Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins

Mitsuo Umetsu[a,b], Kouhei Tsumoto[a,*], Kumar Ashish[a], Shigeki Nitta[a], Yoshikazu Tanaka[a], Tadafumi Adschiri[b], Izumi Kumagai[a,**]

[a] *Department of Biomolecular Engineering, Graduate School of Engineering, Tohoku University, Aobayama 07, Aoba-ku, Sendai 980-8579, Japan*
[b] *Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, 2-1-1, Katahira, Aoba-ku, Sendai 980-8577, Japan*

Received 3 November 2003; revised 26 November 2003; accepted 26 November 2003

First published online 15 December 2003

Edited by Thomas L. James

Abstract **Several recombinant proteins in inclusion bodies expressed in *Escherichia coli* have been measured by Fourier transform infrared and solid-state nuclear magnetic resonance spectra to provide the secondary structural characteristics of the proteins from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 (hyperthermophilic proteins) in inclusion bodies. The β-strand-rich single chain Fv fragment (scFv) and α-helix-rich interleukin (IL)-4 lost part of the native-like secondary structure in inclusion bodies, while the inclusion bodies composed of the hyperthermophilic proteins of which the native form is α-helix rich, are predominated by α-helix structure. Further, the secondary structure of the recombinant proteins solubilized from inclusion bodies by detergent or denaturant was observed by circular dichroism (CD) spectra. The solubilization induced the denaturation of the secondary structure for scFv and IL-4, whereas the solubilized hyperthermophilic proteins have retained the α-helix structure with the CD properties resembling those of their native forms. This indicates that the hyperthermophilic proteins form native-like secondary structure in inclusion bodies. Refolding of several hyperthermophilic proteins from in vivo aggregated form without complete denaturation could be accomplished by solubilization with lower concentration (e.g. 2 M) of guanidine hydrochloride and removal of the denaturant via stepwise dialysis. This supports the existence of proteins with native-like structure in inclusion bodies and suggests that non-native association between the secondary structure elements leads to in vivo aggregation. We propose a refolding procedure on the basis of the structural properties of the aggregated archaeon proteins.**

Key words: Archaeon protein; Inclusion body; Refolding; Solid-state nuclear magnetic resonance; Guanidine hydrochloride

1. Introduction

Bacterial expression of various foreign genes yields recombinant proteins in the periplasm or cytoplasm and also in insoluble particles called inclusion bodies, which are often formed in the bacterial host [1,2]. The in vivo aggregation is recognized as a serious obstacle to utilizing the gene resources supplied from genomic analysis. Completion of decoding enormous genome sequences gave rise to the concept of structural genomics [3–5], which can supply various structures of proteins only by expressing proteins as a soluble form. Inclusion bodies are close-packed aggregates composed mainly of the recombinant proteins and have no known physiological function [6]. The aggregation of recombinant proteins may be due partly to overexpression, which increases the local concentration of the protein, but several other factors also lead to inclusion body formation; for example, non-native hydrophobic interactions among partially folded intermediates, reducing conditions in the cytoplasm, and intermolecular disulfide linking [7–10]. The in vivo aggregation of several proteins can be suppressed by controlling culture conditions [11] and by the coexpression of chaperone molecules [12,13]. However, there is still little comprehensive understanding of the mechanism of inclusion body formation.

Despite the drawbacks of utilizing inclusion body formation, the refolding in vitro has been one of the primary topics in protein engineering because no general system for inhibiting the formation of inclusion bodies has been constructed [14]. The recombinant protein solubilized by a high concentrated strong denaturant is refolded from a completely denatured form by removal of the denaturants, however, it is frequently difficult to attain highly efficient yield due to the competing off-pathway self-association of the folding intermediates in the refolding in vitro [15,16]. Recent studies on unfolding and renaturation of protein supply several techniques for inhibiting the aggregation; introduction of L-arginine [15,17,18] or chaperone [19,20] in the system of removing denaturants, and the refolding on column [21], but the efficient yield in the refolding systems has been strongly dependent upon the intrinsic properties of the target proteins.

The structures of in vivo aggregated proteins have been analyzed by Fourier transform infrared (FT-IR) spectroscopy, X-ray diffraction, electron microscopy (EM), and solid-state nuclear magnetic resonance (NMR). The spectroscopic techniques are in wide use for β-amyloid, in which cross β-structures lead to formation of an insoluble filament (FT-IR [22],

*Corresponding author. Fax: (81)-22-217 7276.
**Corresponding author. Fax: (81)-22-217 6164.
E-mail addresses: tsumoto@mail.cc.tohoku.ac.jp (K. Tsumoto), kmiz@mail.cc.tohoku.ac.jp (I. Kumagai).

Abbreviations: BSA, bovine serum albumin; CD, circular dichroism; CP, cross polarization; CTAC, cetyltrimethylammonium chloride; EM, electron microscopy; FT-IR, Fourier transform infrared; GdnHCl, guanidine hydrochloride; IL, interleukin; IMAC, immobilized metal affinity chromatography; MAS, magic angle spinning; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; PDSD, proton-driven spin diffusion; scFv, single chain Fv antibody fragment; 2D, two-dimensional

0014-5793/03/$30.00
doi:10.1016/S0014-5793(03)01441-8

X-ray [23], EM [24], liquid NMR [25], solid-state NMR [26]); however, fewer approaches have been applied to inclusion bodies because of their heterogeneous structures. For the heterogeneous aggregates, FT-IR spectroscopy is often used to analyze the secondary structure, using the bands corresponding to amide bond stretching [7,9,22]. The FT-IR spectrum of interleukin (IL)-1β (which has only β-sheets in its native form) in inclusion bodies shows a structure rich in β-structures, and deconvolution analysis suggests that the composition of secondary structure is similar in the inclusion body and the native form [9]. In contrast, IL-2 and β-lactamase, both of which have more than 50% α-helix in the native secondary structure, have non-native β-structure in inclusion bodies as well as decreased α-helix content [7,27]. In general, the renaturation of recombinant proteins from inclusion bodies is performed after the solubilization by strong denaturant agents or detergents which are considered to cause complete denaturation [16,28]. Complete denatured proteins, however, should be subjected to complicated refolding procedures for blockading many off-pathways to aggregation. The structural analysis for inclusion bodies would be of importance, particularly when the refolding of the proteins from inclusion bodies is attempted on the basis of structural properties of the aggregated proteins.

In this study, we analyzed the structures of several recombinant proteins in inclusion bodies expressed in *Escherichia coli*: the β-sheet-rich single chain Fv (scFv) fragment from a mouse anti-lysozyme monoclonal antibody (HyHEL-10), IL-4 with more than 50% α-helix structure, and several α-helix-rich recombinant proteins from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. The hyperthermophilic archaeons are one of the major target species for structural genomics [4,5], but the hyperthermophilic proteins are frequently expressed as inclusion bodies in *E. coli* [3,4], due to the difference in the optimal temperature for growth between host cell and hyperthermophilic archaeon. Secondary structure in inclusion bodies was measured by FT-IR and solid-state NMR spectroscopy, and the solubilized form from inclusion bodies was estimated from circular dichroism (CD) spectra. On the basis of results obtained, we show that the hyperthermophilic protein forms rigid native-like secondary structure in inclusion bodies. In addition, some of the hyperthermophilic proteins in inclusion bodies could be solubilized without the degeneration of the secondary structure formed in inclusion bodies by a 2 M guanidine hydrochloride (GdnHCl) solution, and further, the solubilized proteins could be refolded by removal of the denaturant via the stepwise dialysis method. We would discuss the potential for the direct refolding from inclusion bodies, not through complete denatured form.

2. Materials and methods

2.1. Materials

Lyophilized powdered bovine serum albumin (BSA) was purchased from Wako Pure Chemical (Osaka, Japan). A plasmid containing DNA coding for scFv of a mouse anti-lysozyme monoclonal antibody, HyHEL-10 (VL-linker-VH-$(His)_6$), was constructed by inserting the *Nco*I–*Xba*I fragment of pUTN3 [29] into the *Nco*I–*Xba*I-digested plasmid pUT7 [30]. The gene for IL-4 was completely synthesized from several oligonucleotides by an overlap extension polymerase chain reaction (PCR) method as described previously [31], and the full-length DNA was inserted into a T7 promoter-based pUT vector after digestion with *Nco*I and *Sac*II. The genes coding for PH0628, PH0979, and PH1830, which have sequences homologous to methionine aminopeptidase, hydrogenase, and glyceraldehyde-3-phosphate dehydrogenase, respectively, in the hyperthermophilic archaeon *P. horikoshii* OT3 [32], were amplified by PCR and then inserted into pET15b after digestion with *Nde*I and *Bam*HI. These hyperthermophilic proteins were selected because of their high expression as inclusion bodies (more than 50% of the total expressed protein). *E. coli* BL21(DE3) cells transformed with each expression vector [33] were incubated in 2× YT medium at 28°C, and expression of each protein was induced by adding 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). The harvested cells were centrifuged and suspended in 50 mM Tris–HCl (pH 8.0) buffer containing 200 mM NaCl. The suspension was sonicated and then centrifuged at 5800×*g* for 30 min at 4°C. The supernatant was used to prepare soluble recombinant proteins and the pellet to prepare inclusion bodies.

To prepare native soluble hyperthermophilic proteins coded by PH0628, PH0979, and PH1830, the supernatants were subjected to immobilized metal affinity chromatography (IMAC) on His-bind resin (Novagen) and gel filtration column chromatography on Superdex 75 (Amersham Biosciences, Tokyo, Japan).

To purify inclusion bodies, the suspension of the pellet in a 4% Triton X-100 solution (50 mM Tris–HCl, 20 mM NaCl, pH 8.0) and centrifugation (5800×*g* for 30 min at 4°C) were repeated several times, and then the inclusion bodies were washed in water to remove the detergent. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis indicated that the inclusion bodies consisted mainly of expressed protein (more than 90% of all the proteins stained by Coomassie brilliant blue R-250). For the FT-IR experiments, the wet inclusion body pellet was dried under vacuum and then mixed with about 5% (w/w) KBr.

The soluble forms of scFv and IL-4 were obtained from inclusion bodies by refolding as described previously [34]. That is, the denatured, reduced proteins in inclusion bodies were solubilized with a 6 M GdnHCl solution containing 2-mercaptoethanol. The proteins were then refolded by gradually removing GdnHCl with stepwise dialysis from 6 to 3, 2, 1, 0.5 M, and finally 0 M. L-Arginine (final concentration, 0.4 M) was added at the 1 and 0.5 M GdnHCl stages. Redox reagent (oxidized form of glutathione for scFv, a mixture of oxidized and reduced form of glutathione for IL-4) was also introduced at the same stages (final concentration, 50-fold molar excess relative to the protein concentration).

2.2. Refolding of hyperthermophilic protein from inclusion bodies

Each inclusion body was solubilized by a 2 M GdnHCl solution (50 mM Tris–HCl, 20 mM NaCl, pH 8.0), and then the solutions were purified by IMAC and gel filtration column chromatography on Superdex 75. The solubilized hyperthermophilic proteins were refolded by removing GdnHCl with dialysis from 2 M GdnHCl to 0 M; The PH0979-coded protein solution was directly dialyzed from 2 to 0 M, while the PH0628- and PH1830-coded proteins were refolded by stepwise dialysis through 1 and 0.5 M GdnHCl solutions containing 0.4 M L-arginine.

2.3. FT-IR and solid-state NMR experiments

Diffuse reflectance FT-IR spectra were obtained with an FTS-165 (Bio-Rad Lab., USA), scanned 3000 times at a resolution of 2 cm^{-1}. Solid-state NMR spectra were collected at room temperature with a Bruker DRX-400 spectrometer equipped with a 4 mm double resonance probe. The magic angle spinning (MAS) technique was applied at frequencies of 8 kHz. Two-pulse phase modulation was used for proton decoupling, with pulses of 8 μs alternating between two phases separated by an angle of 15° [35]. A cross polarization (CP) sequence with a linear ramp from 100 to 70% was applied for the heteronuclear ^{1}H-^{13}C polarization transfer at high MAS rates during a CP time of 2–4 ms [36].

The two-dimensional (2D) homonuclear ^{13}C-^{13}C dipolar correlation spectra were measured by the proton-driven spin diffusion (PDSD) technique [37] and were collected with time-proportional phase incrementation. The 2D homonuclear ^{13}C spectra were collected for 256 t_1 data points and 1024 t_2 data points, and 256 scans were collected for each t_1 point using a repetition time of 1 s.

For the 2D NMR spectra, the proteins in inclusion bodies were fully or selectively ^{13}C labeled. For the fully ^{13}C-labeled proteins, transformed *E. coli* cells were grown at 28°C in a minimal medium based on M9 medium containing 1 g l^{-1} of fully ^{13}C-labeled D-glucose (Isotec, USA). The selectively labeled proteins were produced based

on the method of Muchmore et al. [38] and amino acid metabolism [39]: 2 ml of an amino acid blend containing 1.5% (w/v) each of asparagine, aspartic acid, methionine, threonine, lysine, isoleucine, glutamic acid, glutamine, proline, and arginine was added to 1 l of cells growing in M9 medium containing 1 g of fully ^{13}C-labeled D-glucose. In the modified M9 medium, the carbons in glycine, alanine, valine, leucine, serine, cysteine, tryptophan, phenylalanine, tyrosine, and histidine should become fully labeled.

2.4. CD spectra

CD spectra were measured with an AVIV CD spectrometer (Proteiron Co., USA): path length, 1.0 mm; resolution, 0.2 nm; average time, 4 s. Recombinant proteins in inclusion bodies were solubilized in 50 mM Tris–HCl (pH 8.0) buffer containing 0.5% cetyltrimethylammonium chloride (CTAC) and 200 mM NaCl. The solubilized proteins were purified by IMAC followed by gel filtration column chromatography.

3. Results

3.1. FT-IR and solid-state NMR spectra

Fig. 1 shows the FT-IR spectra in the amide I region for lyophilized BSA powder (a) and for inclusion bodies containing scFv (b), IL-4 (c), and PH0979-coded protein (d). The IR spectrum for BSA showed a maximum at 1659 cm^{-1} (Fig. 1a), but that for scFv did not show an explicit maximum, due to an increased intensity around 1630 cm^{-1} (Fig. 1b). According to previous reports [9,40,41], the amide I band at 1650–1660 cm^{-1} can be assigned to an α-helix or an irregular loop structure, and the bands around 1620–1630 cm^{-1} are assigned to β-structures. The spectrum for inclusion bodies of IL-4 showed a maximum at 1687 cm^{-1} rather than at 1659 cm^{-1}, and amide I bands corresponding to β-structures appeared around 1630 cm^{-1} (Fig. 1c). Considering the native structure of IL-4 as determined by NMR analysis (55% α-helix and 3% β-structure) [42], the relative amount of β-structure seems higher in the in vivo aggregated IL-4. In contrast, the PH0979-coded hyperthermophilic protein (Fig. 1d), whose native structure from the soluble fraction is 42% α-helix as determined by CD spectroscopy (see Fig. 5d), showed a predominant amide I band at 1660 cm^{-1} (similar to BSA), without any bands clearly attributable to β-structure. Similar results were obtained for the other hyperthermophilic proteins (PH0628 and PH1830, data not shown).

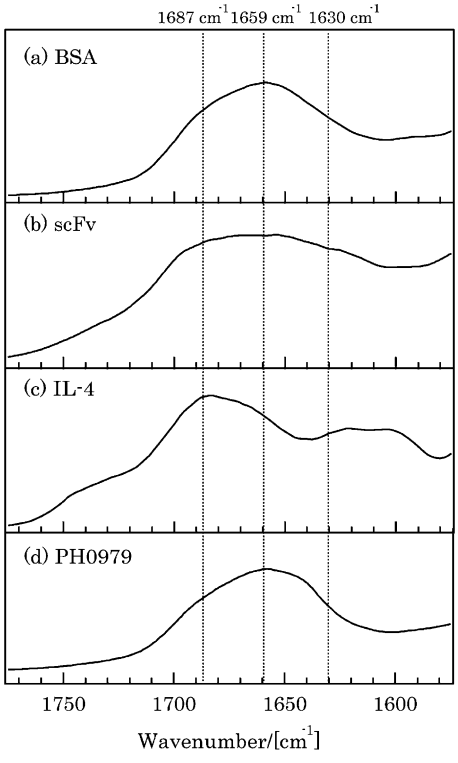

Fig. 1. FT-IR spectra in the amide I region for (a) lyophilized powdered BSA, and for inclusion bodies containing (b) scFv, (c) IL-4, and (d) PH0979-coded protein.

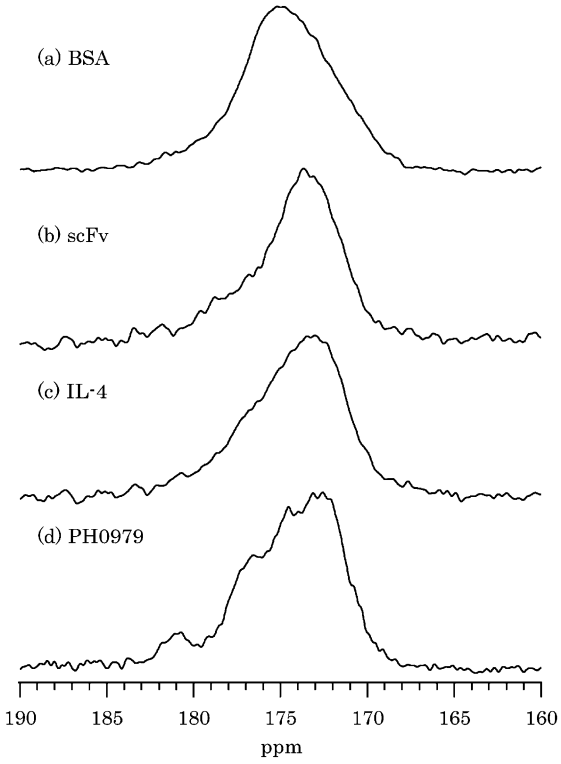

Fig. 2. Solid-state ^{13}C CP/MAS NMR spectra in the region of carbonyl carbon for (a) lyophilized powdered BSA, and for inclusion bodies containing (b) scFv, (c) IL-4, and (d) PH0979-coded protein.

Solid-state ^{13}C CP/MAS NMR spectra were measured in the region of carbonyl carbon resonance, to recognize α-helix structures (Fig. 2), because it is difficult to discern solely from IR spectra whether the band at 1650–1660 cm^{-1} arises from α-helix or irregular/loop structures. The NMR spectrum of BSA had a maximum at 175.5 ppm, which is assigned to α-helix structure (Fig. 2a) [43,44], implying that the amide I band at 1659 cm^{-1} in the FT-IR spectrum can also be attributed to α-helix structure. For scFv, resonance was enhanced at 173 ppm, but the intensity at 175–176 ppm decreased relative to the BSA spectra (Fig. 2b), indicating that the FT-IR spectrum for scFv has little of the component corresponding to α-helix structure. The NMR spectra for the other proteins (IL-4 and PH0979, Fig. 2c and d; PH0628 and PH1830, data not shown) showed a clear shoulder at 176 ppm; therefore, the corresponding FT-IR spectra indicate the presence of α-helix structure. The inclusion bodies used for the FT-IR measurement were dried and mixed with KBr, which might induce the structural change of protein. On the other hand, we measured the solid-state NMR spectra for the dried and hydrate inclusion bodies (dried sample: data not shown, hydrate sample: Fig. 2) and recognized the identical spectra between them for all the proteins. This indicates that the α-helix struc-

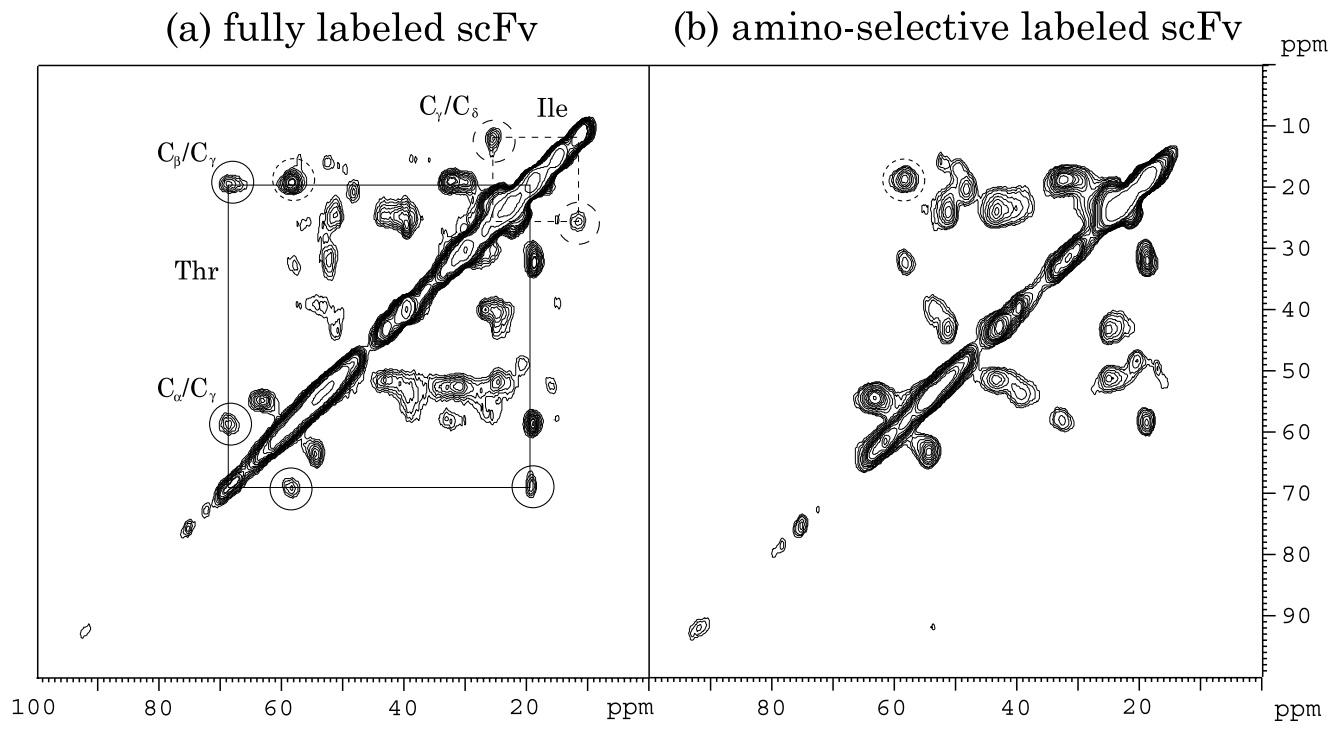

Fig. 3. Contour plot of the 2D CP/MAS PDSD ^{13}C-^{13}C dipolar correlation spectra for (a) uniformly ^{13}C-labeled scFv and (b) amino-selective labeled scFv in inclusion bodies at a magnetic field of 9.5 T. The MAS rotation frequency and the mixing time were 8 kHz and 15 ms, respectively. The solid and dashed circles indicate correlations between C_α, C_β, and C_γ in threonine and between C_γ and C_δ in isoleucine, respectively. The dotted circle represents the correlation derived from valine (see 2D NMR in Section 3).

ture observed in the inclusion bodies is not changed by the treatment by dehydration.

We also measured the 2D homonuclear ^{13}C-^{13}C dipolar correlation NMR spectrum, which also detects α-helix structure. Fig. 3 shows a contour plot of the 2D ^{13}C-^{13}C spectra of fully ^{13}C-labeled scFv (a) and amino-selective labeled scFv (b). Both spectra are much less refined than 2D NMR results reported for optimized proteins in solid-state NMR experiments [45], indicating much heterogeneity in the inclusion body. The contour plot for the fully labeled scFv shows clear correlation signals between C_α, C_β, and C_γ for threonine (solid circles in Fig. 3a) and a cross peak from C_γ/C_δ in isoleucine (dashed circles in Fig. 3a). However, the 2D spectrum for the selectively labeled scFv completely showed no correlation from threonine or isoleucine (Fig. 3b). The correlation between the resonances at 58 and 19 ppm (dotted circle in

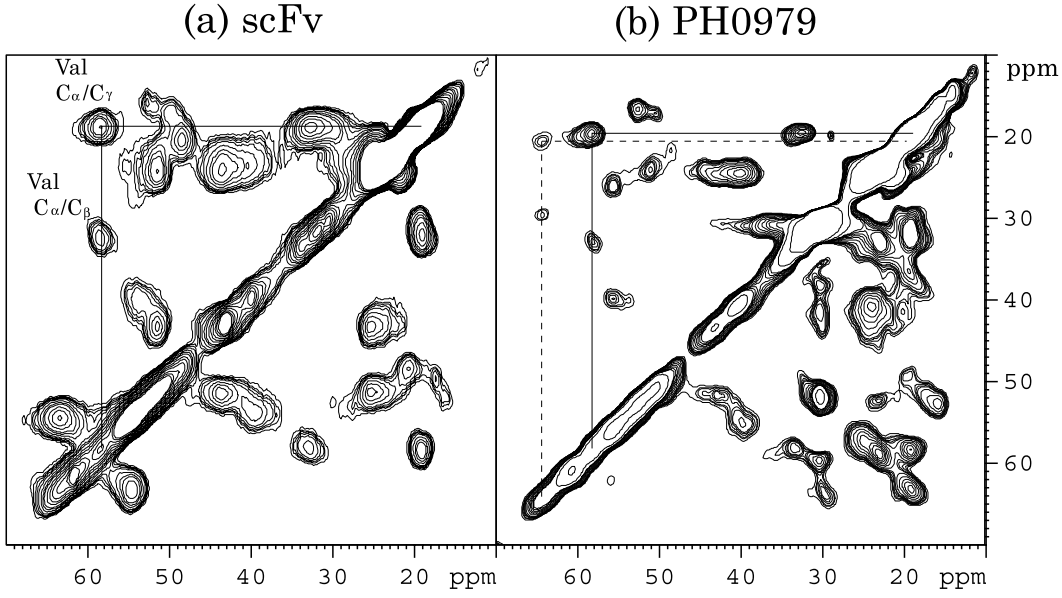

Fig. 4. Contour plot of 2D CP/MAS PDSD ^{13}C-^{13}C dipolar correlation spectra for amino-selective ^{13}C-labeled (a) scFv and (b) PH0979-coded protein in inclusion bodies at a magnetic field of 9.5 T. The MAS rotation frequency and the mixing time were 8 kHz and 15 ms, respectively. The solid and dotted lines indicate the two types of correlation between C_α, C_β, and C_γ in valine.

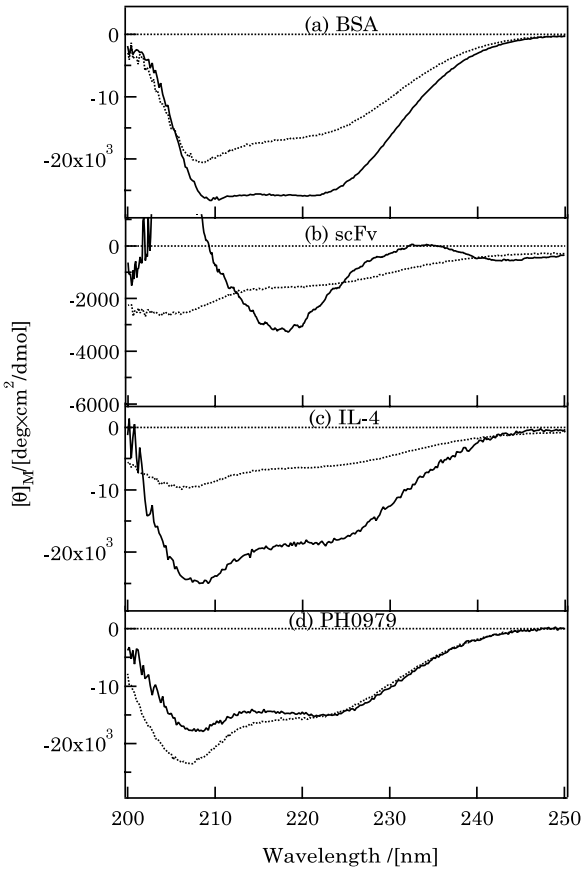

Fig. 5. CD spectra of (a) BSA, (b) scFv, (c) IL-4, and (d) PH0979-coded protein in native and refolded forms (solid line), and solubilized form by CTAC detergent (dotted line). Lyophilized BSA was directly dissolved in 50 mM Tris–HCl (pH 8.0), 200 mM NaCl, 1 mM ethylenediamine tetraacetic acid (EDTA) (solid line) or in Tris–HCl buffer with 0.5% CTAC detergent (dotted line). For scFv and IL-4, the protein that was refolded from inclusion bodies by stepwise dialysis (see Section 2) was measured in the Tris–HCl buffer lacking CTAC detergent (solid lines in b, c), and the native form of PH0979-coded protein was obtained from a soluble cytoplasmic fraction (solid line in d). The solubilized forms (scFv, IL-4, and PH0979-coded protein) were extracted from inclusion bodies by 0.5% CTAC solution and were measured in Tris–HCl buffer with 0.5% CTAC detergent after purification by IMAC and gel filtration (dotted lines in b–d).

Fig. 3a) may be assigned to a transfer of coherence between C_α and C_γ in threonine or valine but the observation of the cross peak in the contour plot for the selectively labeled scFv (dotted circle in Fig. 3b) proves that the correlation is derived from valine.

Fig. 4 shows the ^{13}C-^{13}C dipolar correlation spectrum for the amino-selective labeled scFv (a) and the PH0979-coded hyperthermophilic protein (b). In the spectrum for selectively labeled scFv (Fig. 4a), C_α in valine is resonated at 58 ppm and makes a clear cross peak with C_β and C_γ of valine at 32 and 19 ppm, respectively (solid line). A similar correlation network in valine also occurred in the spectrum for the PH0979-coded protein (solid line in Fig. 4b), but another correlation corresponding to valine was also observed (dotted line in Fig. 4b); in this new network, C_α in valine at 64 ppm makes a correlation with C_β and C_γ of valine at 30 and 20 ppm. Considering that the chemical shift for C_α of valine in an α-helix structure is significantly downfield shifted from the standard value (60–61 ppm) to ~65 ppm [44], the chemical shift at 64 ppm observed only for PH0979-coded protein can be assigned to the C_α of valine in an α-helix structure. These 2D NMR results suggest no α-helix structure in the in vivo aggregated scFv and support that PH0979-coded protein forms an α-helix structure in inclusion bodies.

3.2. CD spectra of native proteins and those solubilized from inclusion bodies by CTAC

Fig. 5 shows the CD spectra for each protein in its native form (solid line) and after solubilization by CTAC (dotted line) from a lyophilized powder (BSA) or from inclusion bodies (the recombinant proteins). BSA showed a typical spectrum for an α-helix structure, with strong intensity at 222 nm (solid line, Fig. 5a); however, the intensity at 222 nm for BSA in 0.5% CTAC was lower (dotted line in Fig. 5a). Thus, CTAC detergent denatured BSA to some extent. For scFv and IL-4, the spectra of the solubilized forms differed from those of the native structures (Fig. 5b,c). The native structure of scFv showed a clear trough at 218 nm, which can be attributed to β-structure (solid line in Fig. 5b), but a trough at 205 nm appears with the decrease of the CD intensity at 218 nm for the solubilized form (dotted line in Fig. 5b). The solubilized form of IL-4 had a much weaker trough at 222 nm than the native form, and the trough at 208 nm for the native form was shifted to 205 nm for the solubilized form (Fig. 5c). The PH0979-coded hyperthermophilic protein had the same intensity at 222 nm as its native structure, although the trough at 208 nm was slightly enhanced (Fig. 5d). Conservation of the 222 nm CD intensity was observed also for the other hyperthermophilic proteins (PH0628 and PH1830, data not shown), suggesting that the solubilized forms of the hyperthermophilic proteins have a rigid native-like α-helix structure. The slight enhancement at 208 nm was also observed for the PH0628-coded protein (data not shown). This might imply the existence of partially non-native structures in addition to the native-like α-helix structure.

3.3. Refolding of hyperthermophilic protein from inclusion bodies without complete denaturation

We also measured the CD spectra for the recombinant proteins solubilized from inclusion bodies by a 2 M GdnHCl solution (Fig. 6a). The solubilized scFv and IL-4 showed no characteristic CD signals derived from secondary structure (data not shown), while the CD spectrum for the solubilized PH0979-coded protein had a large trough at 222 nm (dotted line in Fig. 6a). Although the CD spectrum in the region of 200–212 nm was not measured in the 2 M GdnHCl solution due to high concentration of GdnHCl, the comparable CD intensity of the solubilized form to that of native form at 222 nm indicates that the α-helix structure formed in inclusion bodies has higher tolerance for chaotropic agent than the secondary structure of scFv and IL-4 (Fig. 6a). This suggests that the PH0979-coded proteins in inclusion bodies were solubilized by 2 M GdnHCl solution with the rigid native-like α-helix structure conserved. For the attempt to refold the PH0979-coded protein from the solubilized form without complete denaturation, GdnHCl was removed from the solubilizing solution by direct dialysis to a GdnHCl-free solution. The simple removal of GdnHCl attained the refolding yield of 75% for the PH0979-coded protein, and 67% of the refolded

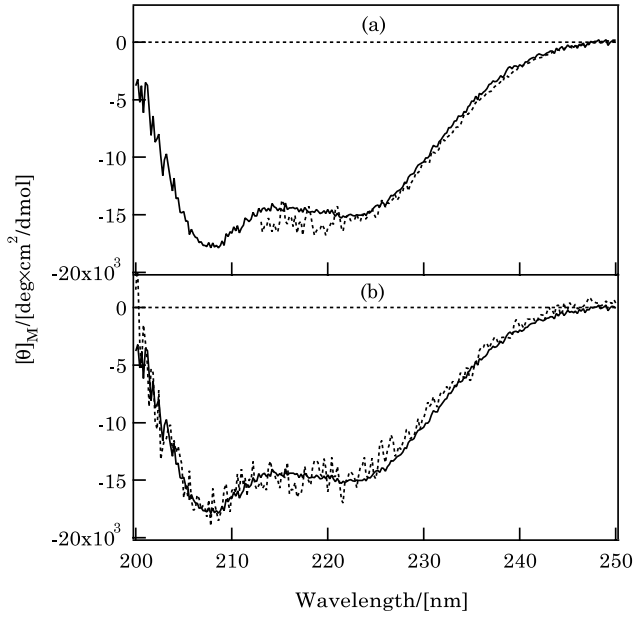

Fig. 6. a: CD spectra of PH0979-coded protein in native form (solid line) and solubilized form by 2 M GdnHCl (dotted line). The solubilized PH0979-coded protein was extracted from inclusion bodies by 2 M GdnHCl and was measured in Tris–HCl buffer with 2 M GdnHCl after purification by IMAC and gel filtration. b: CD spectra of PH0979-coded protein in native form (solid line) and refolded form (dotted line). The refolded form was measured after the solubilization from inclusion bodies by 2 M GdnHCl and removal of GdnHCl by dialysis to GdnHCl-free Tris–HCl buffer. The concentrations of native and refolded forms were 4 and 0.8 μM, respectively.

proteins was thermostable (Table 1). The refolded PH0979-coded protein showed the identical CD profile to that of native protein (Fig. 6b), which has the same troughs at 222 and 208 nm as those of native form. This result supports the suggestion that PH0979-coded protein forms the native-like secondary structure in inclusion bodies.

The PH0628- and PH1830-coded proteins were also solubilized by 2 M GdnHCl solution, and native-like CD signals were recognized for both the solubilized proteins (data not shown); however, most of proteins were aggregated by the direct dialysis from 2 M GdnHCl to 0 M. We also attempted to refold the solubilized PH0628- and PH1830-coded proteins by the stepwise dialysis trough 1 and 0.5 M GdnHCl solutions containing 0.4 M L-arginine. The modest refolding procedure raised the refolding yield up to 62% (PH0628) and 43% (PH1830), and about half of the refolded proteins were thermostable (Table 1). As the case of PH0979-coded protein, CD spectra of these two proteins refolded were identical to those of native proteins (data not shown).

4. Discussion

In this study, we measured FT-IR and solid-state NMR spectra for several recombinant proteins in inclusion bodies and found secondary structural differences between proteins. All our recombinant proteins had substantial secondary structures in inclusion bodies, but the degree of similarity to native structures seemed to vary; scFv and IL-4 in inclusion bodies seemed to have non-native structure, while the secondary structure of α-helix-rich hyperthermophilic proteins used in this study was native-like even in the dried inclusion bodies (Fig. 1). The solid-state NMR spectra for the dried and hydrated inclusion bodies were identical (Fig. 2), indicating that the secondary structures of the proteins in inclusion bodies are too rigid to be affected by the dehydration treatment which might have had the potential to change the secondary structure. Fink et al. [7] reported that aggregation often leads to an increase of the secondary structure indicated by the amide I components around 1630 cm^{-1} in IR spectra, suggesting that the intermolecular interaction leading to aggregation involves β-sheet-like interactions. The formation of non-native β-structures in inclusion bodies has been observed for IL-2 and β-lactamase, which have 57 and 32% α-helix structure, respectively, in their native forms [7,27]. The same result seems to be observed for IL-4, but the hyperthermophilic recombinant proteins in inclusion bodies have α-helix-rich structure with little amide I components around 1630 cm^{-1} observed (Fig. 1). This indicates that the hyperthermophilic proteins are aggregated in vivo after the formation of the secondary structure and that the secondary structure, especially α-helix, is not changed by the aggregation. Interestingly, the α-helix structures were so rigid that a detergent, CTAC, and low concentrated GdnHCl denaturant cannot denature the secondary structure of the proteins (Figs. 5d, 6a), and further the far-ultraviolet (UV) CD properties were comparable to those of native form. The solubilization of native-like secondary structure from inclusion bodies supports the observation of the native-like structure in inclusion bodies and the suggestion that the structure can not be changed by in vivo aggregation.

The formation of inclusion bodies has been reported for P22 tailspike endorhamnosidase [46,47] and human IL-1β [48]. The aggregation behavior of P22 tailspike protein indicates that a thermolabile intermediate formed during folding in vivo leads to the formation of inclusion bodies [8,46,47], and the formation of inclusion bodies of IL-1β has also been explained in terms of the properties of a folding intermediate rather than those of the native state [48–50]. For the hyperthermophilic proteins, therefore, the formation of inclusion bodies might be originated from non-native association between secondary structure elements.

In general, the recombinant proteins in inclusion bodies are solubilized by a high concentrated strong denaturant, such as 6 M GdnHCl or 9 M urea, and refolded by removal of the denaturants via dilution or dialysis [16,28]. The refolding from

Table 1
The refolding ratio of the hyperthermophilic protein from inclusion bodies without complete denaturation

	Refolding ratio (%)[a]	After heat treatment (%)[b]
PH0628	62[c]	42
PH0979	75[d]	67
PH1830	43[c]	69

[a]Each recombinant protein was solubilized from inclusion bodies by 2 M GdnHCl and purified by IMAC and gel filtration, and then the protein–GdnHCl solutions were dialyzed to GdnHCl-free solution.
[b]Each refolded protein was heated at 70°C for 30 min. The ratios were calculated by dividing the amount of the soluble proteins after the heat treatment by the total amount of the refolded protein.
[c]PH0628- and PH1830-coded proteins were refolded by the stepwise dialysis to a simple Tris–HCl solution through the 1 and 0.5 M GdnHCl solutions containing 0.4 M L-arginine.
[d]PH0970-coded protein was refolded by direct dialysis to a simple Tris–HCl solution.

completely denatured form is frequently difficult to attain highly efficient yield due to the competing off-pathway self-association of the folding intermediates in the refolding in vitro [15]. Especially, the denaturation of hyperthermophilic protein is apt to be irreversible perhaps because the relaxation kinetics of folding–unfolding is unusual slow [51,52]; indeed, to our knowledge, completely denatured hyperthermophilic proteins are not often refolded by frequently used methods (M. Umetsu and K. Tsumoto, unpublished data). The fact that the PH0979-coded proteins were refolded from the inclusion bodies with 2 M GdnHCl supports the existence of native-like secondary structure in inclusion bodies, but also proposes a new refolding procedure from the inclusion bodies which can be a cross-cut from aggregated form in inclusion bodies of archaeon proteins to a native form. Recently, we solubilized green fluorescent protein (GFP) from the inclusion bodies expressed in *E. coli* using a low concentrated GdnHCl and L-arginine; the solubilized GFP protein is fluorescent even in the solubilizing solutions, and the GFP protein after the removal of the denaturants attains identical activity to native form [53]. In this study, we could refold the hyperthermophilic proteins from in vivo aggregates without complete denaturation, but the refolding yields were different between the proteins used; half of the solubilized PH0979-coded protein was refolded to a thermostable form by a direct dialysis to a GdnHCl-free solution (Table 1), while most of the PH0628- and PH1830-coded proteins were aggregated by the dialysis (data not shown). Gradual removal of the denaturant via stepwise dialysis increased the refolding yield (Table 1), indicating the importance of promoting the appropriate association between the secondary structure elements for refolding to native form. According to our gel filtration of the native forms, the PH0979-coded protein is in monomeric form while the PH0628- and PH1830-coded proteins are in dimeric and tetrameric forms, respectively. The refolded and thermostable proteins also showed the identical oligomerization (data not shown). The refolding yield might be influenced by the variation of the association between each polypeptide chain.

In conclusion, the recombinant thermostable proteins from hyperthermophilic archaeon in inclusion bodies had native-like rigid secondary structures. The refolding of the hyperthermophilic proteins from the aggregated forms in inclusion bodies could be accomplished by solubilization with lower concentration of GdnHCl and removal of the denaturant, indicating the importance of promoting appropriate association between the secondary structure elements.

Acknowledgements: We express our heartfelt thanks to Prof. Huub J.M. de Groot of Leiden University for advice on 2D NMR measurements and are grateful to Prof. Tsunenori Nozawa and Dr. Zheng-Yu Wang of Tohoku University for discussions of the NMR results. The work was supported by a Grant-in-Aid for the Industrial Technology Research Grant Program in 2002 from the New Energy and Industrial Technology Development Organization (NEDO) of Japan and a Scientific Research Grant from the Ministry of Education, Science, Sports, and Culture of Japan. This research was also partially supported by the Ministry of Education, Science, Sports and Culture, a Grant-in-Aid for the COE project, Giant Molecules and Complex Systems, 2002.

References

[1] Williams, D.C., van Frank, R.M., Muth, W.L. and Burnett, J.P. (1982) Science 215, 687–689.
[2] Georgiou, G., Telford, J.N., Shuler, M.L. and Wilson, D.B. (1986) Appl. Environ. Microbiol. 52, 1157–1161.
[3] Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. (2000) Prog. Biophys. Mol. Biol. 73, 339–345.
[4] Gaasterland, T. (1999) Curr. Opin. Microbiol. 2, 542–547.
[5] Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai1, E.F., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000) Nat. Struct. Biol. 7, 903–909.
[6] Taylor, G., Hoare, M., Gray, D.R. and Marston, F.A.O. (1986) Biotechnology 4, 553–557.
[7] Fink, A.L. (1998) Fold. Des. 3, R9–R23.
[8] Haase-Pettingell, C.A. and King, J. (1988) J. Biol. Chem. 263, 4977–4983.
[9] Oberg, K., Chrunyk, B.A., Wetzel, R. and Fink, A.L. (1994) Biochemistry 33, 2628–2634.
[10] Nieba, L., Honegger, A., Krebber, C. and Plückthun, A. (1997) Protein Eng. 10, 435–444.
[11] Suzuki, T., Imamura, A., Ueguchi, C. and Mizuno, T. (1998) Plant Cell Physiol. 39, 1258–1268.
[12] Amrein, K.E., Takacs, B., Stieger, M., Molnos, J., Flint, N.A. and Burn, P. (1995) Proc. Natl. Acad. Sci. USA 92, 1048–1052.
[13] Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) J. Biol. Chem. 270, 25328–25331.
[14] Lilie, H., Schwarz, E. and Rudolph, R. (1998) Curr. Opin. Biotechnol. 9, 497–501.
[15] Buchner, J. and Rudolph, R. (1991) Biotechnology 9, 157–162.
[16] Misawa, S. and Kumagai, I. (1999) Biopolymers 51, 297–307.
[17] Umetsu, M., Tsumoto, K., Hara, M., Ashish, K., Goda, S., Adschiri, T. and Kumagai, I. (2003) J. Biol. Chem. 278, 8979–8987.
[18] Arakawa, T. and Tsumoto, K. (2003) Biochem. Biophys. Res. Commun. 304, 148–152.
[19] Altamirano, M.M., Garcia, C., Possani, L.D. and Fersht, A. (1999) Nat. Biotechnol. 17, 187–191.
[20] Tsumoto, K., Umetsu, M., Yamada, H., Ito, T., Misawa, S. and Kumagai, I. (2003) Protein Eng. 16, 535–541.
[21] Rogl, H., Kosemund, K., Kuhlbrandt, W. and Collinson, I. (1998) FEBS Lett. 432, 21–26.
[22] Nilsson, M.R. and Raleigh, D.P. (1999) J. Mol. Biol. 294, 1375–1385.
[23] Blake, C. and Serpell, L. (1996) Structure 4, 989–998.
[24] Jiménez1, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M. and Saibil, H.R. (1999) EMBO J. 18, 815–821.
[25] Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H. and Goto, Y. (2002) Nat. Struct. Biol. 9, 332–336.
[26] Tycko, R. and Ishii, Y. (2003) J. Am. Chem. Soc. 125, 6606–6607.
[27] Przybycien, T.M., Dunn, J.P., Valax, P. and Georgiou, G. (1994) Protein Eng. 7, 131–136.
[28] Fisher, B., Sumner, I. and Goodenough, P. (1993) Biotechnol. Bioeng. 41, 3–13.
[29] Tsumoto, K., Nakaoki, Y., Ueda, Y., Ogasahara, K., Yutani, K., Watanabe, K. and Kumagai, I. (1994) Biochem. Biophys. Res. Commun. 201, 546–551.
[30] Uchiyama, H., Perez-Prat, E.M., Watanabe, K., Kumagai, I. and Kuwajima, K. (1995) Protein Eng. 8, 1153–1161.
[31] Asano, R., Kudo, T., Makabe, K., Tsumoto, K. and Kumagai, I. (2002) FEBS Lett. 528, 70–76.
[32] Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. and Kikuchi, H. (1998) DNA Res. 5, 55–76.
[33] Studier, F.W. and Moffatt, B.A. (1986) J. Mol. Biol. 189, 113–130.
[34] Tsumoto, K., Shinoki, K., Kondo, H., Uchikawa, M., Juji, T. and Kumagai, I. (1998) J. Immunol. Methods 219, 119–129.
[35] Bennett, A.E., Rienstra, C.M., Auger, M. and Lakshmi, K.V. (1995) J. Chem. Phys. 103, 6951–6958.

[36] Metz, G., Wu, X. and Smith, S.O. (1994) J. Magn. Reson. Ser. A 110, 219–227.
[37] Bloembergen, N. (1949) Physica 15, 386–426.
[38] Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) Methods Enzymol. 177, 44–73.
[39] Umbarger, H.E. (1978) Annu. Rev. Biochem. 47, 533–606.
[40] Torii, H. and Tasumi, M. (1996) in: Infrared Spectroscopy of Biomolecules (Mantsch, H.H. and Chapman, D., Eds.), pp. 1–18, Wiley-Liss, New York.
[41] Murayama, K., Wu, Y., Czarnik-Matusewicz, B. and Ozaki, Y. (2001) J. Phys. Chem. B 105, 4763–4769.
[42] Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1992) Science 256, 1673–1677.
[43] Saitô, H. (1986) Magn. Reson. Chem. 24, 835–852.
[44] Saitô, H., Tuzi, S. and Naito, A. (1998) Annu. Rep. NMR Spectrosc. 36, 79–121.
[45] Pauli, J., van Rossum, B., Forster, H., de Groot, H.J.M. and Oschkinat, H. (2000) J. Magn. Reson. 143, 411–416.
[46] Yu, M.-H. and King, J. (1988) J. Biol. Chem. 263, 1424–1431.
[47] Villafane, R. and King, J. (1988) J. Mol. Biol. 204, 607–619.
[48] Wetzel, R., Perry, L.J. and Veilleux, C. (1991) Biotechnology 9, 731–737.
[49] Chrunyk, B.A., Evans, J., Lillquist, J., Young, P. and Wetzel, R. (1993) J. Biol. Chem. 268, 18053–18061.
[50] Welzel, R. and Chrunyk, B.A. (1994) FEBS Lett. 350, 245–248.
[51] Ogasahara, K., Nakamura, M., Nakura, S., Tsunasawa, S., Kato, I., Yoshimoto, T. and Yutani, K. (1998) Biochemistry 37, 17537–17544.
[52] Kaushik, J.K., Ogasahara, K. and Yutani, K. (2002) J. Mol. Biol. 316, 991–1003.
[53] Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D. and Arakawa, T. (2003) Biochem. Biophys. Res. Commun. 312, 1383–1386.