K.; Nishiyama, Y.; Nishimura, K.; Aoki, A.; Suzuki, F.; Kaji, H.: Ulrich, A. S.; Williamson, M. P. *Macromolecules* **2015**, *48*, 28-36.

6. Egawa, A.; Fujiwara, T.; Mizuguchi, T.; Kakitani, Y.; Koyama, Y.; Akutsu, H. *Proc. Natl. Acad. Sci.* **2007**,*104*, 790-795.

7. Asakura, T.; Ohgo, K.; Ishida, T.; Taddei, P.; Monti, P.; Kishore, R. *Biomacromolecules* **2005**, *6*, 468-474.

8. Iwadate, M.; Asakura, T.; Williamson, M. P. *J. Biomol. NMR* 1999, **13**, 199-211.

9. Shen, Y.; Bax, A. *J. Biomol. NMR* 2013, **56**, 227-241.

10. VanderHart, D. L.; McFadden, G. B. *Solid State Nucl. Magn. Reson.* **1996**, *7*, 45-66.

11. Demco, D. E.; Johansson, A.; Tegenfeldt, J. *Solid State Nucl. Magn. Reson.* **1995**, *4*, 13-38.

12. Asakura, T.; Yao, J. *Protein Sci.* **2002**, *11*, 2706-2713.

13. Chen, X.; Shao, Z.; Knight, D. P.; Vollrath, F. *PROTEINS: Structure, Function, and Bioinformatics* **2007**, *68*, 223–231.

14. Yamane, T; Umemura, K.: Nakazawa, Y.; Asakura, T. *Macromolecules* **2003**, *36*, 6766-6772.

15. Zhu. Z.; Gong, D.; Liu, L.; Wang, Y. *Anal. Bioanal. Chem.* **2014**, *406*, 2709–2718.

16. Asakura, T.; Ohgo, K.; Komatsu, K.; Kanenari, M.; Okuyama, K. *Macromolecules* **2005**, *38*, 7397-7403.

17. Asakura, T.; Ashida, J.; Yamane, T.; Kameda, T.; Nakazawa, Y.; Ohgo, K.; Komatsu, K. *J. Mol. Biol.* **2001**, *306*, 7397-7403.

18. Asakura, T.; Suzuki, Y.; Yazawa, K.; Aoki, A.; Nishiyama, Y.; Nishimura, K.; Suzuki, F.; Kaji, H., *Macromolecules* **2013**, *46*, 8046-8050.

19. Holland, G. P.; Janelle E. Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L. Chem. Commun. **2008**, 5568–5570.

20. Asakura, T.; Iwadate, M.; Demura, M.; Williamson, M. P. *Int. J. Biol. Macromol.* **1999**, *4*, 167-171.

21. Demura, M.; Minami, M.; Asakura, T.; Cross, T. A. *J. Am. Chem. Soc.* **1998**, *120*, 1300-1308.

22. Drummy, L. F.; Farmer, B. L.; Naik R. R. *Soft Matter* **2007**, *3*, 877–882.

23. Pe´rez-Rigueiro, J.; Elices, M.; Plaza, G. R.; Guinea, G. V. *Macromolecules* **2007**, *40*, 5360-5365.

24. Fraser, R. B. C.; MacRae, T. P. *Conformation in Fibrous Proteins*; Academic Press: New York, 1973; pp 293-343.

25. Martel, A.; Burghammer, M.; Davies, R. J.; Riekel, C. *Biomacromolecules* **2007**, *8*, 3548-3556.

26. Fraser, B.; MacRae, T. P. *Conformation of Fibrous Proteins and Related Synthetic Polypeptides*; Academic Press: **1973**; pp293 - 343.

27. Xu, G; Gong, L.; Yang, Z.; Liu, X. Y. *Soft Matter* **2014**, *10*, 2116-2123.

28. Plaza, G. R.; Corsini, P.; Marsano, E.; Pérez-Rigueiro, J.; Elices, M.; Riekel, C.; Vendrely, C.; Guinea, G. V. *J. Polymer Sci. Part B: Polymer Physics* **2012**, *50*, 455–465.

29. Liang, K.; Gong, Y.; Fu, J.; Yan, S.; Tan, Y.; Du, R.; Xing, X.; Mo, G.; Chen, Z.; Cai, Q.; Sun, D.; Wu, Z. *Internat. J. Biol. Macromol.* **2013**, *57*, 99– 104.

30. Asakura, T.; Yao, J.; Yamane, T.; Umemura, K.; Ulrich, A. *J. Am. Chem. Soc.* **2002**, *124*, 8794-8795.

31. Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 42-49.

32. Strydom, D. J.; Haylett, T.; Stead, R. H. *Biochem. Biopyhs. Res. Comm.* **1977**, *79*, 932-938.

33. Takahashi, Y.; Gehoh, M.;Yuzuriha, K. *Int. J. Biol. Macromolecules* **1999**, *24,* 127-138.

34. Asakura, T.; Nakazawa, Y.; Ohnishi, E.; Moro, F. *Protein Sci.* **2005**, *14*, 2654–2657.

Chapter 6
Conclusion

In this thesis, the author has made up and applied the analytical methods for elucidating the nanometer scale morphology in the heterogeneous *B. mori* Cp-fraction. Advanced points in this thesis are as follows;

1) Through the application of build-up curve approach to the real sample of *B. mori* Cp-fraction, the author has shown that the relaxation behavior is very sensitive to the environment over the local geometry experimentally and essential for understanding detailed morphology which cannot clarify with other analytical instruments. Although there are still some problems to be solved in the application to the candidate models of *B. mori* Cp-fraction with Silk II form, this thesis is advanced in the point of practical aspect of detailed relaxation treatment including ^{13}C-labeing percentage, intra/intermolecular consideration, considerable distance limit and so on.
2) By combining the solid state NMR methods based on homonuclear dipolar coupling or spin diffusion, structures with variety of scales were observed for each domain marker peak. Namely, from 0.15 nm in intra-residue structure up to several ten nm in the whole Cp-fraction. The combination of 2D ^{13}C-^{13}C correlation spectrum used mainly in protein studies and relaxation time measurements used mainly in polymer study is advanced because their combination made it possible to describe the detailed inter-domain structure at atomic level.

In chapter 2 "Assignment of each ^{13}C NMR peak for each domain in Silk II structure of *Bombyx mori* (*B. mori*) silk fibroin Cp-fraction", the author assigned heterogeneous Ser peaks in solid state ^{13}C CP/MAS NMR spectra of the *B. mori* Cp-fraction through 2D ^{13}C-^{13}C DARR measurement at a short mixing time. The Ser peaks contained three β-sheet peaks which consisted of distorted β-sheet A, β-sheet domains A and B, and random coil peak. Moreover, based on T_1^C of each Ser-Cβ peak, possible formation of hydrogen bond at Ser-Cβ-OH and the larger inter-sheet packing in β-sheet domain B were suggested. The obtained chemical shifts will be effective for the validation of possible local structure around Ser by combining with chemical shifts calculation. Then, the information on each Ser-Cβ mobility will also narrower the candidate local structure diversity.

In chapter 3 "Quantitative method to evaluate the structural agreement of tertiary packing structure in a β-sheet domain –Investigation by antiparallel β-sheet alanine tetramer (AP-Ala$_4$) with already-known structure", the author investigated quantitativities of the conventional PDSD and DARR experiments based on spin-diffusion. The quantitativities of effective inter-carbon distances were better in DARR data. Moreover, through the method construction process, the distance limit, which should be considered in the build-up calculation and more cases of DARR application in our experimental condition, has been estimated experimentally. Then, the author has shown the necessity of diagonal T_1^C elements for the quantitative prediction of peak intensities at the longer mixing times which reflect on

longer distance contribution. The fact clarified experimentally in this chapter will be helpful for distance measurement by 2D homonuclear correlation spectrum of other iterative samples.

In chapter 4 "An application of ^{13}C-^{13}C DARR build-up curve approach for the candidate models with chemical shift agreements and method development for the *B. mori* Cp-fraction with Silk II form", the DARR sequence selected in the previous chapter was applied to evaluate the structural agreements of our most previous proposed model, which contains two kinds of β-sheet structures, with the experimental data of *B. mori* Cp-fraction (Silk II). From this analysis, comparable agreement with the permissible range which had been set from the model peptide with already-known structure, but for the complete agreement, consideration of β-sheet distribution in a crystallite and/or relation with turn location, which was discussed in the chapter 5, might be needed. This shows paradoxically that the DARR build-up curve analysis reflects on the both of local packing geometry and more wide-range of structural state. Therefore, this method would be proper for verifying structural agreement including whole structural state.

In chapter 5 "Insight into structure modeling of β-sheet domain in Silk II structure of *B. mori* silk fibroin", contact between the two β-sheet structures A and B within a few nm ranges were clearly shown by the 2D ^{13}C-^{13}C DARR measurement at the longer mixing time of 400 ms. From the same spectrum

and $T_{1\rho}^{H}$ measurement, it was suggested that the two β-sheets consist of separated small domains, between which spin diffusion would be averaged within about 5 nm diameter, in a crystallite or β-sheet region in the *B. mori* Cp-fraction. This kind of knowledge on inter-domain structure in subnano-scale could not be obtained without the detailed characterization of each domain and residue via solid state NMR approach described in chapter 2. Moreover, from the T_1^{H} measurement, the spin-diffusion averaging in the tree domains such as β-sheets A, B and distorted β-turn/random coil domains was estimated. The hierarchical information of the Cp-fraction, crystalline region sample, will be a key to understand structural relation with the other non-crystalline sequence region such as $(GAGAGX)_n$ repetitive region (X = Y or V).

Acknowledgement

The author gratefully acknowledges many advices and comments through this thesis of Prof. T. Asakura at Tokyo University and Agriculture and Technology.

The author also wishes to thank

Prof. A. Asano (National Defense Academy) for helpful discussion, suggestion and many machine times,

Prof. M. P. Williamson (University of Sheffield) for helpful and impressive discussion and suggestion,

Prof. G. S. Boutis (Brooklyn College) for many suggestions and helps for preparing the MATLAB codes for build-up curve analysis.

Then, the author also wants to extend thanks to many contributions of the followings.

・Prof. T. Fujiwara and Dr. A. Egawa (Osaka University): Providing the original fortran program for their relaxation matrix analysis and helpful suggestion.

・Dr. Aoki (TUAT): Valuable discussion on Silk II models, kindly supports and suggestions about fortran program.

・All the members in Asakura Lab of TUAT and Asano Lab of NDA when my belonging.