J Biomol NMR (2007) 38:325–339
DOI 10.1007/s10858-007-9170-x

ARTICLE

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

Yoh Matsuki · Hideo Akutsu · Toshimichi Fujiwara

Received: 22 January 2007 / Accepted: 24 May 2007 / Published online: 6 July 2007

Abstract We describe an approach for the signal assignment and structural analysis with a suite of two-dimensional ^{13}C–^{13}C magic-angle-spinning solid-state NMR spectra of uniformly ^{13}C-labeled peptides and proteins. We directly fit the calculated spectra to experimental ones by simulated annealing in restrained molecular dynamics program CNS as a function of atomic coordinates. The spectra are calculated from the conformation dependent chemical shift obtained with SHIFTX and the cross-peak intensities computed for recoupled dipolar interactions. This method was applied to a membrane-bound 14-residue peptide, mastoparan-X. The obtained C′, C^{α} and C^{β} chemical shifts agreed with those reported previously at the precisions of 0.2, 0.7 and 0.4 ppm, respectively. This spectral fitting program also provides backbone dihedral angles with a precision of about 50° from the spectra even with resonance overlaps. The restraints on the angles were improved by applying protein database program TALOS to the obtained chemical shifts. The peptide structure provided by these restraints was consistent with the reported structure at the backbone RMSD of about 1 Å.

Keywords Signal assignment · Magic-angle spinning · Chemical shift · Spectral simulation

Software availability The program described in this paper is available from the authors upon request.

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

Y. Matsuki
Institute for Bioinformatics Research and Development,
Japan Science and Technology Agency, 5-3 Yonbancho,
Chiyoda, Tokyo 102-8666, Japan

Introduction

The signal assignment is the basis of NMR studies on protein structure and dynamics. A number of automation programs have been developed for the efficient analysis of the solution NMR spectra (Moseley and Montelione 1999). Here, a major bottleneck has been the signal overlap. Most of the algorithms in the early period (Cieslar et al. 1988; Eccles et al. 1991; Kleywegt et al. 1991) analyzed two-dimensional (2D) ^{1}H-NMR spectra with the pattern recognition for identifying the spin systems. However, they often fail in the analysis of proteins larger than 10 kDa (Leopold et al. 1994). The introduction of the triple-resonance and recent higher dimensionality techniques contributed greatly to the improvements in spectral resolution and to the automation of the spectrum analysis (Buchler et al. 1997; Zimmerman et al. 1997; Leutner et al. 1998; Liu et al. 2005).

The signal overlap makes solid-state NMR analysis difficult. The structural heterogeneity of immobilized polypeptides leads to inhomogeneous line broadening due to chemical shift (Cadars et al. 2005; Havlin and Tycko 2005). Proteins often have ^{13}C linewidths of 1–2 ppm even with highly ordered backbone structure. Signal overlap is conspicuous for α-helical membrane proteins which mainly consist of hydrophobic amino acids having small chemical shift dispersion (Kobayashi et al. 2006). This problem is partly alleviated by semi-selective isotope-labeling (Castellani et al. 2003; Kainosho et al. 2006), protein micro-crystallization (Marulanda et al. 2004; Franks et al. 2005; and references therein) and cryoprotection (Jakeman et al. 1998). Nevertheless, except for a few favorable cases (Etzkorn et al. 2007; Lange et al. 2007), the signal overlap due to this line broadening hampers the NMR structural analysis of uniformly labeled non-crystalline proteins in

solids by reducing the spectral resolution and the signal sensitivity.

In this paper, we propose an alternative route that overcomes the signal overlap in solid states. The difference between the experimental and simulated 2D ^{13}C–^{13}C MAS spectra is minimized as a function of atomic coordinates under molecular dynamics (MD) calculation, which we refer to as Simulated-Anneal Spectrum (SAS) fitting. We have made use of recent advances in protein NMR. First, we utilized a relationship between the chemical shifts and the 3D structure of proteins (Wishart and Case 2001). This relationship allows the prediction of chemical shifts from the 3D structure. A number of software routines for the prediction (Meiler 2003; Neal et al. 2003; Wang and Jardetzky 2004) are developed along with the growth in the deposition of protein chemical shifts and structures. The backbone and sidechain chemical shifts of ^{1}H and ^{13}C provide qualitative (Ösapay and Case 1994; Wishart and Nip 1998; and references in Wishart and Case 2001) and quantitative (Cornilescu et al. 1999; Neal et al. 2006; and references therein) structural information on the basis of the NMR experiments and the quantum chemical calculations (Oldfield 2002; Xu and Case 2002). Second, we have calculated the cross-peak intensities by the exact numerical simulation of the dynamics of multispin systems in solids (Bak et al. 2000; Matsuki et al. 2004; Veshtort and Griffin 2006). The calculation gives accurate peak intensities when the magnetic relaxation due to molecular motion is ignorable. Lastly, we have employed simulated anneal structure optimization under NMR restraints (Güntert et al. 1997; Brünger et al. 1998). This optimization searches a global structural space for the purpose of signal assignment and structural analysis. We have applied the SAS-fitting to membrane-bound 14-residue peptide, mastoparan-X (MP-X).

Numerical methods

Simulated annealing for the spectral analysis

Figure 1 shows the SAS-fitting procedure. This procedure is initiated by generating the extended structure $\boldsymbol{r}_0^{(N)}$ for a primary structure $\boldsymbol{p}$ of a target peptide under investigation. The restrained molecular dynamics calculation with the simulated annealing is performed with a standard protocol defined in task file "anneal.inp" attached to CNS (Brünger et al. 1998). Here, we have added FORTRAN routines for a pseudo-energy term, E_{Sp} and the first derivative of E_{Sp} with respect to $\boldsymbol{r}^{(N)}$. The energy E_{Sp} represents the difference between the simulated and experimental spectra. The spectrum calculation from the chemical shifts and the peak intensities are described in the following sections. The spectra are calculated at every MD step as a function of coordinates $\boldsymbol{r}^{(N)} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ of N atoms in a molecule. Thus, the calculated spectra are fitted to experimental ones by simulated anneal molecular dynamics. The energy E_{Sp} is included in the objective function as $E_{\mathrm{Tot}} = E_{\mathrm{CNS}}(\boldsymbol{r}) + E_{\mathrm{Sp}}(\boldsymbol{r}, \boldsymbol{p})$. The first term is the sum of the

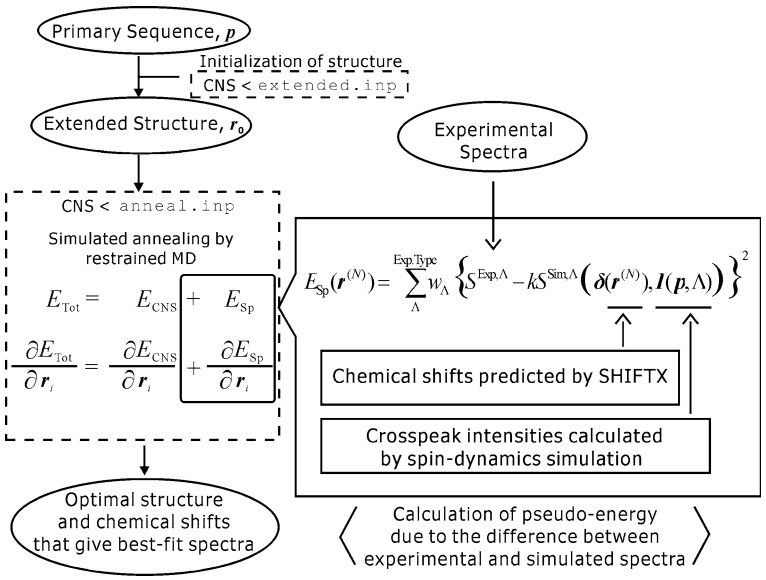

Fig. 1 Scheme for the SAS-fitting analysis. This analysis starts from extending a polypeptide chain by "extend.inp" in CNS. Simulated spectra are fitted to experimental ones by minimizing the total energy E_{Tot} with the simulated annealing MD calculation. The energy E_{Tot} contains pseudo-energy E_{CNS} in CNS and E_{Sp} for spectral fitting. Subroutine programs for calculating the energy E_{Sp} and the differentiation $\partial E_{\mathrm{Sp}}/\partial \boldsymbol{r}^{(N)}$ were added to CNS. The spectra are calculated as a function of atomic coordinates of the peptide $\boldsymbol{r}^{(N)}$

potential energy E_{Pot} and the NMR restraints E_{NMR} defined in CNS. The energy E_{Sp} is defined as

$$E_{\mathrm{Sp}}(\boldsymbol{r}) = \sum_{\Lambda} w_{\Lambda} E_{\Lambda} = \sum_{\Lambda} w_{\Lambda} \begin{cases} 0 & \text{if} \quad |D_{ij}| \le l_1 \\ \sum_{ij} D_{ij}^2 & \text{if} \quad l_1 |D_{ij}| \le l_2 \\ \sum_{ij} \left(A + B/|D_{ij}| + C|D_{ij}|\right) & \text{if} \quad l_2 < |D_{ij}| \end{cases} \tag{1}$$

where $D_{ij} = S_{ij}^{\mathrm{Exp},\Lambda} - kS_{ij}^{\mathrm{Sim},\Lambda}$ is the intensity difference between the experimental and simulated spectra at a point (i, j), and k is a normalizing factor calculated to minimize E_{Sp} by the linear least-squares fitting. The energy terms E_{Λ} and weights w_{Λ} are distinguished by Λ representing an experiment type and a region in a spectrum. The potential has a flat-bottom region at $|D_{ij}|$ less than l_1 and a linear asymptotic region at $|D_{ij}|$ larger than l_2 (Nilges and O'Donoghue 1998). The coefficients A and B were determined so that the potential is continuous and differentiable at the point $|D_{ij}| = l_2$; C is the asymptotic slope.

The SAS-fitting was performed for four spectral regions, $\Lambda = \mathrm{C'C^{\alpha}C^{\beta}}, \mathrm{C^{\alpha}C^{\beta}}, \mathrm{C^{\alpha}(N)C'}$ and $\mathrm{C^{\alpha}(NC')C^{\alpha}}$: (i) a $\mathrm{C'C^{\alpha}C^{\beta}}$ region of a broadband 2D $\mathrm{C'}_i$–$(\mathrm{C^{\alpha}},\mathrm{C^{\beta}})_i$ correlation spectrum, (ii) a $\mathrm{C^{\alpha}C^{\beta}}$ region of a 2D DQ/SQ spectrum for the $\mathrm{DQ^{\alpha\beta}}$–$(\mathrm{C^{\alpha}},\mathrm{C^{\beta}})_i$ correlations between the DQ and SQ frequencies for $^{13}\mathrm{C^{\alpha}}$–$^{13}\mathrm{C^{\beta}}$ pairs; (iii) $\mathrm{C^{\alpha}(N)C'}$ and (iv) $\mathrm{C^{\alpha}(NC')C^{\alpha}}$ regions of a 2D $\mathrm{C^{\alpha}}_i$–$(C',\mathrm{C^{\alpha}})_{i-1}$ correlation spectrum.

At the end of the simulated anneal fitting, a set of $\mathrm{C^{\alpha}}$, $\mathrm{C^{\beta}}$ and C′ chemical shifts δ_{Opt} are obtained with coordinates $\boldsymbol{r}_{\mathrm{Opt}}^{(N)}$. The potential energy E_{Pot} excludes the chemical shifts that are not allowed energetically. The SAS-fitting minimization was iterated with the same initial coordinates $\boldsymbol{r}_0^{(N)}$ and random velocities at a temperature. Each minimization gives δ_{Opt} and $\boldsymbol{r}_{\mathrm{Opt}}^{(N)}$.

Prediction of the chemical shifts

The chemical shifts for calculating 2D ^{13}C spectra were predicted from the atomic coordinates at every MD step with program SHIFTX (Neal et al. 2003). This method employs empirically derived chemical shift hypersurfaces (Le and Oldfield 1994; Wishart and Nip 1998), which relate the chemical shifts to backbone dihedral angles, ring-currents, electric fields, hydrogen bonds and nearest neighbor effects (Wagner et al. 1983; Ösapay and Case 1994). This method enables the rapid prediction of the backbone ^{15}N, $^{13}\mathrm{C^{\alpha}}$, $^{13}\mathrm{C^{\beta}}$, ^{13}C′ and $^{1}\mathrm{H^{N}}$ chemical shifts in proteins. The accuracy in chemical shift predictions was reported about 1.2 and 1.4 ppm for $\mathrm{C^{\alpha}}$ and $\mathrm{C^{\beta}}$, respectively (Heise et al. 2005), which are comparable to the linewidths due to chemical shift dispersion for conformations belonging to a specific secondary structure, because unordered peptides exhibit linewidths exceeding 4 ppm (Havlin and Tycko 2005).

Spin dynamics calculation for the cross-peak intensities

The peak intensities were computed with a program developed in-house for the exact numerical simulation of the spin dynamics under homo- and hetero-nuclear dipolar interactions for amino-acid spin systems in the powdered state. Experimental parameters such as a mixing scheme, RF and static fields, mixing time and sample spinning frequency were the same as those used in the experiments. The calculated intensities stored as a table were referenced in the spectral simulation.

The intensities for C′–$\mathrm{C^{\alpha}}$ and C′–$\mathrm{C^{\beta}}$ cross-peaks, $I^{\mathrm{C7}}_{\mathrm{C'C^{\alpha}}}$ and $I^{\mathrm{C7}}_{\mathrm{C'C^{\beta}}}$, in the $\mathrm{C'C^{\alpha}C^{\beta}}$ region were calculated for the POST-C7 sequence (Hohwy et al. 1998). The intensities for $\mathrm{C^{\alpha}}$ and $\mathrm{C^{\beta}}$ in the $\mathrm{C^{\alpha}C^{\beta}}$ region of the DQ/SQ spectrum (Hong 1999) were calculated from the DQ filter efficiencies, $I_{\mathrm{DQC^{\alpha}}}$ and $I_{\mathrm{DQC^{\beta}}}$, under the POST-C7 with the phase cycling $xy\overline{xy}$ for the reconversion block and +– +– for the signal accumulation. It has been shown that the exact numerical simulation provides accurate signal intensities for the spin dynamics under the homo- and hetero-nuclear dipolar interactions recoupled under magic-angle spinning (Specific CP between ^{15}N and ^{13}C: Baldus et al. 1998; POST-C7: Hohwy et al. 1998; RFDR: Bennett et al. 1998).

The intensities of the $\mathrm{C^{\alpha}}_i$–$\mathrm{C^{\alpha}}_{i-1}$ cross-peaks in the 2D $\mathrm{C^{\alpha}}_i$–$(\mathrm{C'},\mathrm{C^{\alpha}})_{i-1}$ spectrum (Astrof and Griffin 2002; Fujiwara et al. 2004) were calculated by multiplying the magnetization transfer efficiencies for three consecutive transfers, $\mathrm{C^{\alpha}}_i \rightarrow \mathrm{N}_i$, $\mathrm{N}_i \rightarrow \mathrm{C'}_{i-1}$, and $\mathrm{C'}_{i-1} \rightarrow \mathrm{C^{\alpha}}_{i-1}$, as $I^{0}_{\mathrm{C^{\alpha}N}} \times I^{-}_{\mathrm{NC'}} \times I^{0}_{\mathrm{C'C^{\alpha}}}$, and those of the auto-peaks $\mathrm{C^{\alpha}}_i$–$\mathrm{C^{\alpha}}_i$ as $(I^{0}_{\mathrm{C^{\alpha}N}} \times I^{0}_{\mathrm{NC'}} \times I^{0}_{\mathrm{C'C^{\alpha}}}) + (I^{+}_{\mathrm{C^{\alpha}N}} \times I^{-}_{\mathrm{NC'}} \times I^{0}_{\mathrm{C'C^{\alpha}}}) + (I^{0}_{\mathrm{C^{\alpha}N}} \times I^{-}_{\mathrm{NC'}} \times I^{+}_{\mathrm{C'C^{\alpha}}})$. Here, the superscripts 0, – and + stand for the intra-residue transfer and the transfers to the preceding and following residues, respectively. Cross-polarization was used for the transfers $\mathrm{C^{\alpha}} \rightarrow \mathrm{N}$ and $\mathrm{N} \rightarrow \mathrm{C'}$. The RFDR scheme (Bennett et al. 1998) was used for $\mathrm{C'} \rightarrow \mathrm{C^{\alpha}}$. The intensities for cross-peaks $\mathrm{C^{\alpha}}_i$–$\mathrm{C'}_{i-1}$ and $\mathrm{C^{\alpha}}_i$–$\mathrm{C'}_i$ in the $\mathrm{C^{\alpha}(N)C'}$ region were calculated as $I^{0}_{\mathrm{C^{\alpha}N}} \times I^{-}_{\mathrm{NC'}}$ and $(I^{0}_{\mathrm{C^{\alpha}N}} \times I^{0}_{\mathrm{NC'}}) + (I^{+}_{\mathrm{C^{\alpha}N}} \times I^{-}_{\mathrm{NC'}})$, respectively.

The signal intensities were calculated for 13 amino-acid spin systems: G, A, S, C, (R,K,L), P, (E,Q,M), T, (D,N), (F,Y), (H,W), I and V. We can simplify the spin systems by ignoring ^{13}C spins beyond $\mathrm{C^{\gamma}}$ without affecting the backbone and $\mathrm{C^{\beta}}$ signal intensities because the spin dynamics at short mixing time can be calculated accurately for systems consisting of a small number of spins. Thus the simplification permits grouping 20 amino-acid spin systems into

Table 1 Chemical shifts of amino-acid[a] used in spin dynamics simulation

a.a.	$\delta(C^{\alpha})$	$\delta(C^{\beta})$	$\delta(C^{\gamma})$/ppm
G	48	–	–
S	65	70[b]	–
C	58	35[b]	–
A	58	20[d]	–
R,K,L	58	35[b]	25[b]
E,Q,M	58	35[b]	35[b]
P	65	35[b]	25[b]
T	65	70[b]	25[d]
D,N	58	35[b]	180[e]
F,Y	65	35[b]	120[f]
H,W	58	35[b]	120[f]
I	65	35[c]	25[c], 20[d]
V	65	35[c]	20[d], 20[d]

[a] C′ chemical shifts were 178 ppm for all groups

[b] CH_2

[c] CH

[d] CH_3

[e] Carboxyl carbon

[f] Aromatic carbon

13 systems. These spin systems are defined by dipolar and *J* couplings, chemical shifts and the number of spins. The spins and the chemical shifts employed are summarized in Table 1. Those chemical shifts were obtained with SHIFTX for α-helix. We added the amide N atom in residue *i* and the N and C^{α} atoms in residue *i* + 1 to the spin systems to calculate $I_{C^{\alpha}N}$. To calculate $I_{NC'}$, we added N_i, C'_{i-1} and N_{i+1}, and removed all sidechain atoms. We added C'_{i-1} and C^{α}_{i+1} to calculate the other intensities mentioned above. Shielding anisotropies for ^{13}C (Pang and Zuiderweg 2000; Wei et al. 2001; Sun et al. 2002) and ^{15}N spins (Oas et al. 1987; Shoji et al. 1998) were considered.

The calculated intensities are summarized in Fig. 2a. Small chemical shift variations due to the conformation dependence did not affect the intensities. The intensities $I^{-}_{NC'} = 0.65$ and $I^{0}_{NC'} = 0.02$ were used for all the amino-acid groups. Figure 2b compares the calculated $W3^{\alpha}$–$N2^{\alpha}$ and $I6^{\alpha}$–$G5^{\alpha}$ cross-peak intensities $I^{0}_{C^{\alpha}N} \times I^{-}_{NC'} \times I^{0}_{C'C^{\alpha}}$ with the experimental intensities in the Fig. $3d_1$ for the $C^{\alpha}(NC')C^{\alpha}$ region. The agreement between the simulated and experimental intensities confirms the accuracy of the calculation.

A 2D spectrum was expressed as a sum of peaks $s_l : S^{\mathrm{Sim},\Lambda} = \sum_{l=1}^{N_\Lambda} s_l$, where N_Λ is the number of peaks in the region Λ . The peak is a function of frequencies ω_1 and ω_2:

$$s(\omega_1, \omega_2) = I \cdot \frac{\Delta_1}{\Delta_1^2 + (\omega_1 - \delta_1)^2} \cdot \frac{\Delta_2}{\Delta_2^2 + (\omega_2 - \delta_2)^2},$$

where *I* is the peak intensity; (δ_1, δ_2) and (Δ_1, Δ_2) are the chemical shifts and the linewidths. The intensities were truncated at frequency $3\Delta_i$ away from its center $\omega_i = \delta_i$.

Calculation of the first derivative for the pseudo-energy

The computation of the partial derivative of E_{Sp} with respect to $r_{m,l}$, $l = x$, y and z, needs the slope of the intensity $\partial S^{\mathrm{Sim}}_{ij}(\boldsymbol{r})/\partial r_{m,l} = \left(S^{\mathrm{Sim}}_{ij}(\boldsymbol{r}) - S^{\mathrm{Sim}}_{ij}(\boldsymbol{r}^{+}_{l})\right)/\Delta r_{m,l}$ with $\boldsymbol{r}^{+}_{l} = (\boldsymbol{r}_1, \boldsymbol{r}_2 \ldots, \boldsymbol{r}_m + \Delta r_l, \ldots \boldsymbol{r}_N)$ where *m* specifies the backbone N, C^{α}, C^{β} and C′ atoms. The chemical shifts must be calculated for a small displacement, $+\Delta r_m$. To make this computation efficient, we modified the most time-consuming subroutine in SHIFTX to run only for a residue that includes atom *m* and the neighboring residues. We also renewed only peaks whose chemical shifts were varied by the displacement.

We modified SHFITX to make the chemical shifts differentiable as a function of the atomic coordinates. We eliminated all the functional discontinuities in chemical shift by smoothing all the hypersurfaces with the cubic spline interpolation. This permitted incorporating SHIFTX into the restrained molecular dynamics program CNS.

Materials and methods

Sample preparation

The uniformly ^{13}C,^{15}N-labeled MP-X combined with the (dipalmitoylphosphatidylcholine):(dipalmitoylphosphatidylglycerol) = 4:1 deuterated lipid bilayers was packed into a rotor with a 3.2-mm diameter. About 6 mg of the MP-X was mixed with the membranes at a molar ratio, MP-X:lipids = 1:20 and a relative humidity of 32%. Details on the sample preparation are given in our publication (Todokoro et al. 2006).

NMR measurements

The initial ^{13}C magnetization was prepared by ramped-amplitude cross-polarization with a contact time of 2.0 ms. The CH dipolar interactions during the evolution and mixing periods were decoupled with a TPPM sequence at an RF amplitude of 70 kHz and CW RF fields of 85 kHz, respectively, unless otherwise specified (Fujiwara et al. 2004). The relaxation delay between experiments was 3.5 s.

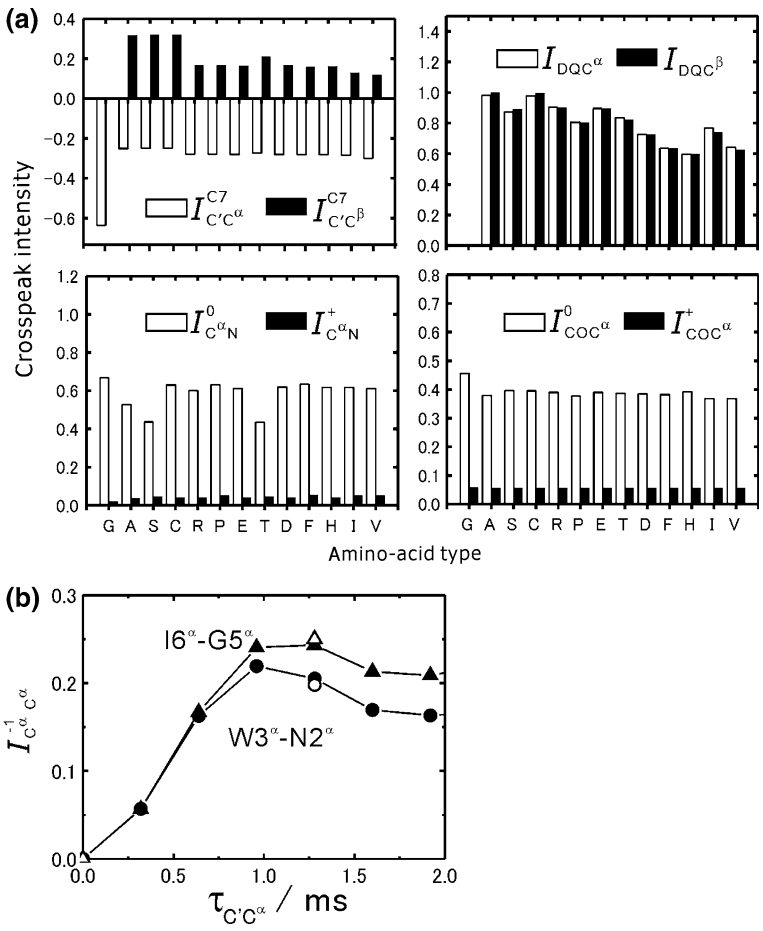

Fig. 2 (**a**) Calculated cross-peak intensities and magnetization transfer efficiencies for amino-acid spin systems. The intensities $I^{C7}_{C'C_z}$ and $I^{C7}_{C'C^\beta}$ are computed for the C′–C$^\alpha$ and C′–C$^\beta$ peaks by the ^{13}C DQ dipolar mixing, and I_{DQC^α} and I_{DQC^β} are for the C$^\alpha$ and C$^\beta$ peaks in the DQ/SQ correlation. The efficiencies $I^0_{C_zN}, I^+_{C_zN}, I^0_{C'C_z}$ and $I^+_{C'C_z}$ are computed for the $C^\alpha_i \rightarrow N_i$, $C^\alpha_i \rightarrow N_{i+1}$, $C'_{i-1} \rightarrow C^\alpha_{i-1}$, and $C'_{i-1} \rightarrow C^\alpha_i$ transfer, respectively. (**b**) Calculated and experimental cross-peak intensities for the $C^\alpha_i(N_iC'_{i-1})C^\alpha_{i-1}$ experiments. Closed circles and triangles give the calculated intensities for the W3$^\alpha$-N2$^\alpha$ and I6$^\alpha$-G5$^\alpha$ correlations as a function of the RFDR mixing time for the $C'_{i-1} \rightarrow C^\alpha_{i-1}$ transfer. Each of the $C^\alpha_i \rightarrow N_i$ and $N_i \rightarrow C'_{i-1}$ mixing times was 6 ms. The open circle and triangle show the experimental intensities in the spectrum (Fig. 3d$_1$) at [$\delta(C^\alpha_i)$, $\delta(C^\alpha_{i-1})$] = [59, 51] and [66, 47] ppm, respectively

A 2D broadband C'_i–$(C^\alpha,C^\beta)_i$ correlation spectrum was obtained with a DQ ^{13}C dipolar mixing under the POST-C7 pulse sequence (Hohwy et al. 1998) at the mixing time of 1.3 ms, a ^{13}C RF amplitude of 112 kHz and a frequency of about 100 ppm. The spectrum obtained with the DQ dipolar mixing has an advantage of distinguishing between correlations C′–C$^\alpha$ and –C$^\beta$ by the positive and negative cross peak intensities, respectively. A DQ/SQ spectrum (Hong 1999) was obtained with the POST-C7 sequence at a DQ excitation and reconversion time of 0.25 ms without a ^{1}H decoupling field (Hughes et al. 2004). A 2D C^α_i–$(C',C^\alpha)_{i-1}$ spectrum was recorded with a mixing period consisting of three steps, $C^\alpha_i \rightarrow N_i$, $N_i \rightarrow C'_{i-1}$, and $C'_{i-1} \rightarrow C^\alpha_{i-1}$ (Fujiwara et al. 2004). Band-selective ^{13}C–^{15}N cross-polarization (Baldus et al. 1998) was used for $C^\alpha_i \rightarrow N_i$ and $N_i \rightarrow C'_{i-1}$ at ^{13}C and ^{15}N RF amplitudes of 32 and 20 kHz with mixing times of 6 ms. The RFDR sequence (Bennett et al. 1998) was used for $C'_{i-1} \rightarrow C^\alpha_{i-1}$ at an RF amplitude of 70 kHz with a mixing time of 1.3 ms.

The 2D correlation experiments were performed on a Varian InfinityPlus 700 spectrometer with a double resonance MAS T3 probe at a ^{1}H resonance frequency of 699.7 MHz, a spinning frequency ν_R = 16 kHz and a probe temperature of –60°C. Only the C^α_i-$(C',C^\alpha)_{i-1}$ experiment was performed on a Varian InfinityPlus 500 spectrometer with a triple resonance MAS probe at a ^{1}H resonance frequency of 499.7 MHz and ν_R = 12.5 kHz. The FIDs were zero-filled to give matrices with 1024 × 1024 complex points and multiplied with the 100-Hz exponential line broadening function in both dimensions. The chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate according to IUPAC recommendation (Morcombe and Zilm 2003). The S/N ratio for cross-peaks was typically about 10 for the intra-residue and 5 for the inter-residue correlations.

Computation for the SAS-fitting

The MD simulation was performed with the PARALLHDG ver4.02 force field attached to CNS. Task file "anneal.inp" was run under the control of our extended CNS for the SAS-fitting. The task file specifies the filenames of the experimental spectra, spectral limits, the number of spectral points, linewidths for simulated peaks as well as parameters for E_{Sp}, l_1, l_2, C and w_Λ in Eq. (1). The spectral

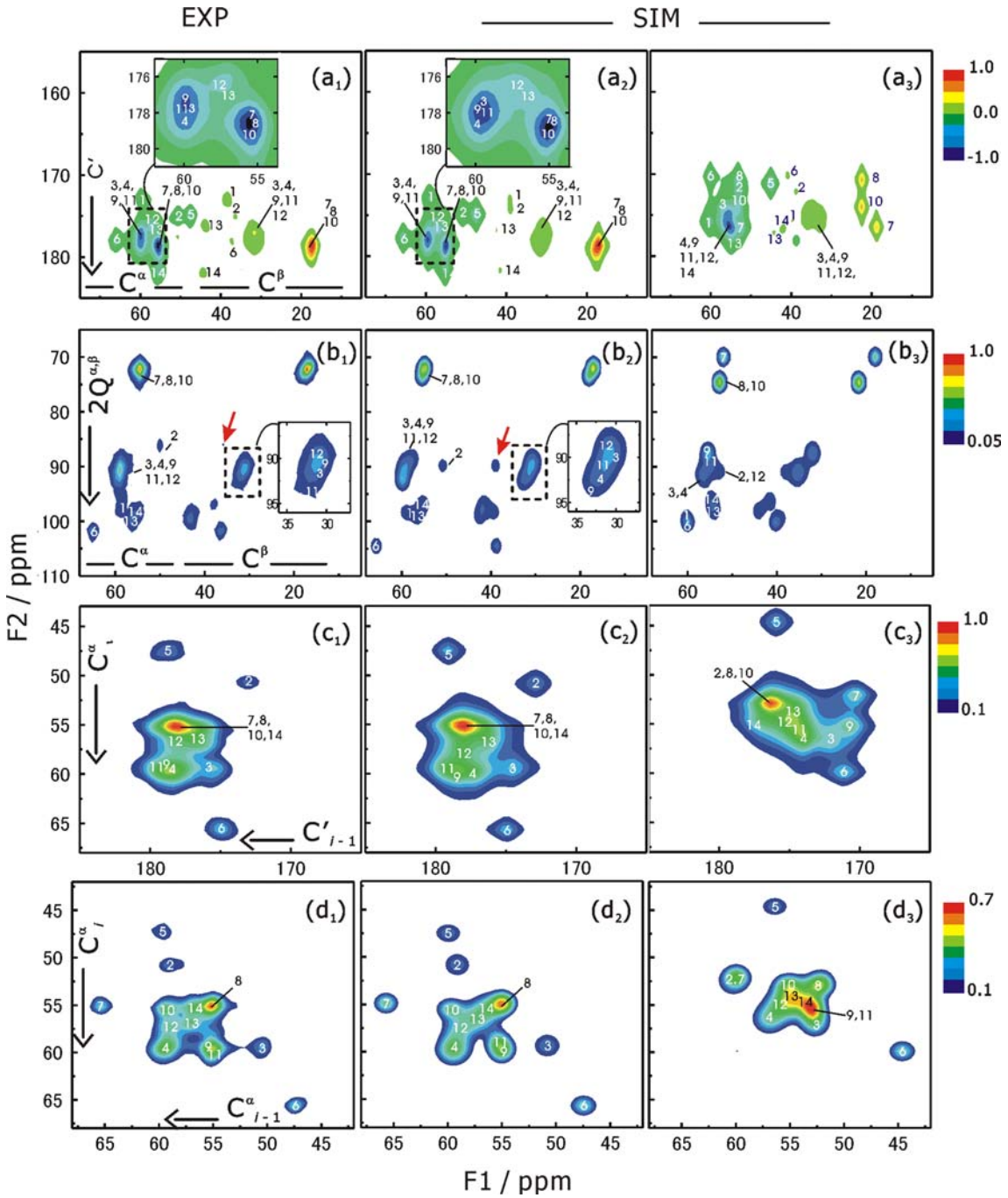

Fig. 3 2D ^{13}C–^{13}C correlation spectra of membrane-bound MP-X. The left and middle columns show the experimental and best-fit spectra, respectively. The spectra simulated at the initial MD step are shown in the right column. The regions boxed with the dashed lines are expanded in the insets. The numbers show the assigned residue numbers. ($\mathbf{a}_{1\text{-}3}$) The $C'C^{\alpha}C^{\beta}$ regions of the 2D C'_i-(C^{α},C^{β})$_i$ spectra. ($\mathbf{b}_{1\text{-}3}$) The $C^{\alpha}C^{\beta}$ regions of the DQ/SQ spectra. The residue numbers are shown only on the $DQ^{\alpha\beta}$-C^{α} peaks. ($\mathbf{c}_{1\text{–}3}$) The $C^{\alpha}_i(N_iC'_{i-1})C'_{i-1}$ regions and ($\mathbf{d}_{1\text{–}3}$) the $C^{\alpha}_i(N_iC'_{i-1})C^{\alpha}_{i-1}$ regions of the 2D C^{α}_i-(C',C^{α})$_{i-1}$ correlation spectra

regions for the fitting were as follows. The $C'C^{\alpha}C^{\beta}$ region: 5–75 ppm with 156 spectral points for the F_2 dimension and 160–190 ppm with 68 points for the F_1; the $C^{\alpha}C^{\beta}$ region: 10–75 ppm with 74 points and 65–110 ppm with 74 points; the $C^{\alpha}(N)C'$ region: 160–190 ppm with 78 points and 40–70 ppm with 78 points; the $C^{\alpha}(NC')C^{\alpha}$ region: 40–70 ppm with 78 points and 40–70 ppm with 78 points. The linewidths were about 1.7–2.5 ppm. Typically, l_1 = 0.05 and l_2 = 0.3 relative to the maximum signal intensity of about 1.0. Default values in "anneal.inp" were used unless otherwise specified. The integration time-step for the high-temperature MD and the cooling stage was 3 fs.

The system was cooled from 50,000 K to 3,000 K for 20–60 ps. The system was further cooled from 3,000 K to 0 K in the Cartesian MD with the time-step of 0.3 fs. The van der Waals energy function was scaled by 0.1, 1.0 and 4.0 at the high-temperature stage, cooling stage in the torsion-angle space, and cooling stage in the Cartesian space, respectively. The scaling factor of 150 for the distance constraints was used throughout the annealing. The relative weights of E_{Sp} for the regions $C'C^{\alpha}C^{\beta}$, $C^{\alpha}C^{\beta}$, $C^{\alpha}(N)C'$ and $C^{\alpha}(NC')C^{\alpha}$ were 0.3, 0.8, 0.8 and 1.0, respectively. The program was compiled with MIPSpro compiler on SGI/IRIX and f2c and c compilers on Linux. A PC cluster with eight 2.8-GHz Xeon processors (IA32S assembled by HPC Systems, Japan) was used for the computation.

Results

Automated signal assignment by the SAS-fitting

We applied the SAS-fitting to the analysis of solid-state NMR spectra of uniformly ^{13}C,^{15}N-labeled MP-X bound to the membranes. Amphipathic peptide MP-X, I-N-W-K-G-I-A-A-M-A-K-K-L-L-NH_2 was located at the interface between the water layer and the hydrophobic region in the bilayer system (Harada et al. 2006). The ^{13}C signal linewidths 1.6–2.1 ppm indicate that the backbone formed ordered conformation. The signals were completely assigned manually by the inter- and intra-residue dipolar spin connectivities obtained with a suite of multidimensional MAS NMR spectra (Todokoro et al. 2006).

We have analyzed the four regions of three 2D ^{13}C–^{13}C spectra shown in Fig. 3a–d by the simultaneous fitting. The residue numbers shown in the experimental spectra in the left column indicate the chemical shifts $\boldsymbol{\delta}_{Man}$ manually assigned. Table 2 shows that the C^{α} signal of K12, C^{β} signals of W3, K4, M9, K11 and K12, and C' signals of W3, M9, K11, K12 and L13 are not resolved. Thus membrane bound MP-X provides an adequate system to check the utility of our program in dealing with spectra with overlapped signals.

As is shown in Fig. 4, the pseudo-energy E_{Sp}/E^0_{Sp} decreased with the time for simulated annealing, where E^0_{Sp} is the initial energy. The insets in Fig. 4 show that all the energy terms for the spectral regions were minimized simultaneously. About 100-ps simulated annealing for MP-X took 90 h on a 2.8 GHz Xeon processor. This simulated annealing was iterated. About 50% of the total 20 minimizations gave the energy E_{Sp}/E^0_{Sp} less than 0.07, and about 30% of the total gave the energy less than 0.05. It took about 1000 h on the processor to run through the whole procedure. Since we computed on a PC cluster having eight processors, we completed the whole procedure with about 150 h.

Fig. 3a_3–d_3 shows the spectra simulated for MP-X with the initial atomic coordinates $\boldsymbol{r}_0^{(N)}$. The simulated spectra

Table 2 Signal resolution of the experimental spectra[a]

Res. No	$\delta(C^{\alpha})$				$\delta(C^{\beta})$		$\delta(C')$	
	$C'C^{\alpha}C^{\beta}$	$C^{\alpha}C^{\beta}$	$C^{\alpha}(N)C'$	$C^{\alpha}(NC')C^{\alpha}$	$C'C^{\alpha}C^{\beta}$	$C^{\alpha}C^{\beta}$	$C'C^{\alpha}C^{\beta}$	$C^{\alpha}(N)C'$
1	O	O	X	O	O	O	O	O
2	O	O	O	O	b	b	O	O
3	X	X	O	O	X	X	X	X
4	X	X	X	O	X	X	X	O
5	O	X	O	O	–	–	O	O
6	O	O	O	O	b	O	O	X
7	X	c	X	O	c	c	c	X
8	X	c	X	O	c	c	c	X
9	X	X	X	O	X	X	X	X
10	X	c	X	O	c	c	c	X
11	X	X	X	O	X	X	X	X
12	X	X	X	X	X	X	X	X
13	X	O	X	X	O	X	X	X
14	O	O	X	X	O	X	O	X

[a] The chemical shifts that can be identified by a cross-peak separated from adjacent peaks by more than and less than 1.5 ppm are marked with O and X, respectively. The chemical shifts resolved in none of the spectra are shaded

[b] Signals were well resolved but have small peak intensities

[c] Chemical shifts are restrained within a range narrower than 1.5 ppm by a single peak-cluster

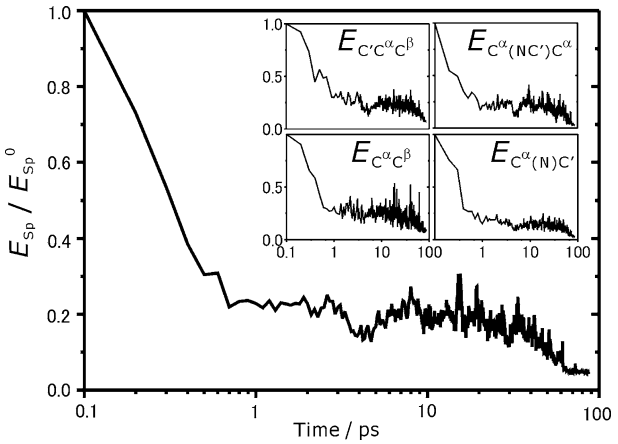

Fig. 4 The pseudo-energy E_{Sp}/E_{Sp}^0 for the fitting of spectra as a function of the simulated-annealing time shown on the logarithmic scale. Energies $E_{C'C^\alpha C^\beta}, E_{C^\alpha C^\beta}, E_{C^\alpha(NC')C^\alpha}$, and $E_{C^\alpha(N)C'}$ in the insets represent the fitting for the 2D C'_i-(C^α, C^β)$_i$ spectrum, the DQ/SQ spectrum, and the $C_i^\alpha(N_iC'_{i-1})C^\alpha_{i-1}$ and the $C_i^\alpha(N_i)C'_{i-1}$ regions of the 2D C^α_i–(C',C^α)$_{i-1}$ correlation spectrum in Fig. 3, respectively

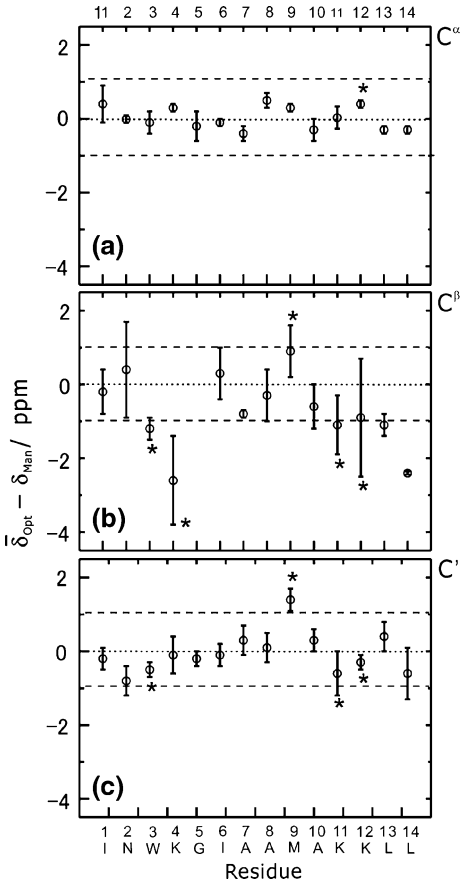

Fig. 5 The deviation of the chemical shifts averaged over the five best SAS-fitted shifts $\bar{\delta}_{Opt}$ from the manually assigned shifts δ_{Man}. The error bars show the RMSD for the five best shifts. The asterisks indicate the residues for which the chemical shifts are determined from unresolved signals in Table 2

do not fit well to the experimental spectra. The best-fit simulated spectra at $E_{Sp}/E_{Sp}^0 =$ 0.04 are shown in Fig. 3a_2–d_2. The residue numbers in the spectra indicate the assignment of the peaks. The best-fit simulated spectra agree with experimental spectra in Fig. 3a_1–d_1 in peak frequency and intensity.

The SAS-fitting minimizations at the five lowest energies E_{Sp}/E_{Sp}^0 less than 0.05 gave the C^α, C^β and C' chemical shifts with RMSDs of about 0.2, 0.7 and 0.4 ppm, respectively. These RMSDs for the residues are indicated by error bars in Fig. 5. The calculated C^α, C^β and C' chemical shifts $\boldsymbol{\delta}_{Opt}$ deviate from manually assigned shifts $\boldsymbol{\delta}_{Man}$ by 0.02 ± 0.3, –0.7 ± 1.1 and –0.06 ± 0.5 on average over all residues (Fig. 5). These deviations were smaller than the linewidths 1.6–2.1 ppm (Todokoro et al. 2006). The β-carbons of W3, K4, M9, K11, K12, L13 and L14 and the carbonyl carbon of M9 have larger deviations. The RMSD was large for C^β of N2, K4 and K12. These deviations and RMSD were mainly due to the unresolved signals except those for C^β of N2, L13 and L14 which are mentioned in Discussion.

Structure analysis by the SAS-fitting with additional structural constraints

We have calculated the structures of MP-X by simulated annealing under the energies ($E_{Pot} + E_{Dis}$), ($E_{Pot} + E_{Sp}$) and ($E_{Pot} + E_{Sp} + E_{Dis}$). Here, E_{Dis} is the energy constraints for experimentally measured distances (Todokoro et al. 2006): 6.0 ±0.4 Å for I1 [^{13}C′]–5G[^{15}N], 4.2 ±0.1 Å for G5[^{13}C′]–A8[$^{13}C^\beta$], 4.2 ± 0.1 Å for A7[^{13}C′]–A10[C^β], 4.2 ±0.1 Å for A10[^{13}C′]–L13[^{15}N], and 4.4 ±0.1 Å for A10[^{13}C′]–L14[^{15}N]. Ramachandran maps show the backbone dihedral angles, (ϕ, ψ) (Fig. 6). The standard deviations σ_ϕ and σ_ψ for ϕ and ψ are given in Fig. 7.

The standard deviations σ_ϕ and σ_ψ were smaller than about 75° at residues W3-A7, A10, K11 and L13 in the structures minimized with $E_{Pot} + E_{Sp}$. The angles at W3, K4, I6, K11 and L13 were restricted in the α-helix region about (ϕ, ψ) = (–60°, –40°). Residue N2 did not have angles in the α-helix region. These dihedral angles are consistent with the previously determined structure of MP-X forming an α-helix except I1 and N2 (Todokoro et al. 2006). However, G5, A7, A8 and A10 can take dihedral angles in the left-handed helix region about (ϕ, ψ) = (50°, 50°) as well. These angle distributions imply that the chemical shift restraints are insufficient to limit ϕ and ψ angles for residues A and G having small sidechains. This is partly because chemical shifts of alanine C^α and C^β for the α-helix are similar to those for the left-handed helix (de Dios et al. 1993; Le et al. 1995).

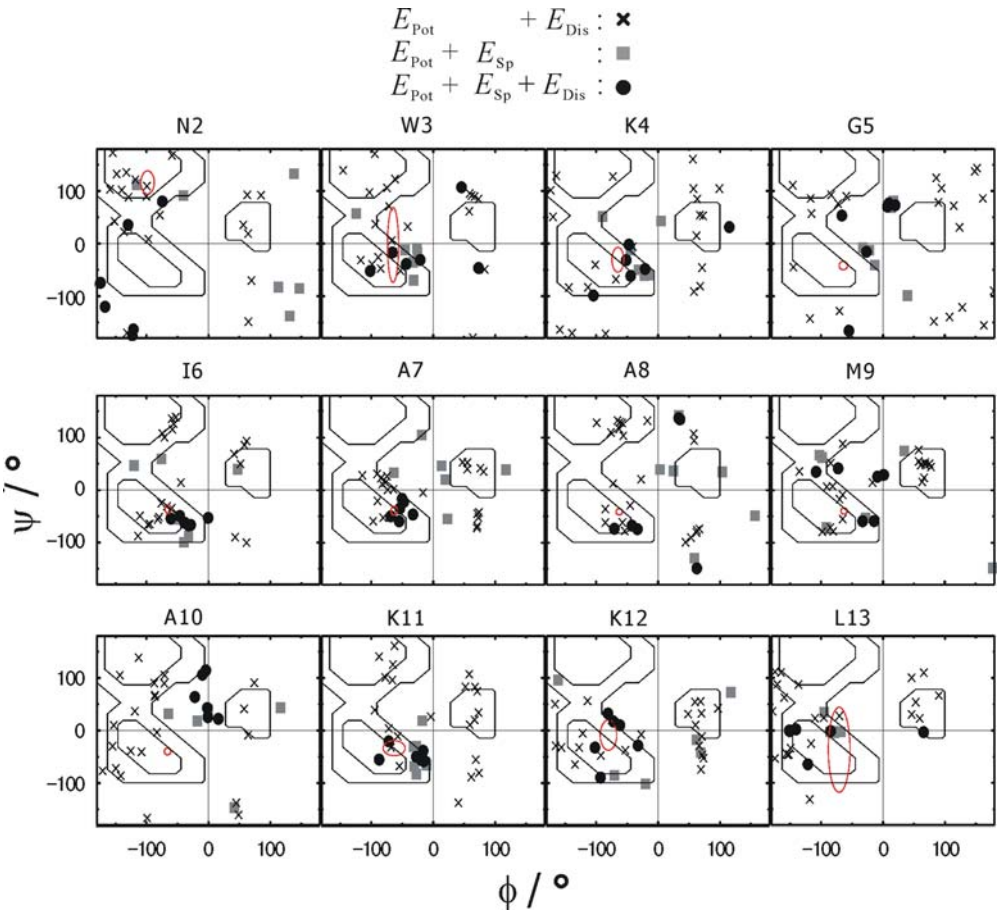

Fig. 6 Ramachandran maps for the residues in MP-X. Crosses show the backbone dihedral angles (ϕ, ψ) in the structures minimized for E_{Pot} + E_{Dis}. The grey squares and black circles show the angles in the structures minimized for E_{Pot} + E_{Sp} and E_{Pot} + E_{Sp} + E_{Dis}, respectively. The red ellipses show the angles predicted by TALOS from the shifts assigned by the SAS-fitting. The size of an ellipse gives the averaged RMSDs in ϕ and ψ. The angles (crosses) were obtained from 20 optimal structures with average energies of about E_{Pot} + E_{Dis} = 50, E_{Pot} = 50 and E_{Dis} = 0 kcal/mol provided by 20 minimization trials. The angles (squares) were obtained from six optimal structures with E_{Pot} + E_{Sp} = 900, E_{Pot} = 180 and E_{Sp} = 750 kcal/mol provided by 20 trials. The angles (circles) were obtained from six optimal structures with E_{Pot} + E_{Sp} + E_{Dis} = 750, E_{Pot} = 130, E_{Sp} = 610 and E_{Dis} = 5 kcal/mol provided by 24 trials

We characterized the structure obtained by minimizing the energy with the distance restrains, E_{Pot} + E_{Sp} + E_{Dis}. The energy E_{Sp}/E^0_{Sp} was decreased to less than 0.05 for about 10% of the minimizations that gave E_{Dis}/E^0_{Dis} less than 0.001. Residues W3-K4, I6-A7, M9 and K11-L13 in these structures had angles for α-helix. Moreover, Figs. 6 and 7c show that I6, A7, M9, A10, K11 and K12 took angles in the ranges narrower than those for the structures minimized under E_{Pot} + E_{Sp}. The standard deviations σ_ϕ and σ_ψ were smaller than 30° for I6, A7 and K11. As a control, the energy without E_{Sp}, E_{Pot} + E_{Dis} was also minimized. As can be seen in Fig. 6, the angles for all residues distributed over broad areas. These distributions are characterized by the standard deviations exceeding 75° as shown in Fig. 7a. Thus, addition of a small number of long-range distance restraints improved the precision of structure obtained by the SAS-fitting.

We have also predicted ϕ and ψ with program TALOS (Cornilescu et al. 1999). Ellipses in Fig. 6 show the standard deviations computed from the fifty (ϕ, ψ) pairs predicted for the five sets of best optimized C^α, C^β and C′ chemical shifts. The RMSDs for the angles given by TALOS were generally smaller than those given by the SAS-fitting. Five, eleven, three and fourteen (ϕ, ψ) outliers for N2, W3, L12 and L13 respectively were excluded in the subsequent structure calculation. Figure 8a shows the optimized structure with the averaged backbone RMSD, 0.8 Å for I1–L14 and 0.3 Å for K4–K12. In Fig. 8b, the optimized structure is compared with the structure colored red, which was determined previously by standard restrained MD under distance and angle restraints (Todokoro et al. 2006). The RMSD between the two structures was 1.0 Å for I1–L14 and 0.4 Å for K4–K12. Residues I1 and N2 take on extended conformation in both structures.

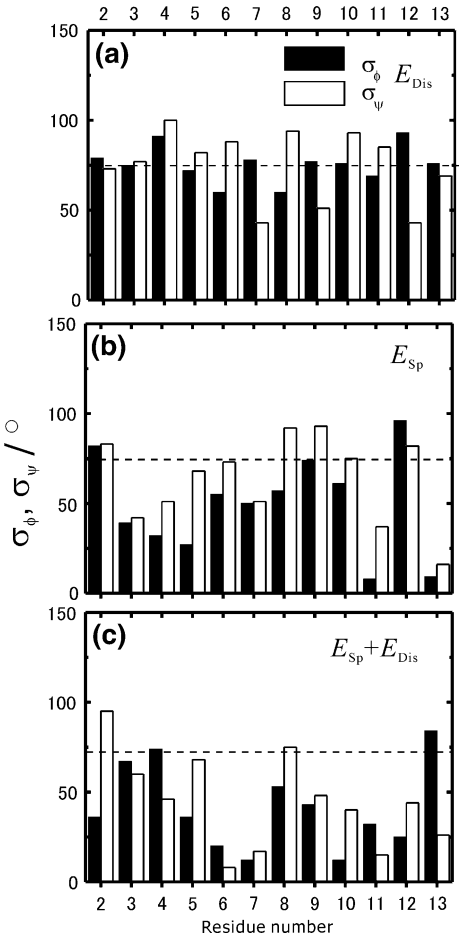

Fig. 7 Standard deviations for the ϕ and ψ angles of the residues in MP-X. The deviations shown in (**a**), (**b**) and (**c**) are calculated by the minimization for the energies $E_{Pot} + E_{Dis}$, $E_{Pot} + E_{Sp}$ and $E_{Pot} + E_{Sp} + E_{Dis}$, respectively

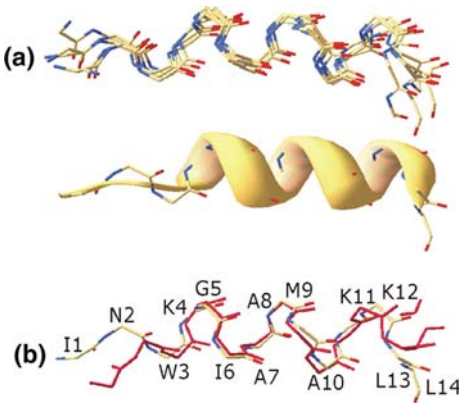

Fig. 8 Backbone structures of MP-X. (**a**) Superposition of the five optimal structures obtained from 100 simulated annealing calculations under the angle restraints given by TALOS from the shifts obtained with the SAS-fitting. The ribbon model shows the structure having the minimum RMSD from the average structure. (**b**) Superposition of the structure in (**a**) obtained with the SAS-fitting and that determined independently from the distances and angle constraints, respectively, shown in grey and red

Discussion

Signal assignment by the SAS-fitting

The precision in the chemical shift measurements primarily depended on the resolution of the spectra. Table 2 shows that the C^α resonances of all residues except K12 are resolved at least in one of the four experimental spectra. The C^α chemical shifts are well resolved in the C^α_i–C^α_{i-1} region of the inter-residue C^α_i–$(C',C^\alpha)_{i-1}$ spectrum because cross-peaks in the region are specified by two C^α chemical shifts of consecutive residues. Correspondingly, the RMSD in δ_{Opt}–δ_{Man} for C^α, 0.2 ppm, was smaller than that for C^β and C'. The $C^\alpha(N)C'$ region of the same inter-residue spectrum gave the RMSD for C', 0.4 ppm. This larger RMSD is partly due to the lower resolution for C': the C' chemical shift dispersion, about 6 ppm in MP-X, is smaller than that for the C^α dispersion, about 18 ppm. The RMSD for C^β, 0.7 ppm, was the largest, because the resolution for the intra-residue correlations that gave C^β chemical shifts is lower than that for the inter-residue correlations.

The signals were assigned by fitting to more than one spectrum. For instance, the DQ/SQ cross-peaks for C^α and C^β resonating at $[\delta(DQ^{\alpha\beta}), \delta(C^\alpha), \delta(C^\beta)] = [73, 55, 18]$ ppm gave the C^α chemical shifts of all alanine residues because only alanine has C^β chemical shift of less than 20 ppm. This allows the assignment of the intensity at $[\delta(C^\alpha_i), \delta(C^\alpha_{i-1})] = [55, 59]$ ppm to the A10$^\alpha$–M9$^\alpha$ correlation as shown by number 10 in Fig. 3d.

The C^α_i–$(C',C^\alpha)_{i-1}$ spectrum provides the C^α_i chemical shift at the two cross-peaks C^α_{i+1}–C^α_i and C^α_i–C^α_{i-1}. This feature enabled the unique assignment of the A10$^\alpha$-M9$^\alpha$ correlation mentioned above. The A10$^\alpha$–M9$^\alpha$ peak was not fit to the intensity at [55, 57] or [55, 66] ppm at the energy E_{Sp}/E^0_{Sp} less than 0.05, because the spectrum did not have peaks at [57, 55] or [66, 55] ppm for the M9$^\alpha$-A8$^\alpha$ correlation.

Many C^β and C' chemical shifts were also assigned by fitting to resolved peaks. In the $C'C^\alpha C^\beta$ region of the C'_i–$(C^\alpha,C^\beta)_i$ spectrum, the C'–C^α peaks at $[\delta(C'), \delta(C^\alpha)] = [175, 47]$ and $[178, 66]$ ppm provided $\delta(C'_5) = 175$ and $\delta(C'_6) = 178$ ppm, respectively, under the restraints $\delta(C^\alpha_5) = 47$ and $\delta(C^\alpha_6) = 66$ ppm given by the $C^\alpha(NC')C^\alpha$ region. Similarly in the DQ/SQ spectrum, the cross-peaks at $[\delta(DQ^{\alpha\beta}), \delta(C^\alpha), \delta(C^\beta)] = [104, 66, 38]$ ppm gave $\delta(C^\beta_6) = 38$ ppm under $\delta(C^\alpha_6) = 66$ ppm.

We found relatively large deviations in $\delta_{Opt} - \delta_{Man}$ for C^β shifts of residues 2, 13 and 14 determined from well-resolved peaks as shown in Fig. 5. A large error bar of ±1.4 ppm for $\delta(C^\beta_2)$ would be caused by the weak DQ/SQ peak intensity indicated by the red arrow in Fig. 3b$_1$. This can be corrected by increasing the weight, $w_{C^\alpha C^\beta}$, for the spectral area. The deviations for $\delta(C^\beta_{13})$ and $\delta(C^\beta_{14})$ are

probably due to inaccurate SHIFTX prediction of chemical shifts for residues near the amidated C-terminal.

The resonance overlap has been a major challenge in the automated signal assignments. This is because it introduces the ambiguity in the chemical shifts during the peak-picking (Thomas et al. 1991). This ambiguity leads to incorrect recognition of the spin system (Leopold et al. 1994) and to divergences of assignment possibilities (Xu et al. 2006). Alternatively, the SAS-fitting directly fits experimental spectra with simulated ones without peak-picking under the potential energy for structure. This fitting can derive chemical shifts from the degenerate resonances. For instance, in the inset of Fig. 3a$_2$, we can simulate the peak clusters at [δ(C′), δ(C$^{\alpha}$)] $\approx$ [178, 61] and [179, 56] ppm, respectively, consisting of the C′–C$^{\alpha}$ peaks for residues (W3, K4, M9, K11) and (A7, A8, A10), in agreement with the experimental results. A close look, however, revealed that the chemical shifts for residues 3, 9 and 11 differ from those manually assigned shown in Fig. 3a$_1$. This difference is due to the insensitivity of the pseudo-energy to the peak rearrangement within a cluster. Nevertheless, the deviations in δ(C′$_3$), δ(C′$_9$) and δ(C′$_{11}$) are 0.5 ± 0.2, 1.4 ± 0.4 and 0.6 ± 0.6 ppm, respectively, which are much smaller than the linewidth of the cluster for C′, about 3 ppm.

Structure analysis by the SAS-fitting with additional structural constraints

Chemical shifts for $^{15}N_H$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ and $^{1}H^{\alpha}$ can restrict the backbone dihedral angles of polypeptides within ranges of about ±40° (Wishart and Case 2001). The angles are restrained not only by the chemical shifts for residue i but also by those for residues $i \pm 1$ in the prediction by SHIFTX (Neal et al. 2003; Heise et al. 2005). The SAS-fitting takes account of the conformational energy as well. We used only the ^{13}C spectra in this work because the ^{15}N and ^{1}H spectra have lower spectral resolution for the analysis. The ϕ and ψ angles, however, were restricted with RMSD of about 50° by the SAS-fitting as shown by Figs. 6 and 7b.

The sidechain conformation defined by the χ angles has small effects on the calculation of ϕ and ψ in the SAS-fitting. This is because the χ angles have minor influences on the shift prediction by SHIFTX: variations in δ(C$^{\alpha}$) as a function of χ from 0° to 180° were about –0.2 and 0.1 ppm at ψ = –40° and 140°, respectively; those in δ(C$^{\beta}$) were –0.4 and 0.0 ppm at ϕ = –60° and –140°; those in δ(C′) were –0.1 and 0.1 ppm at ψ = –40° and 140°.

Simulated annealing was conducted also under the structural energy, distance constraints, and the pseudo-energy for the difference between the experimental and simulated ^{13}C correlation NMR spectra. This simulated annealing permitted the signal assignment at a precision of about of 0.5 ppm with structures having dihedral angle RMSD of about 50°. This accuracy in signal assignments is primarily due to the spectral constraints. This RMSD in the angles indicates that the constraints are not sufficient for structure determination. This structural distribution in the angles may partly contribute to the signal assignment because this distribution allows the search of the chemical shifts over wide chemical shift ranges as a function of atomic coordinates by using SHIFTX. However, simultaneous determination of the accurate high-resolution structure and chemical shifts may be difficult because it would contradict with the limitation in chemical shift prediction with an accuracy of about 1 ppm.

The SAS-fitting searches global conformational space of peptides for the most probable (ϕ, ψ) regions while TALOS searches structure database entries (Cornilescu et al. 1999). Thus, TALOS yielded much narrower angle ranges from the chemical shifts determined by the SAS-fitting as indicated by the ellipses in Fig. 6. The inclusion of such statistical database energy refines the structure as shown in Fig. 8 (Mertens and Gooley 2005; Nederveen et al. 2005). Similar refinement can be found in the X-ray protein structure analysis (Kleywegt 2000).

Applicability of the SAS-fitting

We can evaluate the applicability of this method to other peptide/protein systems from the results obtained for MP-X. Generally, signal assignment and structural analysis by the SAS-fitting depend on the experimental signal resolution and the precision of the chemical shift prediction by SHIFTX. The SAS-fitting gives the RMSDs in the shift and the deviations of the simulated shifts from the experimental ones as shown in Fig. 5. These quantitative measures are useful for evaluating the applicability from the quality of the assignments.

Since β-structure formation uniformly shifts the ^{13}C resonance frequencies, spectra for β-structure show the resolution similar to those for α-helix. For example, backbone ^{13}C and $^{13}C^{\beta}$ signals of a uniformly labeled 23-residue amyloid peptide assuming β-structure were almost completely assigned in a solid state (Iwata et al. 2006). Chemical shifts can be predicted for β-structure as well as α-helix by SHIFTX. Thus we can make signal assignment and structural analysis for β-structure similarly. This is exemplified by the deviations in chemical shifts for residues 1–3 and 13–14 (Fig. 5) that are not in a typical α-helix as illustrated in Fig. 8a. These deviations are not larger than the deviations in the α-helix region.

When the SAS-fitting is applied to larger molecular systems with lower spectral resolution, obtained chemical shifts would have larger RMSD. This would be caused by

the peak rearrangement within a cluster and low signal intensity as mentioned for Fig. 3a$_1$ and 3b$_1$, respectively. When the SAS-fitting is applied to rare structural elements, SHIFTX may not predict chemical shifts correctly. This led to incorrect assignments for $^{13}C^{\beta}$ resonances of Leu residues at amidated C-terminal in this study.

As long as the computation time permits, we can probably perform the signal assignment and structural analysis for larger proteins by the SAS-fitting. Preliminary results have been obtained for a fully ^{13}C,^{15}N-labeled 79-redisue protein, subunit *c* of F_oF_1-ATP synthase (Kobayashi et al. 2006). This protein consists mainly of hydrophobic residues forming α-helices. Thus, 2D ^{13}C–^{13}C correlation spectra exhibited significant signal overlaps because of the small chemical shift dispersion. We applied the quantitative spectral simulation based on the chemical shift prediction to the signal assignment and secondary structure analysis. The combined use of the spectral optimization and manual assignment allowed us to complete the assignment and provided the backbone structural information for all the residues, similarly to the application to MP-X. The spectral simulation was useful also for confirming the manual assignments. Here, the simulated annealing was, however, not employed to minimize the difference between simulated and experimental spectra.

We have applied the SAS-fitting to the spectra obtained at a low temperature in this study. The SAS-fitting would give best performance for the spectra recorded at temperatures lower than about –50°C. This is because the low temperature allows the disregard of the magnetic relaxation which is assumed in the calculation of cross-peak intensities. Solution NMR spectra are not appropriate for the target of the SAS-fitting, either. First, the signal intensities in the solution spectra depend on the *J* couplings and the magnetic relaxation. The current SAS-fitting can not deal with such signal intensities. Second, those solution spectra have signals with linewidths much narrower than 1 ppm. The linewidth of about 1 ppm partly alleviates chemical shift prediction errors smaller than 1 ppm, because the pseudo-energy for the spectra is reduced even with the prediction errors. The solution NMR spectra, however, would not benefit from the linewidths comparable to the prediction errors.

Extension of the SAS-fitting

The application of the SAS-fitting is limited by a long computation time. It needed about 3000 times longer CPU time than the simulated annealing without the SAS-fitting. The rate-limiting step is the calculation of the first derivatives of E_{Sp} that gives force on atoms. This needs the evaluation of 2D spectra computed from the chemical shifts at every MD step. Here the use of the chemical shift hypersurface based on the spline interpolation for the prediction prevents analytical differentiation of the shift. The number of chemical shifts as well as the number of atoms increases linearly with the number of residues N_A in calculating $\partial E_{Tot}(\boldsymbol{r})/\partial r_i$, so that the execution time increases as $\sim N_A^2 n_p$ with the total number of spectral points n_p in 2D spectra. Therefore, the target of analysis is practically restricted to peptides with less than about 20 residues. Future work should concentrate on accelerating the computation. Parallel computation with a PC cluster can be used for the acceleration. We can also reduce the computation time by combining the chemical shift constraints with the SAS-fitting. In this method, some of the signals are assigned at first. These obtained chemical shifts are used for the NMR constraints. The SAS-fitting is applied only to the spectra that are left when the assigned signals are subtracted. This subtraction reduces the computation time by decreasing the number of signals in the target spectra for the SAS-fitting. We have shown such combined use of the spectral optimization and the manual signal-assignment for a 79-residue protein (Kobayashi et al. 2006), though the calculated spectra were not optimized by the simulated annealing.

The current SAS-fitting program is not applicable to signal assignments of sidechain carbons beyond C^{β} for which SHIFTX does not predict their chemical shifts. Although program PROSHIFT has demonstrated the accurate sidechain chemical-shift prediction by the neural network (Meiler 2003), the algorithm is difficult to incorporate into structure optimization routines that require the differentiation of the shifts with respect to the atomic coordinates.

The SAS-fitting can also be applied to the spectra for the distance correlations. The signal intensities can be calculated by a method similar to that used for the back-calculation of NOE intensities in solution NMR. We can optimize the structure by minimizing the difference between the experimental and backcalculated spectra (Thomas et al. 1991). The quantitative deviation of calculated spectra from the experimental ones is represented by E_{Sp} in Eq. (1) for the structure optimization as well as the signal assignments. The energy E_{Sp} can characterize the precision and accuracy of the determined structure like the *R* factor in X-ray crystallography (Kleywegt 2000).

Conclusion

We have developed a method for the non-interactive assignment of the backbone C^{α}, C^{β} and C′ shifts. This method is based on the fitting of the simulated 2D ^{13}C–^{13}C spectra to the experimental ones by the restrained molecular dynamics in program CNS. The spectra were

calculated from conformation-dependent shifts predicted by SHIFTX and the cross-peak intensities computed by the spin dynamics for the multispin systems in powdered solids.

The SAS-fitting was applied to 2D MAS solid-state NMR spectra of uniformly ^{13}C,^{15}N-labeled 14-residue membrane-bound peptide, MP-X. The C^{α}, C^{β} and C′ chemical shifts were obtained respectively with deviations of 0.0 ± 0.3, –0.7 ± 1.1 and –0.1 ± 0.5 from those manually assigned. The SAS-fitting also provided the constraints on the backbone dihedral angles with RMSD of about 50°. The precisions for the dihedral angle restraints were further improved by using a small number of long-range distance constraints. The torsion angle constraints were derived by TALOS from the obtained shifts. These constraints provided the high-resolution structure, which agreed with the previously reported structure at the backbone RMSD of 1.0 Å.

We have shown that the SAS-fitting approach is useful for the signal assignment and structural analysis of immobilized peptide that gives 2D spectra with partial signal overlaps. Improvements in computation speed would automate the structure analysis of larger proteins. Advances in database and quantum mechanical studies of chemical shifts would permit more precise simulation of the spectra from the structure (Wishart and Case 2001). Developments in the theory and numerical analysis of distance-dependent magnetization transfer in solids also allow the quantitative calculation of peak intensities from the structure (Elena et al. 2006; Grommek et al. 2006; Egawa et al. 2007). These improvements would enable the SAS-fitting approach to determine the protein structure from limited NMR experiments.

Acknowledgments We would like to thank Prof. D. S. Wishart for releasing the source program of SHIFTX to us. We also thank Dr. Y. Todokoro for the sample preparation.

References

Astrof NS, Griffin RG (2002) *Soft*-triple resonance solid-state NMR experiments for assignments of U-^{13}C, ^{15}N labeled peptides and proteins. J Magn Reson 158:157–163

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

Bak M, Rasmussen JT, Nielsen NC (2000) SIMPSON: A general simulation program for solid-state NMR spectroscopy. J Magn Reson 147:296–330

Bennett AE, Rienstra CM, Griffiths JM, Zhen W, Lansbury PT Jr, Griffin RG (1998) Homonuclear radio frequency-driven recoupling in rotating solids. J Chem Phys 108:9463–9479

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr D54:905–921

Buchler NEG, Zuiderweg ERP, Wang H, Goldstein RA (1997) Protein heteronuclear NMR assignments using mean-field simulated annealing. J Magn Reson 125:34–42

Cadars S, Lesage A, Emsley L (2005) Chemical shift correlations in disordered solids. J Am Chem Soc 127:4466–4476

Castellani F, van Rossum BJ, Diehl A, Rehbein K, Oschkinat H (2003) Determination of solid-state NMR structures of proteins by means of three-dimensional ^{15}N-^{13}C-^{13}C dipolar correlation spectroscopy and chemical shift analysis. Biochemistry 42:11476–11483

Cieslar C, Clore GM, Gronenborn AM (1988) Computer-aided sequential assignment of protein ^{1}H NMR spectra. J Magn Reson 80:119–127

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

de Dios AC, Pearson JG, Oldfield E (1993) Chemical shifts in proteins: an ab initio study of carbon-13 nuclear magnetic resonance chemical shielding in glycine, alanine, and valine residues. J Am Chem Soc 115:9768–9773

Eccles C, Güntert P, Billeter M, Wüthrich K (1991) Efficient analysis of protein 2D NMR spectra using the software package EASY. J Biomol NMR 1:111–130

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

Elena B, Pintacuda G, Mifsud N, Emsley L (2006) Molecular structure determination in powders by NMR crystallography from proton spin diffusion. J Am Chem Soc 128:9555–9560

Etzkorn M, Martell S, Andronesi OC, Seidel K, Engelhard M, Baldus M (2007) Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. Angew Chem Int Ed 46: 459–462

Franks WT, Zhou DH, Wylie BJ, Money BG, Graesser DT, Frericks HL, Sahota G, Rienstra CM (2005) Magic-angle spinning solid-state NMR spectroscopy of the β^1 immunoglobulin binding domain of protein G (GB1): ^{15}N and ^{13}C chemical shift assignments and conformational analysis. J Am Chem Soc 127:12291–12305

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

Grommek A, Meier BH, Ernst M (2006) Distance information from proton-driven spin diffusion under MAS. Chem Phys Lett 427:404–409

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Harada E, Todokoro Y, Akutsu H, Fujiwara T (2006) Detection of peptide-phospholipid interaction sites in bilayer membranes by ^{13}C NMR spectroscopy: Observation of ^{2}H/^{31}P-selective ^{1}H-depolarization under magic-angle spinning. J Am Chem Soc 128:10654–10655

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

Heise H, Luca S, de Groot BL, Grubmüller H, Baldus M (2005) Probing conformational disorder in neurotensin by two-

dimensional solid-state NMR and comparison to molecular dynamics simulations. Biophys J 89:2113–2120
Hohwy M, Jakobsen HJ, Edén 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
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
Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57
Kleywegt GJ (2000) Validation of protein crystal structures. Acta Cryst D56:249–265
Kleywegt GJ, Boelens R, Cox M, Llinás M, Kaptein R (1991) Computer-assisted assignment of 2D ^{1}H NMR spectra of proteins: basic algorithms and application to phoratoxin B. J Biomol NMR 1:23–47
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
Lange A, Giller K, Hornig S, Martin-Eauclaire M-F, Pongs O, Becker S, Baldus M (2006) Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. Nature 440:959–962
Le HB, Pearson JG, de Dios AC, Oldfield E (1995) Protein structure refinement and prediction via NMR chemical shifts and quantum chemistry. J Am Chem Soc 117:3800–3807
Le H, Oldfield E (1994) Correlation between ^{15}N NMR chemical shifts in proteins and secondary structure. J Biomol NMR 4: 341–348
Leopold MF, Urbauer JK, Wand AJ (1994) Resonance assignment strategies for the analysis of NMR spectra of proteins. Mol Biotechnol 2:61–93
Leutner M, Gschwind RM, Liermann J, Schwarz C, Gemmecker G, Kessler H (1998) Automated backbone assignment of labeled proteins using the threshold accepting algorithm. J Biomol NMR 11:31–43
Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szypersk T (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. Proc Natl Acad Sci USA 102:10487–10492
Marulanda D, Tasayco ML, McDermott A, Cataldi M, Arriaran V, Polenova T (2004) Magic angle spinning solid-state NMR spectroscopy for structural studies of protein interfaces Resonance assignments of differentially enriched *Escherichia coli* thioredoxin reassembled by fragment complementation. J Am Chem Soc 126:16608–16620
Matsuki Y, Akutsu H, Fujiwara T (2004) Precision ^{1}H-^{1}H distance measurement via ^{13}C NMR signals: Utilization of ^{1}H-^{1}H double-quantum dipolar interactions recoupled under magic angle spinning conditions. Magn Reson Chem 42:291–300
Mertens HDT, Gooley PR (2005) Validating the use of database potentials in protein structure determination by NMR. FEBS Lett 579:5542–5548
Meiler J (2003) PROSHIFT: Protein chemical shift prediction using artificial neural networks. J Biomol NMR 26:25–37
Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. J Magn Reson 162:479–486
Moseley HNB, Montelione GT (1999) Automated analysis of NMR assignments and structures for proteins. Curr Opin Struct Biol 9:635–642
Neal S, Berjanskii M, Zhang H, Wishart DS (2006) Accurate prediction of protein torsion angles using chemical shifts and sequence homology. Magn Reson Chem 44:S158–S167
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
Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, Güntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AMJJ (2005) RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59: 662–672
Nilges M, O'Donoghue SI (1998) Ambiguous NOEs and automated NOE assignment. Prog Nucl Magn Reson 32:107–139
Oas TG, Hartzell CJ, Dahlquist FW, Drobny GP (1987) The amide nitrogen-15 chemical shift tensors of four peptides determined from carbon-13 dipole-coupled chemical shift powder patterns. J Am Chem Soc 109:5962–5966
Oldfield E (2002) Chemical shifts in amino acids, peptides, and proteins: from quantum chemistry to drug design. Annu Rev Phys Chem 53:349–378
Ösapay K, Case DA (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. J Biomol NMR 4: 215–230
Pang Y, Zuiderweg ERP (2000) Determination of protein backbone ^{13}CO chemical shift anisotropy tensors in solution. J Am Chem Soc 122:4841–4842
Shoji A, Ozaki T, Fujito T, Deguchi K, Ando I, Magoshi J (1998) ^{15}N chemical shift tensors and conformation of solid polypeptides containing ^{15}N-labeled glycine residue by ^{15}N NMR. J Mol Struct 441:251–266
Sun H, Sanders LK, Oldfield E (2002) Carbon-13 NMR shielding in the twenty common amino acids: comparisons with experimental results in proteins. J Am Chem Soc 124:5486–5495
Thomas PD, Basus VJ, James TL (1991) Protein solution structure determination using distances from two-dimensional nuclear overhauser effect experiments: effect of approximations on the accuracy of derived structures. Proc Natl Acad Sci USA 88:1237–1241
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
Veshtort M, Griffin RG (2006) SPINEVOLUTION: a powerful tool for the simulation of solid and liquid state NMR experiments. J Magn Reson 178:248–282
Wagner G, Pardi A, Wüthrich K (1983) Hydrogen bond length and proton NMR chemical shifts in proteins. J Am Chem Soc 105:5948–5949
Wang Y, Jardetzky O (2004) Predicting ^{15}N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from ϕ, ψ^{i-1}, and χ^{1} torsion angles. J Biomol NMR 28:327–340
Wei Y, Lee DK, Ramamoorthy A (2001) Solid-state ^{13}C NMR chemical shift anisotropy tensors of polypeptides. J Am Chem Soc 123:6118–6126
Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Meth Enzymol 338:3–34
Wishart DS, Nip AM (1998) Protein chemical shift analysis: a practical guide. Biochem Cell Biol 76:153–163

Xu XP, Case DA (2002) Probing multiple effects on ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts in peptides using density functional theory. Biopolymers 65:408–423

Xu Y, Wang X, Yang J, Vaynberg J, Qin J (2006) PASA – A program for automated protein NMR backbone signal assignment. J Biomol NMR 34:41–56

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien CY, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610