J Biomol NMR (2013) 55:189–200
DOI 10.1007/s10858-012-9701-y

ARTICLE

Secondary structural analysis of proteins based on ^{13}C chemical shift assignments in unresolved solid-state NMR spectra enhanced by fragmented structure database

Keisuke Ikeda · Ayako Egawa · Toshimichi Fujiwara

Received: 7 November 2012 / Accepted: 21 December 2012 / Published online: 29 December 2012

Abstract Magic-angle-spinning solid-state ^{13}C NMR spectroscopy is useful for structural analysis of non-crystalline proteins. However, the signal assignments and structural analysis are often hampered by the signal overlaps primarily due to minor structural heterogeneities, especially for uniformly-^{13}C,^{15}N labeled samples. To overcome this problem, we present a method for assigning ^{13}C chemical shifts and secondary structures from unresolved two-dimensional ^{13}C–^{13}C MAS NMR spectra by spectral fitting, named reconstruction of spectra using protein local structures (RESPLS). The spectral fitting was conducted using databases of protein fragmented structures related to $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts and cross-peak intensities. The experimental ^{13}C–^{13}C inter- and intra-residue correlation spectra of uniformly isotope-labeled ubiquitin in the lyophilized state had a few broad peaks. The fitting analysis for these spectra provided sequence-specific C^{α}, C^{β}, and C' chemical shifts with an accuracy of about 1.5 ppm, which enabled the assignment of the secondary structures with an accuracy of 79 %. The structural heterogeneity of the lyophilized ubiquitin is revealed from the results. Test of RESPLS analysis for simulated spectra of five different types of proteins indicated that the method allowed the secondary structure determination with accuracy of about 80 % for the 50–200 residue proteins. These results demonstrate that the RESPLS approach expands the applicability of the NMR to non-crystalline proteins exhibiting unresolved ^{13}C NMR spectra, such as lyophilized proteins, amyloids, membrane proteins and proteins in living cells.

Keywords Signal assignment · Secondary structures · Fragment assembly · Spectral simulation · Signal overlap

Introduction

Magic-angle-spinning (MAS) solid-state NMR spectroscopy is widely used for the study of structures, dynamics and inter-molecular interactions of non-crystalline proteins such as membrane proteins, amyloids and precipitated proteins (Baldus 2007). However, the sequence specific-signal assignments and subsequent structural analysis of the proteins are often quite difficult because ^{13}C-NMR line width of 1–2 ppm due to the structural heterogeneity gives rise to overlaps of signals. Such ^{13}C NMR line widths are smaller than the chemical shift changes due to secondary structure transitions, and are reported for non-crystalline proteins and biomolecules even for well-ordered structures (Kobayashi et al. 2006; Todokoro et al. 2006, 2010; Iwata et al. 2006; Fujiwara et al. 2004; Long and Tycko 1998). This line broadening is conspicuous at low temperatures where jump among low energy structures is suppressed in the glass states (Iben et al. 1989; Kitao et al. 1998). This paper presents a method that derives the chemical shifts and the secondary structures for protein systems even from low-resolution spectra without manual signal assignments. Therefore, the method enables the structure and stability

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9701-y) contains supplementary material, which is available to authorized users.

K. Ikeda · A. Egawa · T. Fujiwara (✉)
Institute for Protein Research, Osaka University,
3-2 Yamadaoka, Suita 565-0871, Japan
e-mail: tfjwr@protein.osaka-u.ac.jp

Present Address:
K. Ikeda
Graduate School of Medicine and Pharmaceutical Sciences,
University of Toyama, Sugitani 2630, Toyama 930-0194, Japan

studies of proteins in states not amenable to the standard sample optimization such as microcrystallization and solubilization, which greatly enhances the NMR applicability.

Several methods are proposed to alleviate difficulties caused by signal overlaps in solid-state NMR spectra of proteins. Crystallization and hydration of proteins decrease line widths by improving the structural order and mobility on the NMR time scale of the line widths (Jakeman et al. 1998; Martin and Zilm 2003). Selective isotope labeling is widely employed to reduce the number of cross-peaks in the spectra and to make signal assignments. One can specifically label amino acid residues with ^{13}C and ^{15}N by chemical synthesis of polypeptides. Semiselective isotope labeling such as amino-acid type selective labeling and stereo-array isotope-labeling can be performed by using bacteria and cell-free systems with labeled media (Castellani et al. 2002; Zech et al. 2005; Kainosho et al. 2006). The selective isotope labeling, however, reduces the number of spins that provides structural information and generally requires further efforts to compensate for the insufficient information especially for the analysis of the whole protein structure. Increasing the dimensionality of NMR improves the spectral resolution at the expense of signal sensitivity because the additional mixing period decreases the amplitude of the magnetization.

Limited experimental NMR data can be fully exploited by the use of informatics for NMR and proteins. The mutual dependence between chemical shifts and structure is well summarized in programs such as SHIFTX (Neal et al. 2003), SPARTA (Shen and Bax 2007), and TALOS (Cornilescu et al. 1999), which allowed the protein structure determination only by chemical shift information integrated in the protein fragment assembly technique (Shen et al. 2008). We demonstrated the determination of uniformly-^{13}C,^{15}N-labeled membrane-bound peptide structure and location in membranes using replica-exchange molecular dynamics simulations and limited solid-state NMR parameters in overlapped signals (Ikeda et al. 2011), The automatic signal assignments for the membrane-bound peptide were also feasible by using molecular dynamics simulation under pseudo-energy for the spectral fitting (Matsuki et al. 2007). In these studies, solid-state NMR spectra were analyzed quantitatively by the numerical spectral simulation from the atomic coordinates. However, a high computational cost of the simulation prevented the application of these methods to larger protein systems.

We have developed an alternative approach to obtaining the chemical shifts and secondary structures of proteins from unresolved 2D ^{13}C–^{13}C solid-state NMR spectra of uniformly ^{13}C,^{15}N-labeled proteins. This method named the reconstruction of spectra using protein local structures (RESPLS) is described in this paper. We focused on chemical shifts of protein fragmented structures depending on the amino acid sequence. Two-dimensional ^{13}C–^{13}C solid-state NMR spectra were simulated on the basis of those chemical shifts. Experimental chemical shifts were assigned by fitting the simulated spectra to experimental ones. The secondary structure was predicted from the obtained chemical shifts (Wishart et al. 1992). We applied the RESPLS approach to experimental 2D ^{13}C–NMR spectra of lyophilized ubiquitin and simulated spectra of five proteins having 46–214 amino acid residues. The structural heterogeneity of the lyophilized state was analyzed from the obtained results.

Materials and methods

The RESPLS procedure is shown by the flowcharts in Fig. 1. This procedure consists of two steps: (1) selecting candidate polypeptides for generating chemical shifts of a target protein and (2) optimizing the combination of the

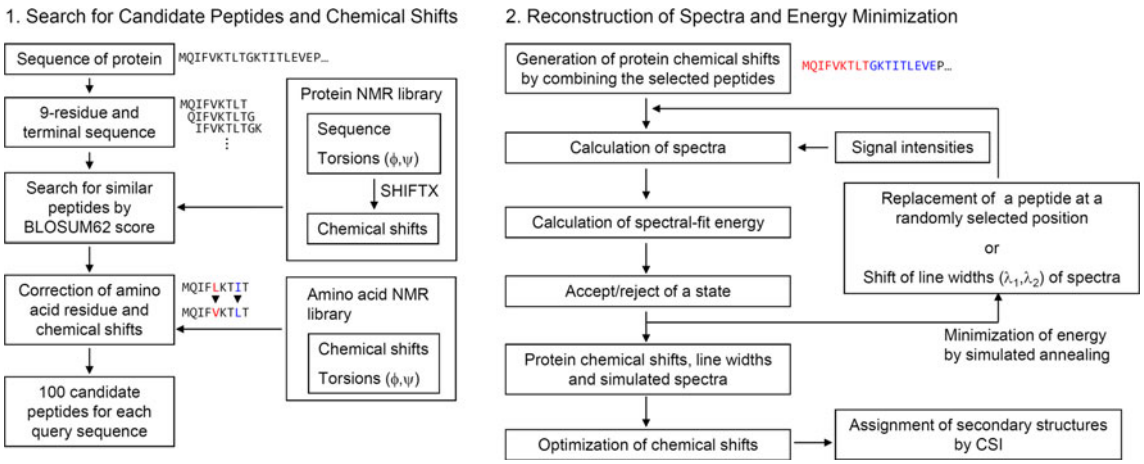

Fig. 1 Schematic representation of the RESPLS approach

peptides and chemical shifts by spectral fitting with simulated annealing algorithm.

Protein NMR library

Non-redundant proteins on the PISCES server were used for generating the polypeptides with backbone ^{13}C and $^{13}C^{\beta}$ chemical shifts (Wang and Dunbrack 2005). We first selected the 2,438 proteins with sequence identities of <40 % in the Protein Data Bank whose structures were determined by X-ray crystallography at resolutions of <1.8 Å. Then, the chemical shifts of the proteins were computed by the program SHIFTX. The amino acid sequence, torsional ϕ, ψ-angles, and $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts of the proteins were stored in the protein NMR library.

Search for peptides with high sequence similarities to the target sequence

The target n-residue protein was initially expressed by connecting peptide sequences selected from the fragmented sequence ensemble composed of nine-residue peptides beginning from residue i (= 1 to n-8) and peptides consisting of less than nine residues at the N- and C-termini. We searched for peptides in the protein NMR library with high sequence similarity to the query peptide sequences in the ensemble. The score for the sequence similarity was defined by the BLOSUM62 matrix representing the evolutional distances of two amino acids (Henikoff and Henikoff 1992). The BLOSUM62 scores for the two amino acids at the same residue number were summed to give the scoring function for the two peptides. The top 100 peptides in the scoring function at each segment were selected from the protein NMR library for the spectral fitting at the next stage.

The selected similar peptides include amino acids different from those in the query sequence. Since backbone chemical shifts primarily depend on types of amino acids and the torsional angles (Wishart et al. 1991), the chemical shifts of them should be modified by replacing the substituted amino acid. To this end, we generated an amino acid NMR library of torsional angles and chemical shifts from the protein NMR library. The chemical shifts of the substituted amino acids in the selected top 100 peptides were replaced by those of the original amino acids with the nearest ϕ and ψ angles in the amino acid NMR library.

Assembly of peptides for generating chemical shifts of the whole protein

Chemical shifts of the target protein were obtained from the protein generated by concatenating the candidate nine-residue peptides from the N-terminus. If the C-terminal peptide consisted of less than nine residues, we used a shorter peptide. We also introduced frame-shifts in the arrangement of the nine-residue peptides by inserting 1–8 residue N-terminal peptides at the N-terminus.

Calculation of cross-peak intensities and ^{13}C–^{13}C spectra

We analyzed a set of experimental two-dimensional spectra: a C_i^{α} - C_i' region of an intra-residue spin diffusion spectrum (CACO) under dipolar assisted rotational resonance (DARR) (Takegoshi et al. 2001), C_i^{α} - C'_{i-1} and C_i^{α} - C_{i-1}^{α} regions of an inter-residue CA(N)COCA spectrum (CANCO, CACA) (Fujiwara et al. 2004), and a $(C^{\alpha}+C^{\beta})_i - C_i^{\alpha}$ region of a double quantum-single quantum spectrum (DQSQ) (Hong 1999).

The magnetization transfer processes in ^{13}C–$^{15}N/^{15}N$–^{13}C cross polarization, excitation of double quantum coherences and conversion to single quantum coherences were calculated for spin systems composed of up to 6 spins for each amino acid residue by integrating time-dependent Hamiltonians as previously described (Matsuki et al. 2007). The ^{13}C–^{13}C magnetization components in the spin diffusion were calculated by a rate matrix analysis (Egawa et al. 2007). The rate of magnetization transfer between a pair of spins was determined by an internuclear distance and a zero-quantum line shape (Kubo and Mcdowell 1988). The factors from the line shapes for ^{13}C–^{13}C spin pairs with and without covalently-bonded protons were assumed to be 1.0×10^{-7} and 2.5×10^{-6}, respectively (Egawa et al. 2007). The distances of the spin pairs were calculated from an all-extended structure of the protein. Spectra for the spin diffusion were calculated from the cross-peak intensities between spins separated by up to two covalent bonds. Although this calculation of signal intensities does not consider tertiary or secondary structures of the protein, it is justified under a short mixing time, where the magnetization is transferred only between spins in proximity. Figure S1 shows observed and calculated cross-peak intensities in a spin diffusion ^{13}C–^{13}C spectrum of MLF tripeptide. The intensities of $C^{\alpha}/C^{\beta} \rightarrow C'$ and $C' \rightarrow C^{\alpha}/C^{\beta}$ transfer calculated for an extended structure of the peptide agreed with the observed ones. We also confirmed that the calculated magnetization amplitudes of ubiquitin for atomic coordinates (PDB: 1 ubq) (Vijay-Kumar et al. 1987) at spin diffusion mixing time ≤20 ms were different only by ≤4 % in RMSD from those obtained for an all-extended protein structure. These differences did not significantly affect the results of the spectral fitting simulation (data not shown). Side-chain carbons other than C^{β} which gave cross-peaks in the spectral regions, e.g. $^{13}C^{\gamma 2}$ of Thr in the DQSQ spectrum, were assumed to have average chemical shifts in the BMRB

statistics (Markley et al. 2008) or shifts determined experimentally. The spectra as a function of frequencies ω_1 and ω_2 were calculated by adding two-dimensional Lorentzian functions with cross-peak intensities I_k, half-width at half-height line widths λ_{1k}, λ_{2k}, and chemical shifts δ_{1k}, δ_{2k} for a cross-peak k,

$$S(\omega_1,\omega_2)=\sum_k I_k \frac{\lambda_{1k}}{\lambda_{1k}^2+(\omega_1-\delta_{1k})^2}\frac{\lambda_{2k}}{\lambda_{2k}^2+(\omega_2-\delta_{2k})^2}.$$

Monte Carlo simulated annealing for energy minimization

Simulated spectra were fitted to experimental spectra by Monte Carlo simulations with simulated annealing algorithm (Kirkpatrick et al. 1983).

First, an initial state with chemical shifts of the protein was generated by connecting candidate peptides each of which was randomly selected at the corresponding peptide segment. The initial line widths were set at 1.0 ppm. Second, we calculated the spectra and the energy function of spectral fit, $E=\sum_j \frac{N_j}{2}\ln\chi_j^2$ with $\chi_j^2=\sum_{i=1}^{N_j}(S_{obs,j,i}-\eta_j S_{sim,j,i})^2$. Here, N_j represents the number of data points for spectrum j = CACO, CANCO, CACA, or DQSQ. $S_{\mathrm{obs},j,i}$ and $S_{\mathrm{calc},j,i}$ denote the spectral intensities of the observed and calculated spectrum j at data point i, respectively. The factor η_j was updated so as to minimize χ_j^2. We used logarithmic functions multiplied by $0.5N_j$ to weight the experimental spectra with different data points, signal-to-noise ratios, and calculation errors as proposed in structure determination by NOE-based distance restraints (Habeck et al. 2006).

A state in the simulation was renewed by replacing a peptide with a one randomly selected from the candidates at the segment which was also randomly selected in the protein. The spectra and the energy were calculated using the chemical shifts of the renewed peptide. The new state was accepted by following the Metropolis criterion (Metropolis et al. 1953) expressed as the probability $P=1$ for $\Delta E\le 0$, and $P=\exp(-\Delta E/T)$ for $0<\Delta E$, where ΔE and T respectively stand for a change in energy of the state and temperature controlled by the simulated annealing algorithm. The initial temperature was set at 1,000. Simulated annealing was performed by decreasing the temperature as $T_{i+1}=\alpha T_i$ for 100,000 steps with $\alpha=0.9999$. The line width was also optimized every 100 steps during the search for the best combination of local structures. All cross-peaks in a 2D spectrum had the same line width for each dimension. Starting from the initial value of 1 ppm, a state was renewed by adding normal random numbers with a standard deviation of 0.1 ppm to the line width and was accepted by following the Metropolis criterion.

Further optimization of spectral-fit energy and chemical shifts

The prediction of chemical shifts by SHIFTX has errors of $\sim$1 ppm in the estimation (Seidel et al. 2009). To compensate for the error, we further optimized chemical shifts by additional energy minimization steps after the minimization described above. Chemical shifts were varied $<\pm$1.5 ppm from the initial values. A random shift with a standard deviation of 1 ppm was added to a randomly selected ^{13}C resonance. Only a change with a reduction of the energy was allowed, and iterated 100,000 times during the minimization.

In total, the whole energy minimization process was conducted 500 times for each of the N-terminal offsets of 0–8 residues. Therefore, we obtained 4,500 resultant chemical shift sets for ubiquitin with $\sim$10 h computation time on 45 CPU cores of Intel Xeon X7542 processors. The program described in this paper is available from the authors upon request.

Estimation of secondary structures by CSI

The obtained 4,500 sets were subjected to CSI programs for assigning the helix (H), extend (E), and coil (C) to amino acid residues (Wishart et al. 1992). The assignments for each residue in the 4,500 states were counted. The most assigned secondary structure was shown as the final result.

Solid-state NMR experiments for lyophilized ubiquitin

About 7 mg of uniformly-^{13}C,^{15}N-labeled ubiquitin purchased from Cambridge Isotope Laboratories (Andover, MA) was packed into a 3.2-mm rotor after the lyophilization. NMR experiments were performed with Varian Infinity-plus spectrometers at ^{1}H resonance frequencies of 600 and 700 MHz equipped with broadband triple resonance MAS probes for a 3.2-mm rotor. Sample spinning frequency was 12.5–16 kHz at a probe temperature of -60 °C. TPPM decoupling was conducted at a ^{1}H RF amplitude of 70–90 kHz (Bennett et al. 1995). A two-dimensional ^{13}C spin diffusion experiment under DARR was carried out at a ^{1}H RF amplitude of 15 kHz with a mixing time of 20 ms. Maximum t_1 and t_2 periods were 10 ms and 12.8 ms with data points 1,000 and 1,024, respectively. Repetition time was 2.0 s. Band-selective cross polarization sequences for $C^{\alpha}\rightarrow N$ and $N\rightarrow C'$ were used for the 2D CA(N)COCA experiment at the ^{13}C and ^{15}N RF amplitudes of 32 and 20 kHz with a mixing time of 4 ms (Baldus et al. 1998). The mixing time for $C'\rightarrow C^{\alpha}/C^{\beta}$ under DARR was 10 ms. The $650(t_1)\times 1{,}024(t_2)$ data matrix was acquired. A DQSQ experiment was performed with the POST-C7 sequences without ^{1}H-decoupling field under 15 kHz MAS (Hohwy et al. 1998; Hughes et al. 2004).

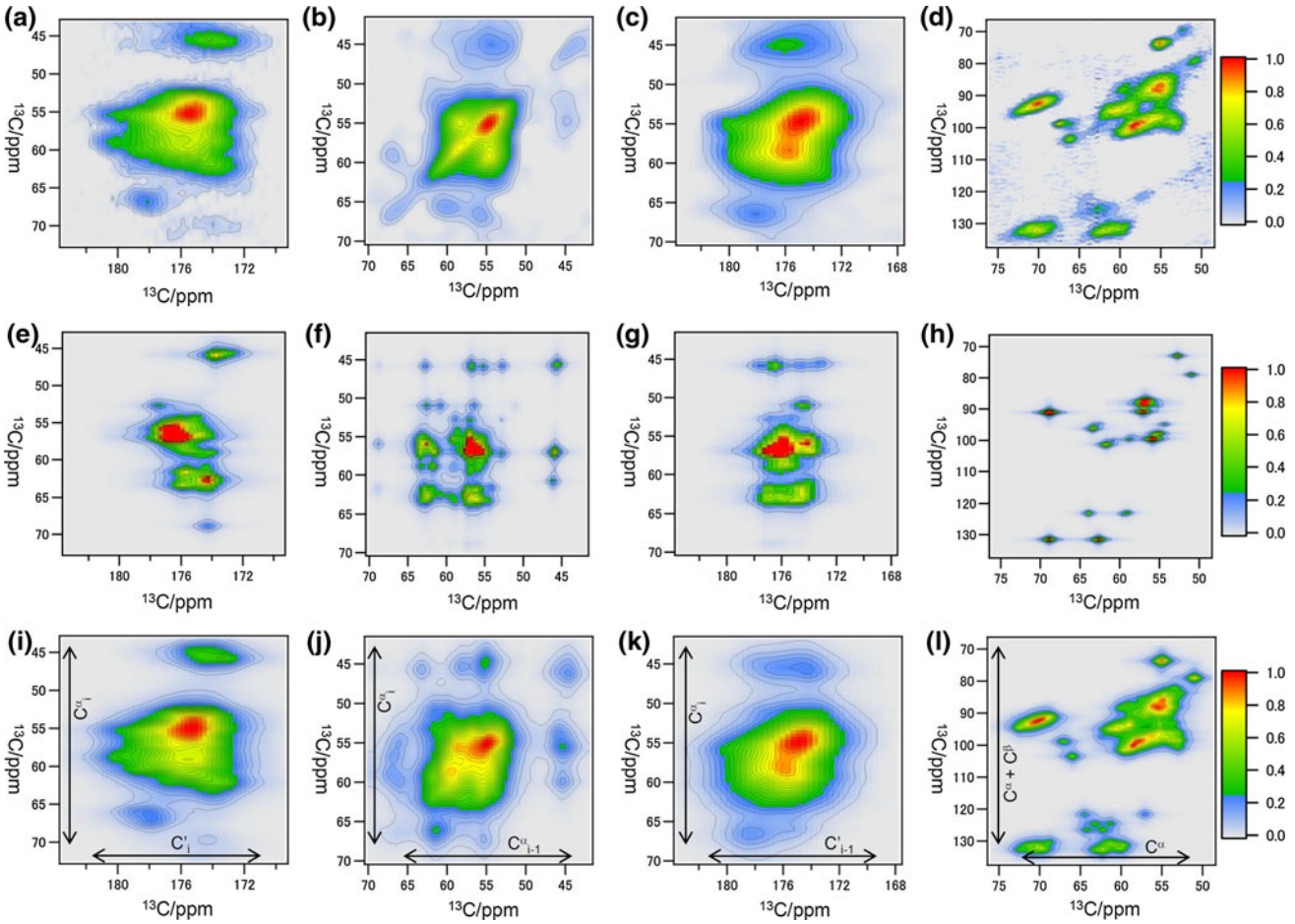

Fig. 2 MAS solid-state NMR spectra of u-[^{13}C,^{15}N] ubiquitin in the lyophilized state. **a**, **e**, **i** CACO spectra. **b**, **f**, **j** CACA spectra. **c**, **g**, **k** CANCO spectra. **d**, **h**, **i** DQSQ spectra. **a–d** Experimental spectra. **e–h** Spectra calculated from the initial chemical shifts and line widths of 1.0 ppm before the spectral fitting. **i–l** Representative calculated spectra after the RESPLS fitting. Spectral intensities larger than 1.0 are shown by *red*. Simulated spectral intensities were adjusted so that the RMSDs between the simulated and experimental spectra gave almost minimum values

Each of the excitation and conversion times for DQ coherences was 267 μs. The 360 × 1,024 data matrix was acquired. The data matrices were zero-filled to 2,048 × 2,048 with Felix 2000 software (Accelrys Inc., San Diego, CA). The free induction decays were multiplied by an exponential function with a dumping rate of 100 Hz. The signal-to-noise ratio was ∼20 for CACO, ∼10 for CACA and CANCO, and ∼15 for DQSQ experiments.

Results

RESPLS fitting of MAS solid-state NMR spectra for lyophilized ubiquitin

To evaluate our approach for estimating ^{13}C chemical shifts and secondary structures, we performed the spectral fitting for ubiquitin in a lyophilized state. The MAS solid-state spectra are shown in Fig. 2a–d. The four spectral regions for intra-residue CACO and DQSQ, and inter-residue CACA and CANCO experiments provide chemical-shift correlations for sequence specific assignments of the C^{α}, C^{β}, and C' signals. The cross-peaks in the spectra, however, are considerably overlapped, which impedes the signal assignments. The line widths of >1.5 ppm were larger than those in microcrystalline ubiquitin of <1.0 ppm (Igumenova et al. 2004; Martin and Zilm 2003; Seidel et al. 2005). However, the CACO spectral intensity seems to agree with the signal distribution of the chemical shifts for the microcrystalline ubiquitin as shown in Fig. 3 (Schubert et al. 2006).

We generated the protein chemical shifts from candidate nine-residue and terminal peptides as indicated in Fig. 1. These peptides ranked among the top 100 in sequence similarity to the target protein. Figures 4a–c show C^{α}, C^{β}, and C' secondary chemical shift distributions of the candidate peptides plotted against the residue number. These secondary shifts are defined by $\Delta\delta = \delta_{obs} - \delta_{coil}$, where

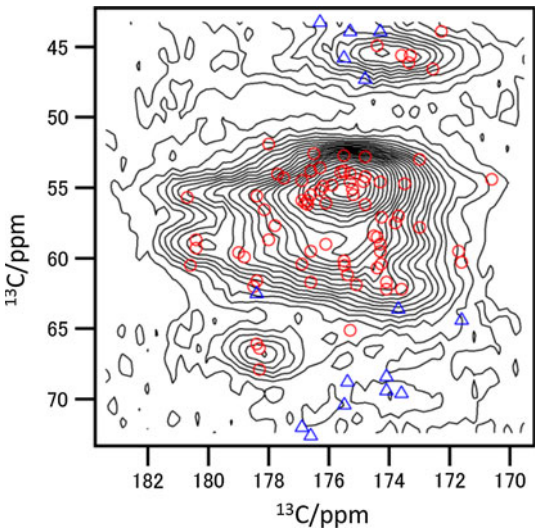

Fig. 3 CACO spectrum of ubiquitin in the lyophilized state. The *symbols* show the resonances for intra-residue chemical-shift correlations in microcrystalline ubiquitin (BMRB: 7111). *red circles* C^{α}–C' correlations. *blue triangles* C^{β}–C' correlations

δ_{coil} is the random coil value. The shifts $\Delta\delta$ were distributed around zero with large standard deviations of about 2 ppm, suggesting that the sets of the peptides included a variety of local structures. The correlation coefficients r between the average secondary chemical shifts of the peptides and the microcrystalline ubiquitin (Fig. 4g–i) were 0.35, 0.50, and 0.18 for C^{α}, C^{β}, and C', respectively. These weak correlations show that the ensemble selected only by the sequence similarity does not provide the secondary structures and chemical shifts similar to those of the target protein. Note that we removed entries for ubiquitin and its homolog proteins with sequence identities of >40 % from the protein NMR library to conduct the fitting simulation without losing general applicability.

The spectral fitting was performed by selecting the peptides whose chemical shifts were compatible with the experimental spectra of ubiquitin. This fitting calculation by the Monte Carlo simulated annealing with 4,500 initial states gave an ensemble of chemical shifts and the line widths. Spectra simulated at the beginning of the spectral

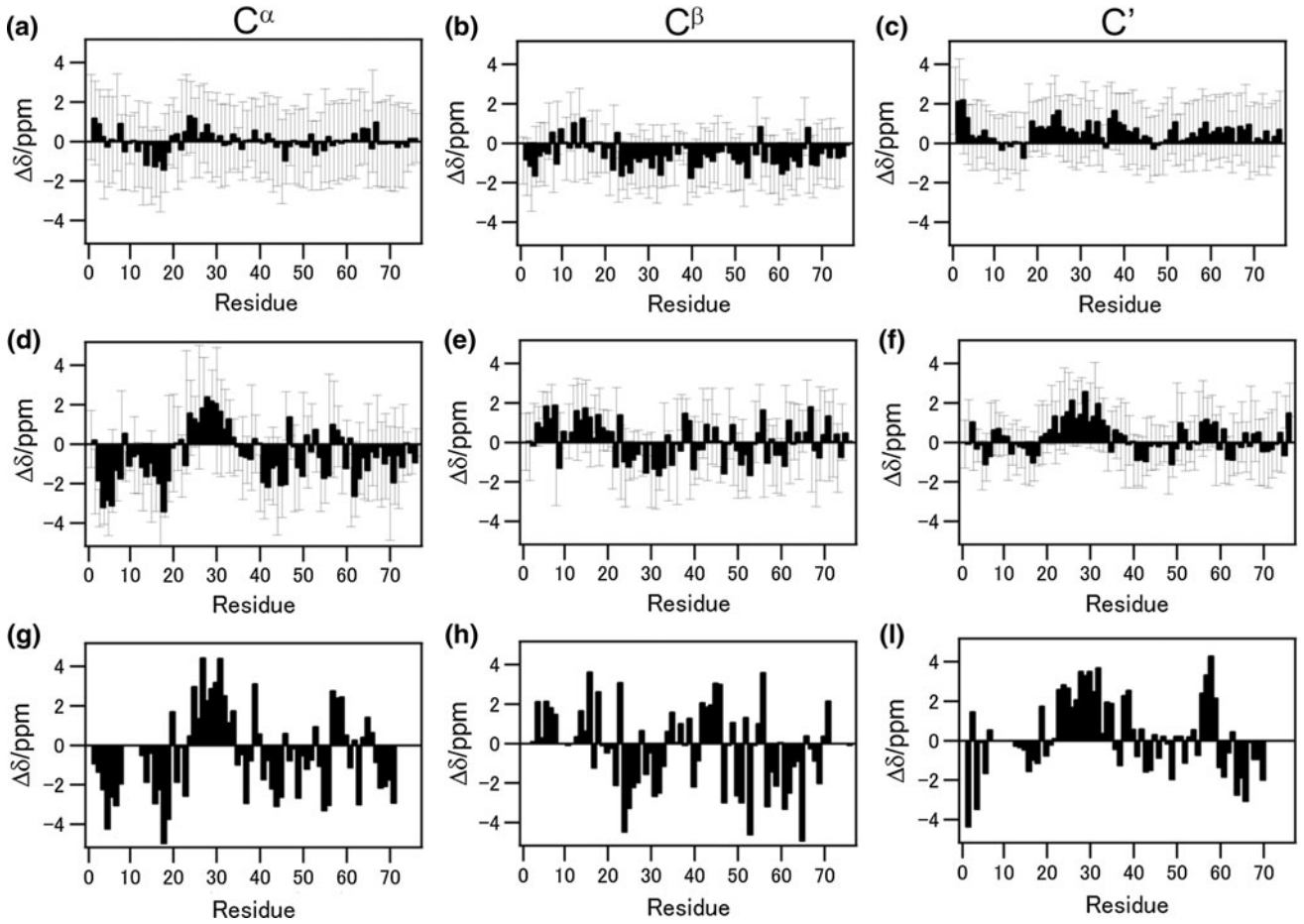

Fig. 4 Secondary chemical shifts of ubiquitin. **a–c** Chemical shifts calculated from the distributions of peptides selected by the sequence similarity score. **d–f** Chemical shifts from the distributions after the RESPLS fitting to the experimental spectra. **g–i** Chemical shifts of the microcrystalline ubiquitin. Chemical shifts for C^{α} (**a**, **d**, **g**), C^{β} (**b**, **e**, **h**) and C' (**c**, **f**, **i**) are shown. *Solid bars* show the secondary shifts, $\Delta\delta = \delta_{obs} - \delta_{coil}$. *Error bars* indicate standard deviations of the secondary shifts in the ensemble of the 4,500 estimates

fitting are shown in Fig. 2e–h. The root-mean-square deviations (RMSD) of these initial spectra from observed ones were 6.8, 3.6, 5.8, and 3.3 for CACO, CACA, CANCO, and DQSQ spectra, respectively. These values were much larger than the spectral RMS noises of 0.8, 0.7, 0.7, and 1.0, indicating the poor fit. The representative spectra of ubiquitin after the optimization by the RESPLS fitting are shown in Fig. 2i–l. The simulated spectra agree with the experimental ones in the extent of spectral noises. The average RMSDs between the observed and the 4,500 converged spectra were 1.3, 1.4, 1.3, and 1.4 with the standard deviations of <0.1 for CACO, CACA, CANCO, and DQSQ spectra, respectively. The RMSDs slightly larger than the noises would show calculation errors of the spectral simulations. The estimates of line widths were converged to 1.5–2.7 ppm with standard deviations of 0.04–0.19 ppm (Table 1). The resulting distributions of 4,500 chemical shift sets are shown in Fig. 4d–f. The obtained chemical shifts were deposited to the database BMRB (accession number 11512). Correlation coefficients of the obtained secondary shifts with those in the crystal state were relatively large: 0.7, 0.6, and 0.6 for C^{α}, C^{β}, and C' (Fig. 4d–i and Fig. S2a–c). Chemical shift RMSDs (CS-RMSDs) between the chemical shifts obtained δ_{pred} and those in the crystal state δ_{obs} are defined as $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\delta_{\mathrm{obs},i}-\delta_{\mathrm{pred},i})^2}$, where N denotes the number of C^{α}, C^{β}, and C'. The CS-RMSDs of ubiquitin were calculated at 1.44, 1.63, 1.43, and 1.50 ppm for C^{α}, C^{β}, C', and total C spins, respectively. These CS-RMSDs were comparable to the line widths of the recorded spectra, indicating the good estimations.

Chemical shifts of backbone and β carbon atoms in proteins depend on the backbone structures. The differences between the secondary shifts of C^{α} and C^{β} are used for estimating the residue-specific secondary structures in immobilized proteins (Luca et al. 2001). The obtained shifts of ubiquitin in the lyophilized state agreed with those in the crystal with $r = 0.8$ (Fig. 5a–b and Fig. S2d). The positive values at residues 22–31 and negative values at residues 2–9, 12–15, 13–15, 40–48, and 60–70 indicate α-helix and β-sheet structures, respectively. This backbone structure agreed well with that for the crystalline ubiquitin (Fig. 5c). The backbone structure was further analyzed with CSI for the C^{α}, C^{β}, and C' chemical shifts. The populations of helix (H), extend (E), and coil (C) structures in the 4,500 estimations are plotted in Fig. 5d. The most selected secondary structure for each residue was assigned as the residue-specific secondary structure. An accuracy score is defined as $Q_{SS} = \frac{N_{ss,correct}}{N_{residue}}$, where $N_{\mathrm{SS,correct}}$ and N_{residue} represent the number of correctly predicted residues and the number of residues, respectively. The score for ubiquitin was 0.79. The positions of the long α-helix and β-sheets in the lyophilized state agreed with those in the crystal with a high probability.

Table 1 Full-width at half-maximum (ppm) of ^{13}C–^{13}C signals of ubiquitin estimated by the RESPLS approach

	F1	F2
CACO[a]	1.55 ± 0.09	1.86 ± 0.09
CACA[b]	2.66 ± 0.19	2.35 ± 0.15
CANCO[c]	2.53 ± 0.15	2.51 ± 0.13
DQSQ[d]	2.33 ± 0.06	1.48 ± 0.04

[a] C^{α}–C' region of ^{13}C–^{13}C spin diffusion spectrum

[b] C^{α}–C^{α} region of CA(N)COCA spectrum

[c] C^{α}–C' region of CA(N)COCA spectrum

[d] (C^{α},C^{β})–C^{α} region of DQSQ spectrum

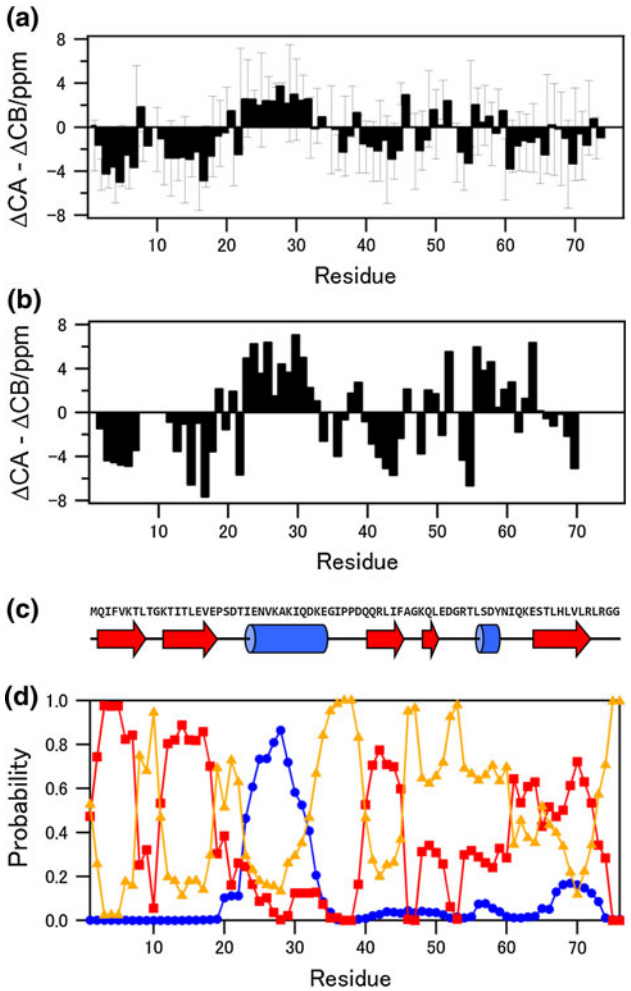

Fig. 5 Secondary structure of ubiquitin. **a** Differences between C^{α} and C^{β} secondary chemical shifts estimated by the RESPLS fitting. **b** Differences for the chemical shifts in the microcrystalline state. **c** Secondary structure in the ubiquitin X-ray structure (PDB: 1 ubq). **d** Estimated secondary structure probabilities by CSI: *blue circles* for Helix; *red squares* for Extend; *yellow triangles* for Coil

The torsion angles can also be estimated directly from those of the selected fragments by the RESPLS fitting. The angles for the 4,500 results were compared with those in the PDB structures. We classified (ϕ, ψ) angles into three regions; region 1: ($-38° < \phi < -160°$, $-22° < \psi < -70°$), region 2: ($-50° < \phi < -180°$, $10° < \psi < 180°$), and region 3: the outside of these regions. The regions 1 and 2 approximately correspond to helical and extended structures, respectively. The amino acid residues were assigned to the most populated regions in the 4,500 angles. The accuracy of the secondary structure is defined as $Q_{TA} = \frac{N_{TA,correct}}{N_{residue}}$, where the $N_{\mathrm{TA,correct}}$ represents the number of the residues whose torsional angle regions are correctly assigned. The Q_{TA} was 0.83, showing the comparable accuracy with Q_{SS} from CSI using the chemical shifts. An alternative approach for predicting torsional angles from the chemical shifts was presented as the TALOS+ program (Shen et al. 2009). We calculated Q_{TA} from the torsion angles obtained by TALOS+ with the averaged chemical shifts. The Q_{TA} of 0.76 was also comparable to Q_{SS}, indicating that our library correctly associated the chemical shifts with protein local structures as the library for TALOS+.

To estimate the contributions of component spectra to the structural analysis, we performed the spectral fitting with a smaller number of the spectra in the data set (Fig. 6). In general, removal of a spectrum decreased the accuracy of the secondary structure Q_{SS}. The elimination of the DQSQ spectrum especially affected the fitting, suggesting that this spectrum provided critical information in selecting combinations of the fragmented peptides probably owing to the relatively dispersed spectral pattern.

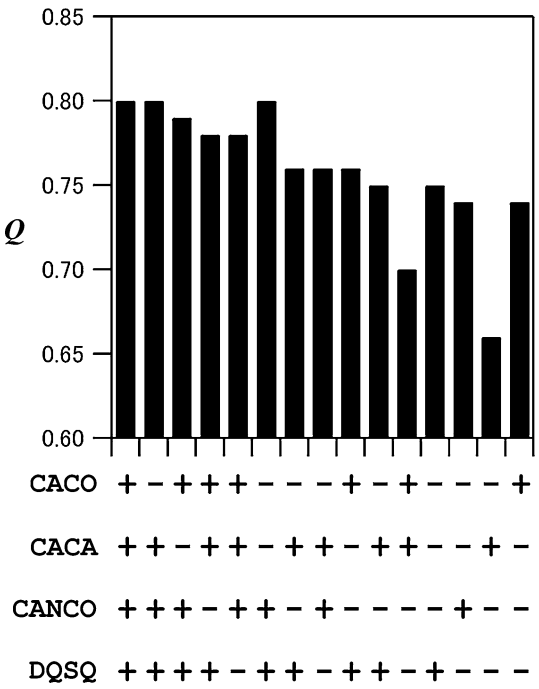

Fig. 6 Accuracy of RESPLS analysis with different combinations of the ubiquitin spectra. Accuracy Q_{SS} indicates the agreement of the estimated secondary structure with the crystal structure

Test for simulated data of proteins

To evaluate the applicability of the RESPLS fitting, we tested this method with 2D spectra simulated for five proteins whose chemical shifts were deposited in the database BMRB (Table 2). We computed the target spectra for CACO, CACA, CANCO and DQSQ with line widths of 2.2 ppm. The C^{α}, C^{β}, and C′ chemical shifts not reported in BMRB were calculated by SHIFTX with the structures deposited in PDB. Noise was added to mimic experimental spectra with S/N ratio of ~20. The RESPLS fitting was conducted for several structural types of proteins such as α/β, α-helix-rich and β-sheet-rich proteins with 46–214 residues. Similarly to the fitting for ubiquitin, entries for homolog proteins were removed from the library.

Figures S3–S12 and Table 2 show the test results. We observed considerable signal overlaps in the target spectra as in the experimental spectra of lyophilized ubiquitin. For all the test proteins, calculated spectra agreed well with the target ones after the RESPLS fitting simulations, judging from the spectral-fit RMSDs comparable to the spectral noises. β1 immunoglobulin binding domain of Protein G (GB1) (Figs. S3–S4) and HPr-like protein crh (Figs. S5–S6) are α/β proteins like ubiquitin (Franks et al. 2006; Juy et al. 2003). The obtained and target chemical shifts were correlated with $r = 0.86$ and 0.66 for GB1 and crh, respectively (Bockmann et al. 2003; Zhou et al. 2007). The total CS-RMSDs were 1.76 and 1.41 ppm. The locations of α-helix and β-sheet regions were estimated with accuracy of 70–80 %. Figures S7 and S8 show the analysis of α-helical Pf1 coat protein (Goldbourt et al. 2007; Welsh et al. 2000). A relatively high correlation coefficient 0.85, a small RMSD in chemical shifts 0.86 ppm, and the accuracy of secondary structure prediction $Q_{\mathrm{SS}} = 0.98$ exhibit the almost perfect estimation. This is probably because Pf1 takes on simple structure with a single α-helical region and a disordered N-terminal structure. We previously showed that the chemical shifts of a 14-residue α-helical peptide were precisely determined with errors of <1 ppm by spectral fitting combined with conformational energy optimization (Matsuki et al. 2007). These results indicate that the spectral analysis is easily performed for relatively small peptides and proteins with simple structures. The simulation was also carried out for the relatively large α-helix-rich 214-residue protein, Het-s(11–224) (Figs. S9–S10) (Greenwald et al. 2010; Schuetz et al. 2010). We found that the obtained pattern of cumulative α-helices was similar to that in the crystal structure. Finally, we

Table 2 Fitting results for the simulated spectra of five proteins

Protein	PDB	BMRB	$N_{residue}$	Structure	CS-rmsd (ppm)[a]	r^b	Q^c_{TA}	Q^d_{SS}
crh	1mu4	5,757	87	α/β	1.76	0.66	0.74	0.69
GB1	2gi9	15,156	56	α/β	1.41	0.86	0.82	0.77
Pf1 coat protein	1ql1	15,138	46	α-rich	0.86	0.85	0.89	0.98
Het-s(217-289)+ his-tag	2kj3	11,064	79	β (amyloid)	1.30	0.60	0.75	0.72
Het-s(11-224)	2wvn	16,965	214	α-rich	1.51	0.55	0.69	0.69

[a] RMSD between the averages of 4,500 estimated chemical shifts and the target values

[b] Correlation coefficient between estimated $\Delta\delta(C^\alpha)$–$\Delta\delta(C^\beta)$ secondary chemical shifts and the target values

[c] Accuracy of ϕ and ψ angles of the selected fragments in the 4,500 structures

[d] Accuracy of estimated secondary structure from the 4,500 chemical shifts by CSI

verified the usefulness of the RESPLS fitting for structural analysis of an aggregated protein, Het-s(217–289) as shown in Figs. S11–S12 (Lange et al. 2009; Van Melckebeke et al. 2010). The obtained chemical shift pattern was consistent with the target shifts. The locations of β-structures and loops were assigned $Q_{TA} = 0.82$ except for the six residues at the N-terminus. This mis-assignment at the N-terminus derived not from the obtained chemical shifts but from the structure assignment by CSI, because the target shifts gave almost the same structural assignment. These results suggest that the method can be applied to amyloidogenic proteins with non-native structures in fibrils or aggregates like amyloids.

The test simulations showed that the CS-RMSD, correlation coefficient r, and the accuracy of structure prediction Q_{SS} were 1.3–1.8 ppm, 0.6–0.9 and 70–80 %, respectively (Table 2). The accuracy for torsion angles Q_{TA} in the fragments selected by RESPLS was 0.7–0.9, which was comparable to the results of ubiquitin. The torsion angles predicted by TALOS+ from the average chemical shifts also gave similar accuracy (Data not shown). Thus, the RESPLS procedure can analyze the spectra of the wide range of proteins with 46–214 residues and different supra-secondary structures.

The secondary chemical shifts obtained by RESPLS are about 1/0.7 times as small as the target shifts for the five proteins. This reduction is expressed by the factor R that is defined as the ratio of the chemical shift difference $|\Delta C^\alpha - \Delta C^\beta|$ from the RESPLS analysis to that from the target shifts (Table S1). Hence the chemical shifts by RESPLS should be corrected by multiplying a factor of about 1/0.7. The line width obtained by RESPLS, however, does not need a correction, because the obtained average line width 2.3 ppm is almost the same as that for the target spectra 2.2 ppm (Table S2).

Discussion

RESPLS allowed us to analyze static structure of ubiquitin in the lyophilized state at a low temperature of −60 °C. The low spectral resolution has been hindered sequence specific structural analysis by NMR (Kennedy and Bryant 1990). Chemical shifts are useful for monitoring the immobilized structure as in the lyophilized state (Havlin and Tycko 2005). As shown in Fig. 5a, the secondary chemical shift as a function of the residue number is similar to that in the crystal structure. The chemical shift in this figure should be multiplied by 1/0.7 as described in Results to be compared with that in Fig. 5b. This similarity indicates the structure in the lyophilized state is almost the same as that in the crystal state. The average secondary chemical shift difference $|\Delta C^\alpha - \Delta C^\beta|$ obtained for the lyophilized sample S/0.7 = 2.6 ppm is close to that for the crystal 2.8 ppm (Table S1), which confirms the structural similarity. The structural homogeneity in the lyophilized state can be estimated from the line width. The line width obtained for the ubiquitin was about 2 ppm. This line width is caused by J_{C-C} couplings of 55–30 Hz (Igumenova and McDermott 2003), exponential line broadening of 100 Hz, and the field inhomogeneity and residual CH dipolar couplings 50 Hz estimated from NMR of valine (Khitrin et al. 2003). These factors account for a line width of about 1 ppm at ^{13}C resonance frequency of 175 MHz. Another 1 ppm is probably due to inhomogeneous broadening. Such a distribution in chemical shift is induced by the structural heterogeneity characterized by a backbone RMSD of about 0.1 nm (Fujiwara et al. 2004; Long and Tycko 1998). Structural trajectory in molecular dynamics simulation for a globular protein in aqueous solution exhibited a backbone RMSD of about 0.1 nm at room temperature (Steinbach and Brooks 1993). Thus the protein in the lyophilized state should have a static structural distribution given by the protein in a solution state at room temperature. Note that the narrow NMR lines for a protein solution only indicate the uniform structure averaged on a time scale of the line width and do not negate the presence of various short-lived structures (Kay 2005). In lyophilized states, the heterogeneities of native structure in aqueous solutions would be preserved under heterogeneous intermolecular interactions.

This analysis suggests that the lyophilized state is employed as a sample for studying the native structure and its fluctuation. NMR analyses of proteins have often been hampered by requirements in the sample preparation such as solubilization at high concentrations. Since the lyophilized proteins are easily prepared, the RESPLS analysis of the lyophilized states should enhance the applicability of the secondary structure estimation by NMR.

The procedure in the present method is characterized by the use of fragmented polypeptides in common with de novo protein structure prediction by fragment assembly like ROSETTA (Bradley et al. 2005; Simons et al. 1997). The fragmented peptides matching a fraction of a protein sequence in question restrict a conformational search space to plausible local structures. We applied this approach to spectral fitting and signal assignments for MAS solid-state NMR spectra. In ROSETTA and CS-ROSETTA nine-residue fragments were used for the fragment assembly because of the strong correlations between the local sequences and the structures (Shen et al. 2008; Simons et al. 1997). Thus we chose nine-residue fragments. We have shown that the C^{α}, C^{β}, and C' chemical shifts and secondary structure of ubiquitin were predictable by the RESPLS approach. The prediction tests for proteins up to 214 residues revealed that the method RESPLS can estimate chemical shifts and secondary structures with accuracy of 70–80 % for various protein folds and sequences. This method can be combined with fragment-based 3-D structure determination techniques such as CS-ROSETTA and CS23D which enabled the structure determination of small proteins only by using chemical shifts (Shen et al. 2008; Wishart et al. 2008).

The performance of the method primarily relies on the selection of peptides assuming conformations similar to the protein local structures. As seen in Fig. 2a–c, the peptides selected on the basis of the local sequence similarity had a wide range of secondary chemical shifts. Therefore, the candidate peptides had a variety of local structures which the amino acid sequences can access. The good agreements between the estimated and target chemical shifts mean that the top 100 nine-residue peptides in the similarity score are sufficient to express the chemical shift space of the proteins. The RESPLS analysis depends also on experimental spectra for fitting. We employed the four types of chemical shift correlation spectra. The line widths exceeded 1.5 ppm, resulting in signal overlaps with no isolated cross-peaks at the S/N ratio of $\sim$20. However, these spectra gave the ^{13}C chemical shifts with the accuracy of 1–2 ppm, and allowed the secondary structure estimation with the accuracy of $>\sim$70 %. As shown in Fig. 6, increase of the spectral data in the analysis should improve the accuracy and precision of the estimation. Addition of ^{15}N-edited spectra such as NCOCX and NCACX would enhance the performance.

There are some limitations in the RESPLS procedure. First, the structures of sample proteins should assume a unique structure especially in main-chain torsion angles. When the ϕ and ψ angles distribute across the different regions in the Ramachandran plot, backbone ^{13}C chemical shifts disperse in a range 4–8 ppm (Wishart et al. 1991). The present approach cannot properly express systems containing heterogeneous secondary structures such as a mixture of folded and unfolded proteins. Heterogeneous systems may be analyzed by introducing an additional parameter like residue-dependent line width. The second limitation is due to protein dynamics which reduces cross-peak intensities mainly by averaging dipolar couplings. Since peak intensities are calculated for static structures in the RESPLS fitting, the estimated chemical shifts and structure would not be reliable for dynamic systems. Thus the present method is applicable to immobile systems such as lyophilized or low-temperatures proteins. The RESPLS fitting is affected also by inaccuracy in the functional relationship between the ^{13}C chemical shifts and protein structures. Since the RMS error in predicting the ^{13}C chemical shifts by SHIFTX is $\sim$1.0 ppm (Seidel et al. 2009), experimental chemical shifts better than the precisions 1 ppm would not much improve the structure estimation. Therefore, the fitting performance should hardly be deteriorated even for the analysis of cross-peaks with line widths 1–2 ppm. The accuracy of the secondary structure estimation with RESPLS is limited also by that for CSI, which is reported to be $\sim$90 % on average (Wishart et al. 1992).

Conclusion

This article proposes a new method for determining secondary structures of non-crystalline proteins from ^{13}C chemical shifts in unresolved spectra using fragmented peptide databases. The method analyzes the solid-state NMR spectra without any peak-picking and manual signal assignments. We revealed that the protein ubiquitin maintained the native secondary structure in the lyophilized state. The secondary structures in five different types of protein were also obtained with the accuracy 70–80 % from the unresolved NMR spectra. Thus, this analysis can target proteins in the dry powders like freeze dried formulations and in low temperature glass states as used for DNP NMR study (Bayro et al. 2011; Pikal et al. 2007). We hope that the applicability of this method RESPLS is tested with other proteins having different folds and larger residue numbers. Structures of proteins in living cells and in natural membrane environments have been studied by solid-state NMR (Reckel et al. 2012; Renault et al. 2012). The NMR spectra of these systems have partially resolved

cross-peaks, where signals are assigned partially to specific residues and to amino acid spin types. Our method can be applied to these protein systems with minor heterogeneities and open a new avenue for structural biology of complex molecules by solid-state NMR.

Acknowledgments This work was supported by funding from the Targeted Proteins Research Program and KAKENHI of the Ministry of Education, Culture, Sports, Sciences and Technology of Japan.

References

Baldus M (2007) Magnetic resonance in the solid state: applications to protein folding, amyloid fibrils and membrane proteins. Eur Biophys J 36(Suppl 1):S37–S48

Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. Mol Phys 95:1197–1207

Bayro MJ, Debelouchina GT, Eddy MT, Birkett NR, MacPhee CE, Rosay M, Maas WE, Dobson CM, Griffin RG (2011) Intermolecular structure determination of amyloid fibrils with magic-angle spinning and dynamic nuclear polarization NMR. J Am Chem Soc 133:13967–13974

Bennett AE, Rienstra CM, Auger M, Lakshmi KV, Griffin RG (1995) Heteronuclear decoupling in rotating solids. J Chem Phys 103:6951–6958

Bockmann A, Lange A, Galinier A, Luca S, Giraud N, Juy M, Heise H, Montserret R, Penin F, Baldus M (2003) Solid state NMR sequential resonance assignments and conformational analysis of the 2×10.4 kDa dimeric form of the *Bacillus subtilis* protein Crh. J Biomol NMR 27:323–339

Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. Science 309:1868–1871

Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. Nature 420:98–102

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Egawa A, Fujiwara T, Mizoguchi T, Kakitani Y, Koyama Y, Akutsu H (2007) Structure of the light-harvesting bacteriochlorophyll c assembly in chlorosomes from *Chlorobium limicola* determined by solid-state NMR. Proc Natl Acad Sci USA 104:790–795

Franks WT, Wylie BJ, Stellfox SA, Rienstra CM (2006) Backbone conformational constraints in a microcrystalline U-^{15}N-labeled protein by 3D dipolar-shift solid-state NMR spectroscopy. J Am Chem Soc 128:3154–3155

Fujiwara T, Todokoro Y, Yanagishita H, Tawarayama M, Kohno T, Wakamatsu K, Akutsu H (2004) Signal assignments and chemical-shift structural analysis of uniformly ^{13}C, ^{15}N-labeled peptide, mastoparan-X, by multidimensional solid-state NMR under magic-angle spinning. J Biomol NMR 28:311–325

Goldbourt A, Gross BJ, Day LA, McDermott AE (2007) Filamentous phage studied by magic-angle spinning NMR: resonance assignment and secondary structure of the coat protein in Pf1. J Am Chem Soc 129:2338–2344

Greenwald J, Buhtz C, Ritter C, Kwiatkowski W, Choe S, Maddelein ML, Ness F, Cescau S, Soragni A, Leitz D, Saupe SJ, Riek R (2010) The mechanism of prion inhibition by HET-S. Mol Cell 38:889–899

Habeck M, Rieping W, Nilges M (2006) Weighting of experimental evidence in macromolecular structure determination. Proc Natl Acad Sci USA 103:1756–1761

Havlin RH, Tycko R (2005) Probing site-specific conformational distributions in protein folding with solid-state NMR. Proc Natl Acad Sci USA 102:3284–3289

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919

Hohwy M, Jakobsen HJ, Eden M, Levitt MH, Nielsen NC (1998) Broadband dipolar recoupling in the nuclear magnetic resonance of rotating solids: a compensated C7 pulse sequence. J Chem Phys 108:2686–2694

Hong M (1999) Solid-state dipolar INADEQUATE NMR spectroscopy with a large double-quantum spectral width. J Magn Reson 136:86–91

Hughes CE, Luca S, Baldus M (2004) Radio-frequency driven polarization transfer without heteronuclear decoupling in rotating solids. Chem Phys Lett 385:435–440

Iben IET, Braunstein D, Doster W, Frauenfelder H, Hong MK, Johnson JB, Luck S, Ormos P, Schulte A, Steinbach PJ, Xie AH, Young RD (1989) Glassy behavior of a protein. Phys Rev Lett 62:1916–1919

Igumenova TI, McDermott AE (2003) Improvement of resolution in solid state NMR spectra with J-decoupling: an analysis of lineshape contributions in uniformly C-13-enriched amino acids and proteins. J Magn Reson 164:270–285

Igumenova TI, McDermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ (2004) Assignments of carbon NMR resonances for microcrystalline ubiquitin. J Am Chem Soc 126:6720–6727

Ikeda K, Kameda T, Harada E, Akutsu H, Fujiwara T (2011) Combined use of replica-exchange molecular dynamics and magic-angle-spinning solid-state NMR spectral simulations for determining the structure and orientation of membrane-bound peptide. J Phys Chem B 115:9327–9336

Iwata K, Fujiwara T, Matsuki Y, Akutsu H, Takahashi S, Naiki H, Goto Y (2006) 3D structure of amyloid protofilaments of beta2-microglobulin fragment probed by solid-state NMR. Proc Natl Acad Sci USA 103:18119–18124

Jakeman DL, Mitchell DJ, Shuttleworth WA, Evans JNS (1998) Effects of sample preparation conditions on biomolecular solid-state NMR lineshapes. J Biomol NMR 12:417–421

Juy M, Penin F, Favier A, Galinier A, Montserret R, Haser R, Deutscher J, Bockmann A (2003) Dimerization of Crh by reversible 3D domain swapping induces structural adjustments to its monomeric homologue Hpr. J Mol Biol 332:767–776

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Guntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Kay LE (2005) NMR studies of protein structure and dynamics. J Magn Reson 173:193–207

Kennedy SD, Bryant RG (1990) Structural effects of hydration—studies of lysozyme by C-13 solids NMR. Biopolymers 29:1801–1806

Khitrin AK, Fujiwara T, Akutsu H (2003) Phase-modulated heteronuclear decoupling in NMR of solids. J Magn Reson 162:46–53

Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Kitao A, Hayward S, Go N (1998) Energy landscape of a native protein: jumping-among-minima model. Proteins 33:496–517

Kobayashi M, Matsuki Y, Yumen I, Fujiwara T, Akutsu H (2006) Signal assignment and secondary structure analysis of a uniformly [^{13}C, ^{15}N]-labeled membrane protein, H^+-ATP synthase subunit c, by magic-angle spinning solid-state NMR. J Biomol NMR 36:279–293

Kubo A, Mcdowell CA (1988) Spectral spin diffusion in polycrystalline solids under magic-angle spinning. J Chem Soc Faraday Trans 1(84):3713–3730

Lange A, Gattin Z, Van Melckebeke H, Wasmer C, Soragni A, van Gunsteren WF, Meier BH (2009) A combined solid-state NMR

and MD characterization of the stability and dynamics of the HET-s(218–289) prion in its amyloid conformation. ChemBioChem 10:1657–1665

Long HW, Tycko R (1998) Biopolymer conformational distributions from solid-state NMR: alpha-helix and 3(10)-helix contents of a helical peptide. J Am Chem Soc 120:7039–7048

Luca S, Filippov DV, van Boom JH, Oschkinat H, de Groot HJM, Baldus M (2001) Secondary chemical shifts in immobilized peptides and proteins: a qualitative basis for structure refinement under Magic Angle Spinning. J Biomol NMR 20:325–331

Markley JL, Ulrich EL, Berman HM, Henrick K, Nakamura H, Akutsu H (2008) BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. J Biomol NMR 40:153–155

Martin RW, Zilm KW (2003) Preparation of protein nanocrystals and their characterization by solid state NMR. J Magn Reson 165:162–174

Matsuki Y, Akutsu H, Fujiwara T (2007) Spectral fitting for signal assignment and structural analysis of uniformly ^{13}C-labeled solid proteins by simulated annealing based on chemical shifts and spin dynamics. J Biomol NMR 38:325–339

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. J Biomol NMR 26:215–240

Pikal MJ, Rigsbee DR, Roy ML (2007) Solid state chemistry of proteins: I. Glass transition behavior in freeze dried disaccharide formulations of human growth hormone (hGH). J Pharm Sci 96:2765–2776

Reckel S, Lopez JJ, Lohr F, Glaubitz C, Dotsch V (2012) In-cell solid-state NMR as a tool to study proteins in large complexes. ChemBioChem 13:534–537

Renault M, Tommassen-van Boxtel R, Bos MP, Post JA, Tommassen J, Baldus M (2012) Cellular solid-state nuclear magnetic resonance spectroscopy. Proc Natl Acad Sci USA 109:4863–4868

Schubert M, Manolikas T, Rogowski M, Meier BH (2006) Solid-state NMR spectroscopy of 10% ^{13}C labeled ubiquitin: spectral simplification and stereospecific assignment of isopropyl groups. J Biomol NMR 35:167–173

Schuetz A, Wasmer C, Habenstein B, Verel R, Greenwald J, Riek R, Bockmann A, Meier BH (2010) Protocols for the sequential solid-state NMR spectroscopic assignment of a uniformly labeled 25 kDa protein: HET-s(1–227). ChemBioChem 11: 1543–1551

Seidel K, Etzkorn M, Heise H, Becker S, Baldus M (2005) High-resolution solid-state NMR studies on uniformly [^{13}C,^{15}N]-labeled ubiquitin. ChemBioChem 6:1638–1647

Seidel K, Etzkorn M, Schneider R, Ader C, Baldus M (2009) Comparative analysis of NMR chemical shift predictions for proteins in the solid phase. Solid State Nucl Magn Reson 35:235–242

Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38:289–302

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Simons KT, Kooperberg C, Huang E, Baker D (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268:209–225

Steinbach PJ, Brooks BR (1993) Protein hydration elucidated by molecular-dynamics simulation. Proc Natl Acad Sci USA 90:9135–9139

Takegoshi K, Nakamura S, Terao T (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 344:631–637

Todokoro Y, Yumen I, Fukushima K, Kang SW, Park JS, Kohno T, Wakamatsu K, Akutsu H, Fujiwara T (2006) Structure of tightly membrane-bound mastoparan-X, a G-protein-activating peptide, determined by solid-state NMR. Biophys J 91:1368–1379

Todokoro Y, Kobayashi M, Sato T, Kawakami T, Yumen I, Aimoto S, Fujiwara T, Akutsu H (2010) Structure analysis of membrane-reconstituted subunit c-ring of *E. coli* H^{+}-ATP synthase by solid-state NMR. J Biomol NMR 48:1–11

Van Melckebeke H, Wasmer C, Lange A, Ab E, Loquet A, Bockmann A, Meier BH (2010) Atomic-resolution three-dimensional structure of HET-s(218–289) amyloid fibrils by solid-state NMR spectroscopy. J Am Chem Soc 132:13765–13775

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol 194:531–544

Wang G, Dunbrack RL Jr (2005) PISCES: recent improvements to a PDB sequence culling server. Nucleic Acids Res 33:W94–W98

Welsh LC, Symmons MF, Marvin DA (2000) The molecular structure and structural transition of the alpha-helical capsid in filamentous bacteriophage Pf1. Acta Crystallogr D Biol Crystallogr 56:137–150

Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222:311–333

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31:1647–1651

Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–W502

Zech SG, Wand AJ, McDermott AE (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. J Am Chem Soc 127:8618–8626

Zhou DH, Shea JJ, Nieuwkoop AJ, Franks WT, Wylie BJ, Mullen C, Sandoz D, Rienstra CM (2007) Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. Angew Chem Int Ed 46: 8380–8383