



Title	Structure determination of uniformly C-13, N-15 labeled protein using qualitative distance restraints from MAS solid-state C-13-NMR observed paramagnetic relaxation enhancement
Author(s)	Tamaki, Hajime; Egawa, Ayako; Kido, Kouki; Kameda, Tomoshi; Kamiya, Masakatsu; Kikukawa, Takashi; Aizawa, Tomoyasu; Fujiwara, Toshimichi; Demura, Makoto
Citation	Journal of biomolecular NMR, 64(1), 87-101 https://doi.org/10.1007/s10858-015-0010-0
Issue Date	2016-01
Doc URL	http://hdl.handle.net/2115/63937
Rights	The final publication is available at www.springerlink.com via http://dx.doi.org/10.1007/s10858-015-0010-0 .
Type	article (author version)
File Information	e.Proofing.pdf



[Instructions for use](#)

Structure determination of uniformly ^{13}C , ^{15}N labeled protein using qualitative distance restraints from MAS solid-state ^{13}C -NMR observed paramagnetic relaxation enhancement

Hajime Tamaki¹

Ayako Egawa²

Kouki Kido¹

Tomoshi Kameda³

Masakatsu Kamiya⁴

Takashi Kikukawa⁴

Tomoyasu Aizawa⁴

Toshimichi Fujiwara²

Makoto Demura^{4,*}

Phone +81-11-706-2771

Email demura@sci.hokudai.ac.jp

¹ Graduate School of Life Science, Hokkaido University, Sapporo, Japan

² Institute for Protein Research, Osaka University, Osaka, Japan

³ Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

⁴ Faculty of Advanced Life Science, Hokkaido University, Sapporo, Japan

Abstract

Magic angle spinning (MAS) solid-state nuclear magnetic resonance (NMR) is a powerful method for structure determination of insoluble biomolecules. However, structure determination by MAS solid-state NMR remains challenging because it is difficult to obtain a sufficient amount of distance restraints owing to spectral complexity. Collection of distance restraints from paramagnetic relaxation enhancement (PRE) is a promising approach to alleviate this barrier. However, the precision of distance restraints provided by PRE is limited in solid-state NMR because of incomplete averaged interactions and intermolecular PREs. In this report, the backbone structure of the B1 domain of streptococcal protein G (GB1) has been successfully determined by combining the CS-Rosetta protocol and qualitative PRE restraints. The derived structure has a C α RMSD of 1.49 Å relative to the X-ray structure. It is noteworthy that our protocol can determine the correct structure from only three cysteine-EDTA-Mn²⁺ mutants because this number of PRE sites is insufficient when using a conventional structure calculation method based on restrained molecular dynamics and simulated annealing. This study shows that qualitative PRE restraints can be employed effectively for protein structure determination from a limited conformational sampling space using a protein fragment library.

Keywords

Solid-state NMR
Magic-angle spinning
Protein structure
Paramagnetic relaxation enhancement
CS-Rosetta

Electronic supplementary material

The online version of this article (doi:10.1007/s10858-015-0010-0) contains supplementary material, which is available to authorized users.

Introduction

Magic angle spinning (MAS) solid-state nuclear magnetic resonance (NMR) has developed rapidly as a powerful method for structure determination of insoluble biomolecules, such as membrane proteins and amyloid fibrils. Improvements in MAS solid-state NMR have arisen because of advances in technologies, such as high-field magnets, spectrometers and probes, new radio frequency (RF) schemes/techniques, and sample preparation. Recently, MAS solid-state NMR achieved a significant milestone for protein structure determination with studies describing the structure of heptahelical membrane proteins (Park et al. 2012; Wang et al. 2013). Nonetheless, despite those technical developments, protein structure determination is still a challenging task by solid-state NMR. A major bottleneck is the collection of through space distance correlations via dipolar couplings, which are very important for determining protein folds.

The primary reason for this bottleneck is that dipolar couplings between ^1H , ^{13}C and ^{15}N spins are scaled with the inverse third power of the internuclear distance. Thus, the signal intensities from these spatial correlations are usually weak, because small through space dipolar couplings are quenched by the simultaneous presence of much larger dipolar couplings across chemical bonds. This phenomenon is called dipolar truncation (Grommek et al. 2006). To alleviate or avoid dipolar truncation, many techniques have been proposed. ^{13}C spin dilution can reduce the quenching effect by directly bonded ^{13}C neighbors and is used widely in structure determination (Castellani et al. 2002; Loquet et al. 2013; Wang et al. 2013; Demers et al. 2014). The combination of ^1H - ^1H spin diffusion and indirect detection of ^{13}C or ^{15}N (e.g., CHHC/NHHC) is a powerful method for the detection of spatial ^1H - ^1H contacts (Lange et al. 2002). Proton-assisted recoupling (PAR) avoids homonuclear dipolar truncation by the third spin-assisted recoupling mechanism (De Paëpe et al. 2008). However, a drawback of such methods is that the experiments are time consuming.

The second drawback is spectral complexity. A spatial correlation by dipolar recoupling between a nuclei pair gives rise to one or two peaks in the spectrum. In general, multi-spin recoupling is used for collection of spatial correlations, leading to the appearance of hundreds or thousands of peaks in a spectrum.

Chemical shift ambiguity caused by inhomogeneous line broadening prevents accurate and correct assignment of the resonances. Automatic assignment methods of ambiguous distance restraints (e.g., ARIA and CANDID) have been introduced for solid-state NMR data analysis using methods that were originally developed for solution-state NMR (Manolikas et al. 2008; Loquet et al. 2008). Some protein structures have been solved using these methods; however, the success of assigning ambiguous restraints relies on the precision of the chemical shifts. To assign chemical shifts with high precision, the peak line widths should be narrow. Micro-crystallization is a promising method to suppress inhomogeneous line broadening (Pauli et al. 2000). However, this approach is not applicable, for example, to lipid-bound membrane proteins and protein fibrils. Therefore, the success of the automatic assignment depends highly on sample properties. Based on the above shortcomings, a method for obtaining long-range distance restraints without increasing the number of cross peaks would be highly advantageous in structure determination by solid-state NMR.

The measurement of nuclear paramagnetic relaxation enhancement (PRE) is a powerful method to obtain long-range ($\sim 20\text{--}25$ Å) distance restraints without increasing the number of cross peaks. PRE distance restraints are obtained easily using covalently linked paramagnetic tags at specific sites in a protein. The unpaired electron enhances the relaxation rates of NMR active nuclei, leading to faster decay of peak intensities and resonance line broadening. Therefore, distance restraints can be obtained from simple chemical shift correlation spectra once the backbone assignment process is complete. PRE restraints have been shown to be effective for global protein fold determination with a limited NOE (nuclear Overhauser effect) data set using solution-state NMR (Battiste and Wagner 2000; Liang et al. 2006; Gottstein et al. 2012). Additionally, PRE restraints improve the accuracy of automated NOE assignments (Furuita et al. 2015). Recently, structure determination methods using PRE distance restraints have been introduced to solid-state NMR (Sengupta et al. 2012; Wang et al. 2012; Wang et al. 2013).

The major drawback of PRE is the distance information derived from PREs has a lower precision when compared with other techniques. In solid-state NMR, the distal unpaired electron-nucleus dipolar coupling can be interfered by homo- and heteronuclear dipolar couplings between ^1H and ^{13}C spins (Lewandowski 2013). To avoid this precision problem, Jaroniec and co-workers used

Cu^{2+} - ^{15}N longitudinal PREs to improve precision (Nadaud et al. 2009). The smaller gyromagnetic ratio of the ^{15}N nucleus and its lower abundance in proteins is preferable for nuclear dipolar coupling suppression. Additionally, the small spin quantum number and single electronic relaxation time of Cu^{2+} are favorable to prevent undesired signal intensity decay during the direct and indirect detection periods. However, this approach is rather difficult to use when trying to obtain side chain restraints. This approach is also time consuming owing to the slow ^{15}N longitudinal relaxation rate. In addition, intermolecular PREs caused by dense molecular packing in solid-state samples also reduce the precision. Thus, a method that can readily obtain PRE profiles and use low precision distance restraints is attractive for protein structure analysis by solid-state NMR.

In this paper, an approach for structure determination of proteins that combines qualitative transverse PRE restraints and CS-Rosetta is described. The qualitative transverse PRE profiles are easily obtained from the peak intensity ratio of paramagnetic and diamagnetic samples. Using the CS-Rosetta fragment assembly, it is possible to compensate the quality of the PRE restraints from solid-state NMR measurements. Rosetta is a successful method for de novo protein structure prediction of small proteins (Rohl et al. 2004; Bradley 2005). NMR researchers have been extending the Rosetta protocol, for example, fragment picking restrained by backbone chemical shifts (Shen et al. 2008), conformational sampling combined with NOE restraints (Lange and Baker 2011) and automated NOE assignment (Lange 2014). Currently, Rosetta is a suitable method for determining protein structure from limited restraint data sets, e.g., backbone chemical shifts, methyl–methyl/methyl–amide/amide–amide NOEs and residual dipolar couplings (RDCs) of deuterated proteins in solution (Lange et al. 2012). Here, we demonstrate the backbone structure determination of the EDTA- Mn^{2+} complex attached B1 domain of streptococcal protein G (GB1) by Rosetta fragment assembly using backbone chemical shifts and relaxation profiles of ^{13}C - ^{13}C correlation spectra by short DARR mixing (Takegoshi et al. 2001).

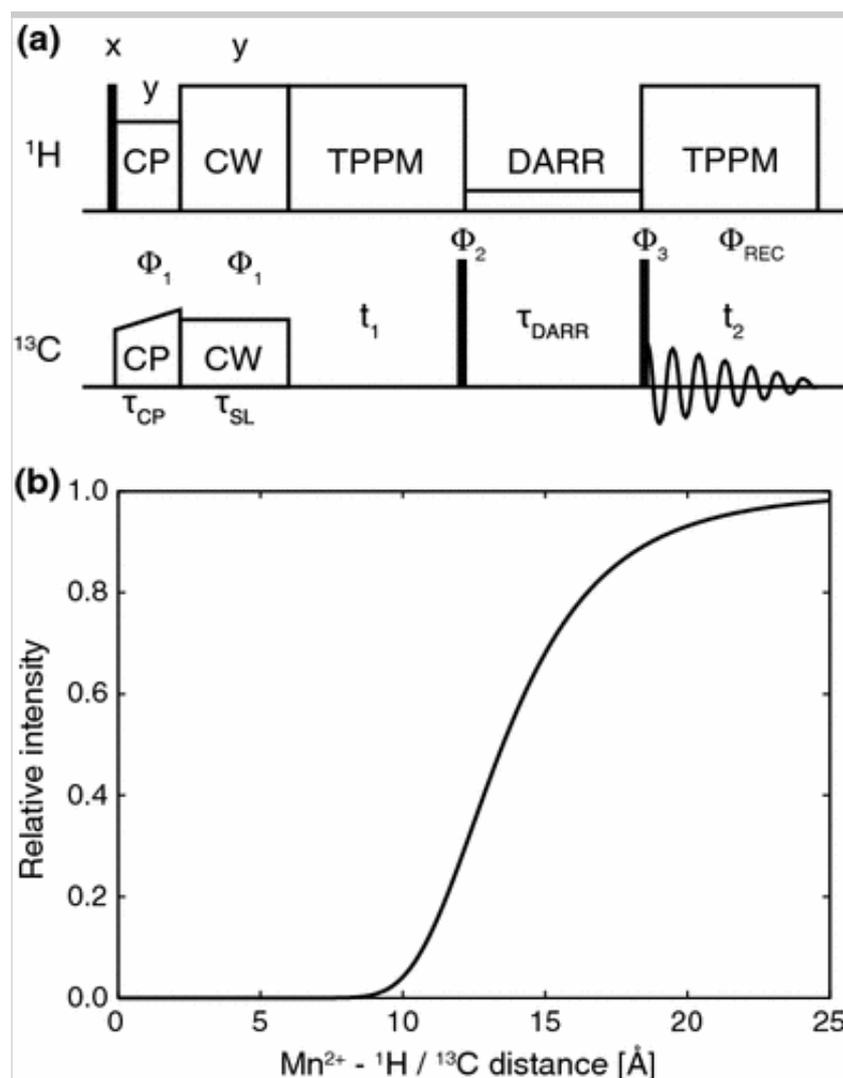
Paramagnetic signal decay

PRE restraints were obtained using the pulse sequence shown in Fig. 1 a. The ^{13}C - ^{13}C spectra using short DARR mixing are widely used to obtain 1- or 2-

bond correlations. We appended a ^{13}C spin-locking period to enhance the signal decay by the longitudinal relaxation rate in the rotating frame. During the ^{13}C spin-locking period, the PRE peak intensity decay can be encoded without restriction of the fast longitudinal relaxation in the rotating frame of ^1H and spectral complexity can be increased using long-time ^{13}C mixing.

Fig. 1

a The ^{13}C - ^{13}C correlation pulse sequence. The *filled rectangles* represent $\pi/2$ pulses. The following phase cycling was used: $\Phi_1 = (-y, y)$, $\Phi_2 = (-x, -x, x, x)$, $\Phi_3 = (x, x, x, x, -x, -x, -x, -x)$ and $\Phi_{\text{REC}} = (-x, x, x, -x, x, -x, -x, x)$. Quadrature phase detection in t_1 is accomplished using the States scheme by $\pi/2$ phase shifting of Φ_1 . **b** The signal decay profile of the pulse sequence by PRE of Mn^{2+} ($s = 5/2$, $T_{1e} = 10$ ns) calculated using Eq. (5). τ_{CP} , τ_{SL} and τ_{DARR} in Eq. (2) were set to 0.32, 1.04 and 20.0 ms, respectively. The *intrinsic line width* was set to 75 Hz



The restraints were obtained from the peak intensity ratios of paramagnetic- and diamagnetic-labeled samples. Thus, relaxation during all periods of the pulse sequence was considered.

The paramagnetic effect of ^{13}C magnetization growth during the cross polarization (CP, Pines et al. 1973) period was estimated by a kinetic model. The ^{13}C magnetization, I_{CP} is given as:

$$I_{CP}(t) = \frac{\exp[\lambda_1(\Gamma_{1\rho,H}, \Gamma_{1\rho,C})t] + \exp[\lambda_2(\Gamma_{1\rho,H}, \Gamma_{1\rho,C})t]}{1 - \exp(-2R_{IS}t)} \quad 1$$

where t is the contact time. λ_1 and λ_2 are parameters dependent on the paramagnetic longitudinal relaxation rate in the rotating frame of ^1H and ^{13}C (denoted as $\Gamma_{1\rho,H}$ and $\Gamma_{1\rho,C}$, respectively). R_{IS} is a rate constant of magnetization transfer between ^1H and ^{13}C spins. The details of the kinetic model derivation and formulae of $\lambda_{1,2}$ are described in the SI text.

In the spin-locking and DARR-mixing periods ^{13}C magnetization is reduced exponentially by the unpaired electron. The reductions I_{SL} and I_{DARR} are expressed as:

$$I_{SL}(t) = \exp(-\Gamma_{1\rho,C}t) \quad 2$$

$$I_{DARR}(t) = \exp(-\Gamma_{1,C}t) \quad 3$$

$\Gamma_{1,C}$ is the paramagnetic longitudinal relaxation rate of ^{13}C , and t is the spin-locking and DARR-mixing times.

In the evolution and detection periods, the signal intensity (peak height) H is proportional to $H \propto 1/R$, where R is a transverse relaxation rate. Thus, the signal intensity of diamagnetic and paramagnetic samples in the direct and indirect detected dimensions are proportional to $H_{dia} \propto 1/R_{2,dia}$ and $H_{para} \propto 1/R_{2,para}$, respectively. $R_{2,dia}$ is the intrinsic transverse relaxation rate of the ^{13}C spin and is estimated from the line width at half-height (w_{dia}) using the equation, $R_{2,dia} = \pi w_{dia}$. The transverse rate of the paramagnetic sample, $R_{2,para}$ is expressed as a sum of the intrinsic ($R_{2,dia}$) and paramagnetic contributions

$(\Gamma_{2,C})$, $R_{2,para} = R_{2,dia} + \Gamma_{2,C}$ (Battiste and Wagner 2000). Thus, the peak intensity reduction during the evolution and detection periods, I_{t_1} and I_{t_2} are expressed as:

$$I_{t_1} = I_{t_2} = \frac{H_{para}}{H_{dia}} = \frac{1/(R_{2,dia} + \Gamma_{2,C})}{1/R_{2,dia}} = \frac{\pi w_{dia}}{\pi w_{dia} + \Gamma_{2,C}} \quad 4$$

Finally, the total signal decay, $I = I_{para}/I_{dia}$ is expressed using Eq. (1), (2), (3) and (4) as:

$$I = \frac{I_{para}}{I_{dia}} = I_{CP}(\tau_{CP}) \cdot I_{SL}(\tau_{SL}) \cdot I_{t_1} \cdot I_{DARR}(\tau_{DARR}) \cdot I_{t_2} \quad 5$$

The paramagnetic relaxation rates, Γ_1 , $\Gamma_{1\rho}$, and Γ_2 are given by the Solomon–Bloembergen equation (Bloembergen 1950; Solomon 1955; Koenig 1982).

$$\Gamma_1 = \frac{2}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_n^2 g_e^2 \beta^2 s(s+1)}{r^6} \left(\frac{3\tau_{c1}}{1 + \omega_n^2 \tau_{c1}^2} + \frac{7\tau_{c2}}{1 + \omega_e^2 \tau_{c2}^2} \right) \quad 6$$

$$\Gamma_{1\rho} \sim \Gamma_2 = \frac{1}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_n^2 g_e^2 \beta^2 s(s+1)}{r^6} \times \left(4\tau_{c1} + \frac{3\tau_{c1}}{1 + \omega_n^2 \tau_{c1}^2} + \frac{13\tau_{c2}}{1 + \omega_e^2 \tau_{c2}^2} \right) \quad 7$$

where μ_0 is the permeability of free space, γ_n is the nuclear gyromagnetic ratio, g_e is the electron g-value, β is the Bohr magneton, S is the electron spin quantum number, r is the unpaired electron-nucleus distance, and ω_n and ω_e are the Larmor frequencies of the nucleus and electron, respectively. The correlation times, τ_{c1} and τ_{c2} were approximated to the longitudinal electron spin relaxation time constant T_{1e} (Nadaud et al. 2009).

The peak intensity decay profile is dependent on the unpaired electron-nucleus distance as shown in Fig. 1 b. τ_{CP} , τ_{SL} , and τ_{DARR} are set to 0.32, 1.04, and 20.0 ms, respectively. The paramagnetic parameters, s and T_{1e} are set to 5/2 and 10 ns, respectively, to evaluate the effect from Mn^{2+} used in our experiments ((Bertini et al. 2001), see “Materials and methods”). w_{dia} is set to 75 Hz (ca. 0.5 ppm at 14.1 T). At $r = 10 \text{ \AA}$, the reductions, I_{CP} , I_{SL} , I_{DARR} , and $I_{t_1} \cdot I_{t_2}$ are

0.45, 0.68, 0.89, and 0.15, respectively. Thus, the PRE signal decay occurs primarily during the direct and indirect detection periods. We performed the PRE experiments for a series of spin-locking time (0–2.08 ms), however the obvious spectral changes did not appear (Fig. S2). The value of I_{SL} is 0.47 at the condition of $r = 10 \text{ \AA}$ and $\tau_{SL} = 2.08 \text{ ms}$, and it is greater than that of $I_{t_1} \cdot I_{t_2}$. The values of I_{SL} and $I_{t_1} \cdot I_{t_2}$ support our experimental results. Under our experimental condition, the contribution of the decay during the ^{13}C spin-locking is limited because spin-locking time is short, which prevents R.F. sample heating. For this purpose, the RF ^1H decoupling can be replaced with high-speed MAS. The value of I_{SL} is 0.11 at the condition of $r = 10 \text{ \AA}$ and $\tau_{SL} = 6.0 \text{ ms}$, thus the I_{SL} decay becomes dominant with the ^{13}C spin-locking time longer than 6.0 ms. The spin-locking time can control the distance that the effective PRE signal decay can be observed. For example, at $r = 15 \text{ \AA}$ and $\tau_{SL} = 1.04 \text{ ms}$, value of the total decay, I is 0.68. By changing the τ_{SL} to 20 ms, the value of I decreases to 0.37. Thus, more long-range distance restraints can be obtained.

Mn^{2+} also causes paramagnetic anisotropic shift dispersion in immobile solid-state samples because of its large electron spin quantum number and decreases the signal intensities because the efficiency of CP is reduced in addition to PRE (Wickramasinghe and Ishii 2006). We evaluated the influence of paramagnetic anisotropic shift dispersion by spin dynamics simulations, and the results are included in our experimental analysis (see SI text for further detail).

Additionally, our findings are transferable to the case of using nitroxide spin labels as the paramagnetic probe, because the constant part of the Solomon-Bloembergen equation of the nitroxide spin label [$s = 1/2$ and $T_{1e} \approx 100\text{ns}$ (Kosen 1989; Bertini et al. 2001)] is similar to that of Mn^{2+} . Indeed, the values of I at $r = 15 \text{ \AA}$ are 0.68 and 0.73 when using Mn^{2+} and nitroxide spin labels, respectively, under our experimental conditions. We adopted Mn^{2+} because paramagnetic transition metal ions have more stable unpaired electrons than nitroxide radicals.

Materials and methods

Numerical simulation of intra-/intermolecular PREs

The intra-/intermolecular PREs from Mn^{2+} were calculated by the Monte Carlo method using the X-ray structure of GB1 (PDB ID: 2GI9). The coordinates of 14 crystal neighbors were calculated using its symmetric entries. The peak intensity ratio of paramagnetic to diamagnetic labeled samples in the pulse sequence used in our experiments (Fig. 1 a), I_{whole} , is given by:

$$I_{\text{whole},i} = \frac{1}{N} \sum_{\mathbf{p} \in \text{pattern}} I(\Gamma'_{1,C}(i, \mathbf{p}), \Gamma'_{1\rho,H}(i, \mathbf{p}), \Gamma'_{1\rho,C}(i, \mathbf{p}), \Gamma'_{2,C}(i, \mathbf{p})) \quad 8$$

where i is the residue index ($i \in 1, \dots, 56$), pattern is the group of paramagnetic labeling pattern vectors ($\mathbf{p} = [p_A, p_B, p_C, \dots, p_O]$, p_A, \dots, p_O are the labeling states of the each chain) derived from the paramagnetic molar ratio. The labeling state of the central chain, p_A was always set to 1 to calculate intramolecular PREs. p_k ($k \in B, \dots, O$) is stochastically given by

$$p_k = \begin{cases} 1, & q \geq s \\ 0, & \text{otherwise} \end{cases} \quad 9$$

where q is the paramagnetic molar ratio and s is a uniformly distributed random number with an interval $[0, 1]$. N is the number of the labeling pattern vectors. I is the signal decay defined by Eq. (5). $\Gamma'_{1,X}$, $\Gamma'_{1\rho,X}$ and $\Gamma'_{2,X}$ are apparent $\Gamma_{1,X}$, $\Gamma_{1\rho,X}$ and $\Gamma_{2,X}$ of X-nuclei ($X = {}^1\text{H}, {}^{13}\text{C}$), respectively. These three terms can be calculated from the labeling pattern and distances between $\text{C}\alpha_i$ atoms in chain A (the central monomer) and $\text{C}\beta$ atoms of paramagnetic-labeled residues (pseudo paramagnetic center) as follows:

$$\Gamma'(i, \mathbf{p}) = \sum_{j \in A, \dots, O} p_j \Gamma(r_{ij}) \quad 10$$

where j is the chain index of the PDB data ($j \in A, \dots, O$), Γ is the paramagnetic relaxation rates given by Eqs. (6) and (7), and r_{ij} is the distance between a $\text{C}\alpha_i$ atom in chain A and the pseudo paramagnetic center in chain j .

One thousand labeling pattern vectors were prepared for each molar ratio. The calculation was independently performed 100 times and the convergence of the results was confirmed (the maximum deviation of the signal intensities was ca. 0.016). N8, E19, and T53 were selected as the paramagnetic labeled residues (they are referred to as GB1-N8 Mn^{2+} , GB1-E19 Mn^{2+} , and GB1-T53 Mn^{2+}). The

other parameters were similarly set to describe “Paramagnetic signal decay”.

Protein expression and purification

The cDNA of GB1-T2Q/F52Y (referred to as GB1) was introduced into a pET-22b(+) vector. Three single cysteine mutants N8C, E19C, and T53C were constructed using quick-change site-directed mutagenesis. Natural abundance GB1-wt was expressed in *E. coli* BL21(DE3) in LB medium and the U-¹³C, ¹⁵N single cysteine mutants were expressed in M9 minimal medium containing 1 g/L ¹⁵NH₄Cl and 3 g/L ¹³C-glucose. The unlabeled GB1-wt and U-¹³C, ¹⁵N single cysteine mutants were purified using the same protocol, except the buffers used for the purification of the single cysteine mutants contained 5 mM DTT. The cell suspension was incubated in water for 10 min at 80 °C and the supernatant was loaded onto a HiTrap DEAE FF column (Li et al. 2013). The purities of the samples were checked by SDS-PAGE.

Ligation of EDTA-metal complex to single cysteine mutants

Ligation of the EDTA-metal (EDTA-M, M = Mn²⁺, Zn²⁺) complex to single cysteine mutants was performed as described previously (Nadaud et al. 2009). Solutions of the single cysteine mutants were eluted through a PD-10 desalting column equilibrated with 50 mM sodium phosphate, pH 6.5. Immediately thereafter, the proteins were incubated overnight at 4 °C with a five-fold molar excess of *N*-[S-(2-pyridylthio)cysteaminy]EDTA, delivered as an aqueous solution and preloaded with 1.1 mol equiv. of Mn²⁺ or Zn²⁺ using MnCl₂ or ZnCl₂. Excess *N*-[S-(2-pyridylthio)cysteaminy]EDTA-Mn²⁺/Zn²⁺ was removed by a PD-10 desalting column equilibrated with 50 mM sodium phosphate, pH 5.5. MALDI-TOF MS was used to confirm the incorporation of EDTA-Mn²⁺/Zn²⁺ side chains.

Microcrystallization of GB1 samples

Microcrystals of GB1 were prepared as described previously (Franks et al. 2005; Nadaud et al. 2009). A solution of unlabeled GB1 was dialyzed with 50 mM sodium phosphate, pH 5.5. Protein solutions were concentrated to 30 mg/mL. To suppress intermolecular PREs, the single cysteine mutants were mixed with three-fold molar excess of unlabeled GB1-wt. Those solutions were mixed with a three-fold volume of precipitation solution (2-methyl-2,4-

pentanediol:2-propanol = 2:1, v/v) three times, and then incubated for 3 days at 18 °C. Finally, the precipitant was collected by centrifugation at 2000g for 10 min and packed into 3.2 mm Varian standard-wall and JEOL RESONANCE zirconia rotors.

Solid-state NMR spectroscopy

Solid-state NMR experiments were performed with Varian Infinity-plus 600 MHz and JEOL RESONANCE ECA 600 MHz II spectrometers equipped with MAS probes for 3.2-mm rotors. The sample spinning frequency was 12.5 kHz at a probe temperature of −10 °C. The chemical shifts were indirectly referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid by adjusting the position of the ¹³C adamantane downfield peak to 40.49 ppm (Morcombe and Zilm 2003). The pulse sequence used in the experiments is shown in Fig. 1 a. The $\pi/2$ pulse widths were 3.2 μ s for ¹H and ¹³C. The ¹H–¹³C CP employed 320 μ s contact time at 50 kHz ¹H RF field and with the ¹³C lock field ramped linearly around the $n = 1$ Hartmann–Hahn condition (ramp gradients were 12.5 kHz). The contact time was determined experimentally to maximize the aliphatic carbon signal intensities. To enhance the contrast in peak intensities because of the paramagnetic interaction, 37.5 kHz and 1.04 ms of ¹³C spin locking was performed under 78 kHz of CW ¹H decoupling followed by a CP period. ¹³C–¹³C short-range mixing was performed for 20 ms under the $n = 1$ DARR condition (Takegoshi et al. 2001). TPPM decoupling (Bennett et al. 1995) was performed for t_1 and t_2 periods at a ¹H RF field of 78 kHz. Maximal t_1 and t_2 periods were 10 ms. All spectra were processed by NMRPipe (Delaglio et al. 1995) and analyzed in Sparky (Goddard and Kneller 2015).

Structure calculation of GB1

PRE structural restraints were derived from C α –C' cross peak intensities of the DARR spectra. Strong PREs with relative intensities smaller than 0.4 to the diamagnetic reference were converted into distances between the C α atom of the PRE affected residue and the C β atom of the EDTA-M attached residue with an upper bound of 16.0 Å, otherwise, the distance restraints were set with a lower bound of 10.0 Å. C'–C α cross peaks were not used to obtain restraints because the signal-to-noise ratio was low due to the short CP. A total of 43 upper- and 60 lower-bound restraints were obtained from N8, E19 and

T53EDTA-M samples. For CS-Rosetta structure calculations, a non-homologous fragment library was generated using the amino acid sequence of GB1 and its backbone chemical shifts deposited in the BMRB database (entry 15156). Fragment picking was performed by CS-Rosetta toolbox version 3 (www.csrosetta.org). The ^1H and ^{15}N chemical shifts were ignored. The $\text{C}\alpha$, $\text{C}\beta$ and C' chemical shifts of Y52 were calculated from chemical shifts of F52 (BMRB entry) and random coil shifts of phenylalanine and tyrosine (Wishart et al. 1995). We performed low-resolution (centroid) modeling with and without PRE restraints. For the unrestrained calculation, the modeling protocol followed the default protocol of *AbinitioRelax* in Rosetta 3.5 (Leaver-Fay et al. 2011). For the restrained calculation, the protocol was the same as the unrestrained calculation, except that the upper and lower bound PRE restraints were combined. The upper and lower PRE restraints, E_{ub} and E_{lb} were expressed by:

$$E_{ub} = \begin{cases} 0 & \text{if } r \leq r_{ub} \\ (r - r_{ub})^2 & \text{if } r_{ub} < r \end{cases} \quad 11$$

$$E_{lb} = \begin{cases} (r - r_{lb})^2 & \text{if } r \leq r_{lb} \\ 0 & \text{if } r_{lb} < r \end{cases} \quad 12$$

where r is the distance between the $\text{C}\alpha$ atom of a PRE affected residue and the $\text{C}\beta$ atom of the EDTA-M modified residue, and r_{ub} and r_{lb} are set to 16.0 and 10.0 Å, respectively. A weighting factor of the PRE restraints was set to 10. 5000 structures were calculated. The full-atom refinements and re-scoring by chemical shifts (Shen et al. 2008) were not performed.

Conventional structural calculations based on restrained molecular dynamics and simulated annealing were performed using Xplor-NIH (Schwieters et al. 2003). PRE and dihedral restraints were used. The PRE restraints were the same restraints used by the CS-Rosetta calculation. Dihedral restraints were obtained from TALOS+ (Shen et al. 2009) using the backbone chemical shifts employed for the fragment picking of CS-Rosetta calculations. For predicted dihedral angles with uncertainties of less than $\pm 20^\circ$, the uncertainties were set to $\pm 20^\circ$ for the structure calculations (Sengupta et al. 2012). Dihedral angles that were not classified as *good* were ignored. A thousand structures were calculated.

Results

Influence of the intermolecular PREs

Severe intermolecular PREs caused by dense molecular packing in the solid-state prevent structure determination. The dilution of paramagnetic labeled molecules by natural abundance and diamagnetic molecules reduces intermolecular PREs; however, it remains impossible to suppress all intermolecular PREs. Consequently, the accuracy of the apparent intramolecular PRE is reduced. Therefore, evaluation of intermolecular PREs is highly important for determining the optimum paramagnetic molar ratio and the shape of the restraint function. Jaroniec and co-workers estimated the influence of the intermolecular PREs for analysis of Cu^{2+} - ^{15}N couplings from measurements of various dilution ratio samples, and concluded that: (1) intermolecular PREs are effectively quenched at ca. 0.10–0.14 paramagnetic molar ratios; and (2) can be avoided by using purely repulsive restraints for atoms distal from the paramagnetic center in relatively low dilution ratio conditions (Nadaud et al. 2011; Sengupta et al. 2012). Using Mn^{2+} as the paramagnetic center, intermolecular PREs are more severely affected, because the electron spin quantum number of Mn^{2+} is five times larger than Cu^{2+} . The influence of intermolecular PREs from Mn^{2+} was assessed by numerical calculations using the X-ray structure of GB1 and its crystal neighbors (Fig. 2a). Minimal and maximal distances between $\text{C}\alpha$ atoms in a central chain and the pseudo-paramagnetic center in other chains are summarized in Table S1. The minimum of the maximal distances is 27.6 Å, and the signal decay by PRE cannot be observed in our model. We concluded that further increasing the size of the crystal-packing model was insignificant.

Fig. 2

a Crystal packing of GB1 used in the intermolecular PRE simulation (PDB ID: 2GI9). The central monomer used in the intramolecular PRE calculation is colored orange. The *red-* and *blue-colored spheres* indicate $\text{C}\beta$ atoms of E19 used in the intramolecular and intermolecular PREs, respectively. **b** Maximal $I_{\text{para}}/I_{\text{dia}}$ differences between data calculated in the absence and presence of intermolecular PREs with various paramagnetic molar ratios

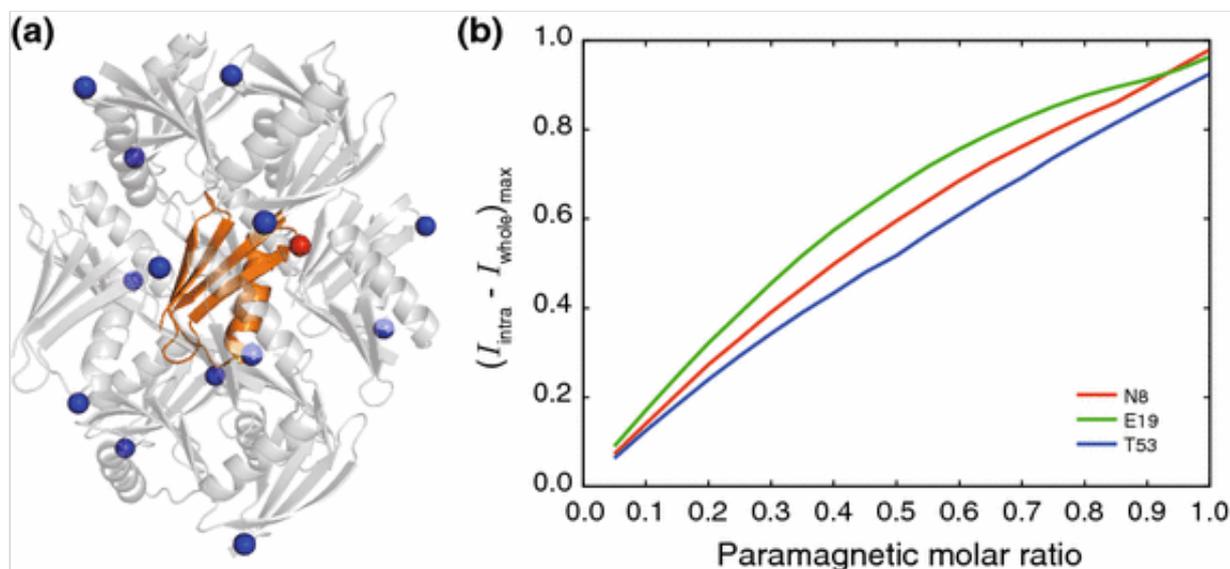
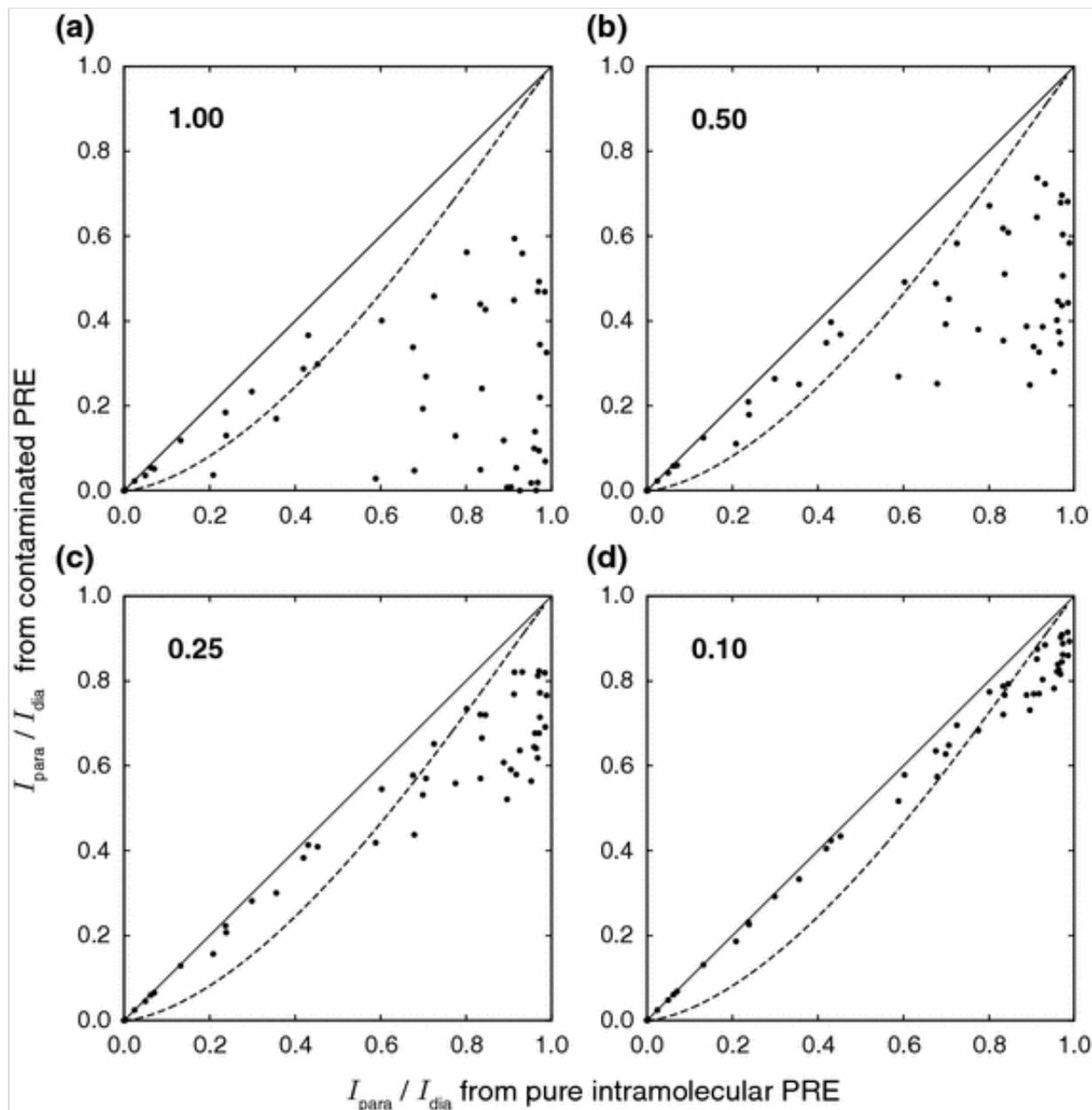


Figure 2b shows the largest influence of the intermolecular PRE of GB1-N8, E19 and T53Mn²⁺ for each paramagnetic molar ratio. Clearly, correlations of the no-dilution condition are largely affected by intermolecular PREs and decreasing the paramagnetic molar ratio reduces these intermolecular PREs in an essentially linear manner. The trends are similar between the three paramagnetic centers, suggesting that the influence of the intermolecular PREs is nearly independent of the paramagnetic labeling site for GB1. We also tested intermolecular PREs using GB1-K28, E42 and D46Mn²⁺, and obtained similar results (Fig. S4). Figure 3 shows the correlation of paramagnetic/diamagnetic peak intensity ratios of GB1-E19Mn²⁺ between purely intramolecular PREs and PREs contaminated by intermolecular PREs at various dilution ratios. Under conditions of 1.0 and 0.5 paramagnetic molar ratios, many residues have errors of >1 Å from the pure intermolecular PREs. Under conditions where the molar ratio is <0.25, residues whose peak intensity ratio from contaminated PRE is <0.4 (corresponding to ca. 13 Å, as estimated by the signal decay model (Fig. 1b) do not have an error exceeding 1 Å. This error range is smaller than the error caused by other factors (for example, the average standard deviation of the C α atoms to the Mn²⁺ distance of the GB1-E19EDTA-Mn²⁺ sample in a MD simulation trajectory is 2.6 Å. See SI, Fig. S11 and Table S5) and thus, intermolecular PREs can be ignored when the ratio is in this range. Residues that give rise to a peak intensity ratio from a pure intramolecular PRE more than 0.4 are highly affected by intermolecular PREs when the molar ratio is 0.25. When the molar ratio is 0.1, residues whose peak intensity ratio from contaminated intermolecular PRE is <0.7 (corresponding to ca. 15 Å) do not

have an error $>1 \text{ \AA}$. However, from the property of the peak intensity ratio changes depends on distance between the nucleus and paramagnetic center, the estimation of this distance for the residues whose peak intensity ratios are high (>0.6 – 0.7) is susceptible to error owing to various factors (e.g., tag flexibility and spectral noise). Additionally, similar results were obtained for GB1-N8Mn²⁺ and GB1-T53Mn²⁺ (Fig. S5 and S6). Finally, we concluded that a paramagnetic molar ratio of 0.25 was appropriate and struck a balance between sensitivity and the residual influence of intermolecular PREs.

Fig. 3

Comparison of $I_{\text{para}}/I_{\text{dia}}$ of GB1-E19Mn²⁺ in the absence and presence of intermolecular PREs for paramagnetic molar ratios of **a** 1.0, **b** 0.5, **c** 0.25 and **d** 0.1. The *dashed lines* indicate an error of 1 \AA between the pure intramolecular and contaminated PREs

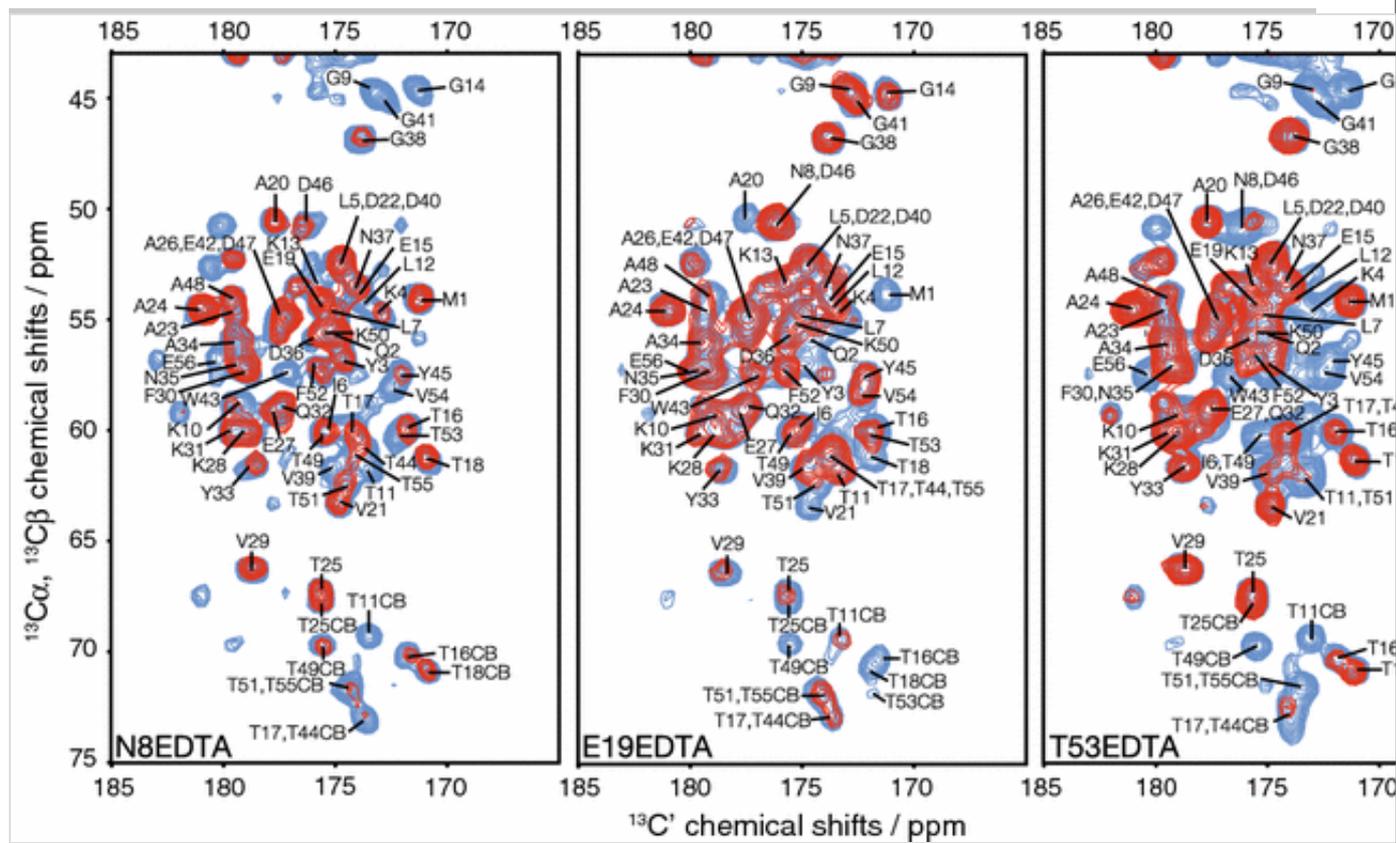


Structural restraints from PREs

2D ^{13}C - ^{13}C spectra of GB1-N8, E19 and T53EDTA-M are shown in Fig. 4. Since the chemical shifts of the resonances are in close agreement with spectral data for GB1-wt, the EDTA-M label does not significantly affect the structure (Fig. S7). Spectra of these labeled mutants clearly showed that Mn^{2+} bleached out signals arising from some residues. This is because Mn^{2+} has a large electron spin quantum number and a single electron relaxation time. Resonance line broadening was not observed for the visible peaks of the Mn^{2+} attached samples.

Fig. 4

$\text{C}\alpha\text{-C}'$ regions of 2D-DARR spectra of N8, E19 and T53EDTA-MGB1. Signals from Mn^{2+} and Zn^{2+} attached samples are shown in *red* and *blue*, respectively



The absence of particular peaks reflects the GB1 fold. For example, the resonance representing M1 of E19-EDTA- Mn^{2+} was absent, whereas in the N8 and T53-EDTA- Mn^{2+} samples, the resonance of M1 was observed. The M1 $\text{C}\alpha\text{-N8C}\beta$, -E19 $\text{C}\beta$ and -T53 $\text{C}\beta$ distances in the X-ray structure (PDB ID: 2GI9) are 23.3, 5.7 and 17.9 Å, respectively (Fig. S8(a)). The resonance of G9 in the N8 and T53-EDTA- Mn^{2+} spectra was missing, whereas in the spectrum of the E19-EDTA- Mn^{2+} sample, the resonance of G9 was observed. The G9 $\text{C}\alpha\text{-N8C}\beta$, -E10 $\text{C}\beta$ and -T53 $\text{C}\beta$ distances in the X-ray structure are 4.9, 23.4 and 9.0 Å, respectively (Fig. S8(b)).

The effect of intermolecular PREs was assessed by examining the peak intensities of G38 and G41 of the GB1-T53EDTA- Mn^{2+} sample as probes to confirm the results of the numerical simulations (see “Influence of the intermolecular PREs”). The peak intensity of G38 is strong, whereas the peak intensity of G41 is very weak. This observation indicates the distance between G41 $\text{C}\alpha$ and the tagged Mn^{2+} at position 53 is closer to the PRE probe than

G38C α . The intramolecular T53C β -G38C α and T53C β -G41C α distances in the GB1 microcrystal structure determined by MAS solid-state NMR (PDB ID: 2KWD, chain A (Nieuwkoop and Rienstra 2010)) are 15.3 and 9.9 Å, respectively, thereby supporting the PRE observation. However, the intermolecular T53C β -G38C α and T53C β -G41C α distances of the neighboring monomer (chain B) are 10.7 and 16.8 Å, respectively (Fig. S8(c)). If the influence of intermolecular PRE were not negligible, both peaks would be completely bleached. Therefore, we concluded that the influence of intermolecular PREs is negligible by diluting the Mn²⁺ attached protein with three molar-fold of non-labeled protein.

Figure 5 shows the relative cross-peak intensities in EDTA-Mn²⁺/EDTA-Zn²⁺ spectra as a function of residue number (a, c, e) and location within the tertiary protein structure for N8, E19 and T53EDTA-M (b, d, f). The peak intensity modulations appear to correlate with the distance between the corresponding residues and the EDTA-M attached residue. In Fig. 6, we compared the observed peak intensities with the calculated values using the X-ray structure and Eq. (5). Clearly, the observed decay in peak intensities from PRE has a large error and cannot be treated as quantitative distance restraints. Taking this into consideration, we applied purely attractive restraints as defined by Eq. (11) for the strong PREs observed and repulsive restraints defined by Eq. (12) for the weak PREs observed. The strong PRE (signal decay is <0.4, corresponds to ca. 13 Å from the paramagnetic center) observed residues have a moderate upper bound error (~3 Å). The other residues have large upper and lower bound errors. However, the signal decay profile given by Eq. (5) suggests the weak PRE observed residues are ca. 10 Å distal from the paramagnetic center (corresponding to a signal decay of ca. 0.04). Therefore, using attractive and repulsive restraints can suitably handle the low-precision and qualitative PRE information. Additionally, this approach avoids issues associated with intermolecular PREs. Strong PREs with relative intensities smaller than 0.4 were then converted into distances between C α atoms of PRE affected residues and the pseudo-paramagnetic center with an upper bound of 16.0 Å, otherwise, the distance restraints were set with a lower bound of 10.0 Å (because CS-Rosetta cannot treat the EDTA-M attached residues, the influence of this approximation is described in SI). A total of 40 upper and 65 lower bound restraints were obtained from N8, E19 and T53EDTA-M samples (Table 1).

Comparisons of the restraints and distances between C α atoms to the pseudo-paramagnetic center on the X-ray structure are shown in Fig. S11. The restraints agree fully with the distances between the C α atoms to the pseudo-paramagnetic center.

Fig. 5

a, c, e C α -C' peak intensity ratios of N8, E19 and T53EDTA-Mn²⁺ to Zn²⁺, respectively. The *orange bars* indicate EDTA-M attached residues. The *gray bars* indicate overlapped peaks. **b, d, e** Crystal structure of GB1 (PDB ID: 2GI9). The peak intensity ratios of *red colored* residues are >0.4, whereas those of *blue colored* residues are <0.4

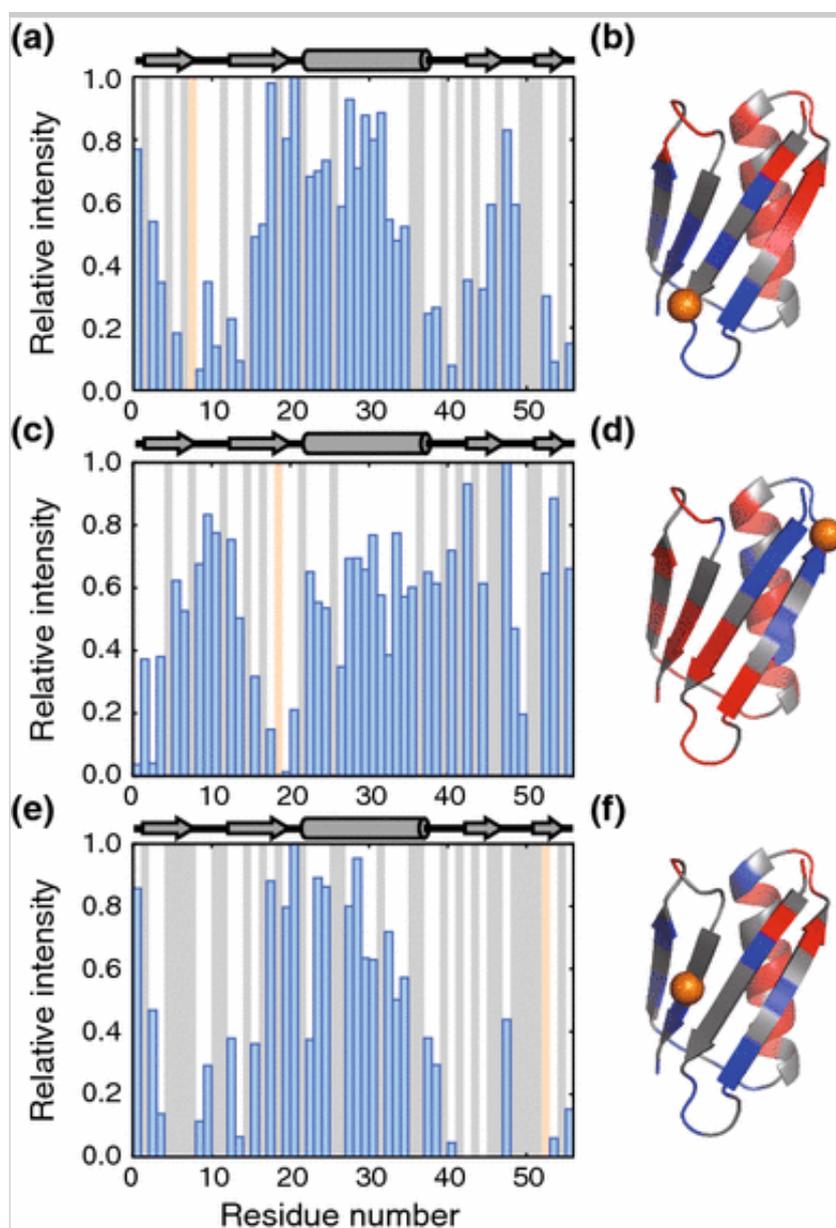
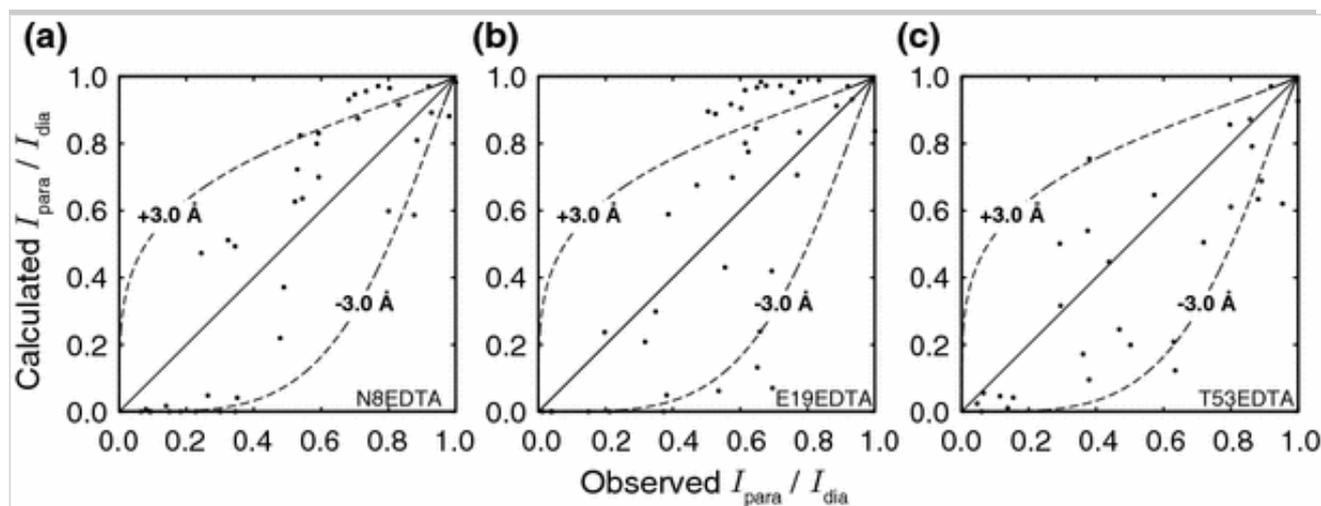


Fig. 6

Comparison of $I_{\text{para}}/I_{\text{dia}}$ of **a** GB1-N8EDTAMn²⁺, **b** GB1-E19EDTAMn²⁺ and **c** GB1-T53EDTAMn²⁺ to the calculated values from the C α -pseudo paramagnetic center distances in the X-ray structure (PDB: 2GI9) using Eq. (5). The *dashed lines* indicate the ± 3 Å error threshold between observed and calculated signal decays. The same plots with *error bars* are illustrated in Fig. S9

**Table 1**

Summary of the CS-Rosetta structure calculation of GB1 with and without PRE restraints

	With PRE	Without PRE
PRE restraints		
N8EDTA-M		
Upper-bound	15	
Lower-bound	22	
E19EDTA-M		
Upper-bound	11	
Lower-bound	28	
T53EDTA-M		
Upper-bound	14	
Lower-bound	15	

RMSD of the lowest score (Å)	1.49	1.52
Lowest RMSD of the sampling (Å)	1.14	1.17
Convergence (Å)	1.22	1.74

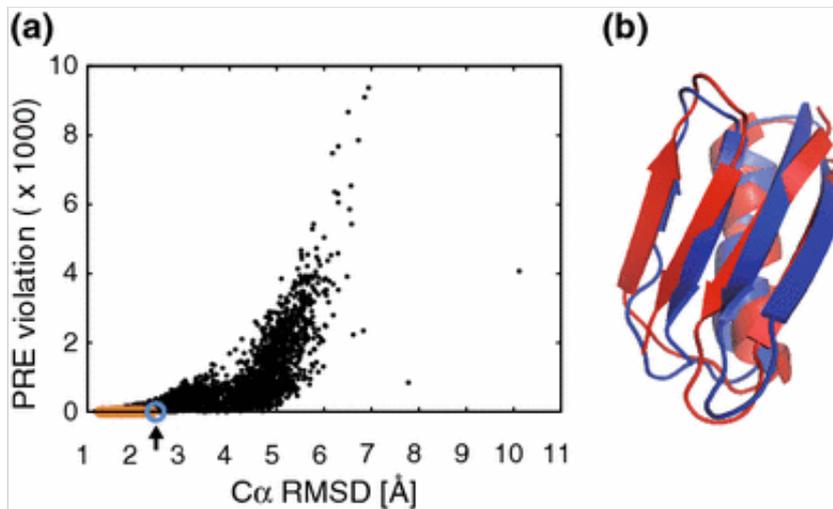
Structure calculations

To evaluate the availability of the qualitative PRE restraints, we performed structure calculations using PREs and dihedral restraints from the backbone chemical shifts based on restrained molecular dynamics and simulated annealing. However, the native structure was not obtained because of structural energy degeneration (Fig. S12). We then performed the CS-Rosetta calculation.

To examine whether the qualitative PRE restraints can identify the protein fold, we performed low-resolution CS-Rosetta calculations without the PRE restraints, and then re-scored with the PRE restraints. Figure 9b presents the energy profiles of the CS-Rosetta calculations. A funnel shape energy profile is observed and the RMSD between the C α atoms of the lowest score structure and the X-ray structure (PDB ID: 2GI9, this RMSD is referred to as C α RMSD) was 1.52 Å. These results indicate the CS-Rosetta calculation was successfully performed. The PRE violation profile of the CS-Rosetta calculation (Fig. 7a) clearly shows a positive correlation between the C α RMSD and the PRE violation. The largest C α RMSD for structures with no PRE violations is 2.4 Å, and the overall structural topology is the same as the native fold (Fig. 7b). This result indicates that qualitative PRE restraints obtained from solid-state NMR measurements can be used to identify protein fold from the CS-Rosetta structure pool.

Fig. 7

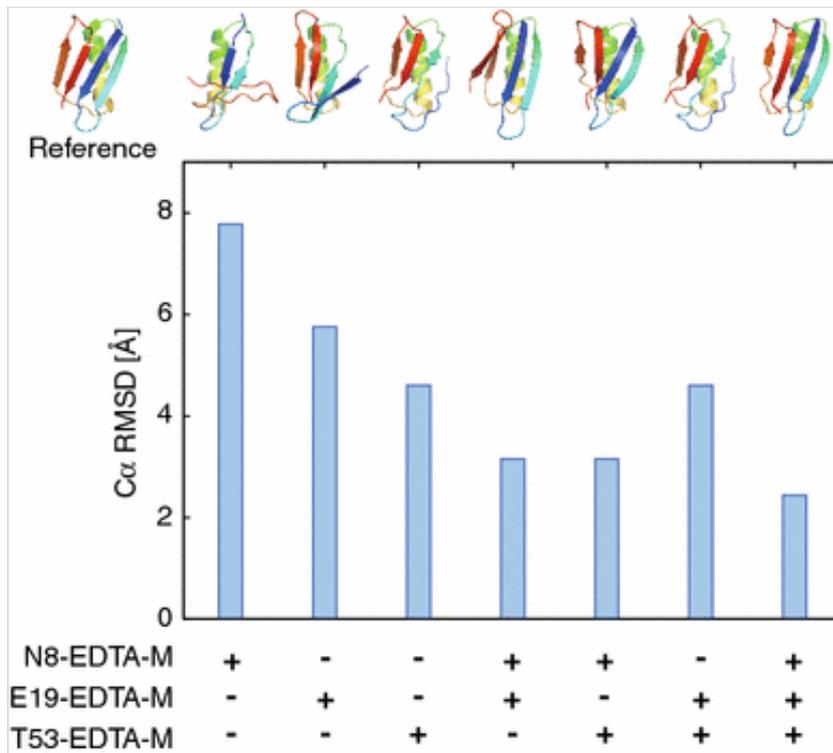
a PRE violation score from three EDTA-M tags versus the C α RMSD to the X-ray structure of GB1 (PDB ID: 2GI9). The structures without any violations are marked with *orange* and *blue colored circles*. The *blue circle* (indicated by an *arrow*) indicates the largest RMSD structure. **b** The X-ray structure is represented in *red*. The structure indicated by the *arrow* in **(a)** is represented in *blue*



To evaluate the contribution of the EDTA-M number and position in determining the correct protein fold, we re-scored the CS-Rosetta structure pool by all combinations of the labeled samples. Figure 8 presents the largest $C\alpha$ RMSD for structures that have no PRE violations. Clearly, the number of EDTA-M labeled samples and the position of the label are important in obtaining the correct fold. For example, the combination of N8 and E19EDTA-M restraints improves the RMSD when compared with using each set of restraints individually. In the case of using the N8EDTA-M restraints, arrangement of the first and second β -sheets was correct. However, when using the T53EDTA-M restraints, arrangement of the third and fourth β -sheets was correct. In the case of using all or, N8 and T53-EDTA-M PRE restraints, the correct protein fold was successfully identified. In our labeling scheme, the N8 and T53EDTA-M labeled samples were required for correct protein fold identification.

Fig. 8

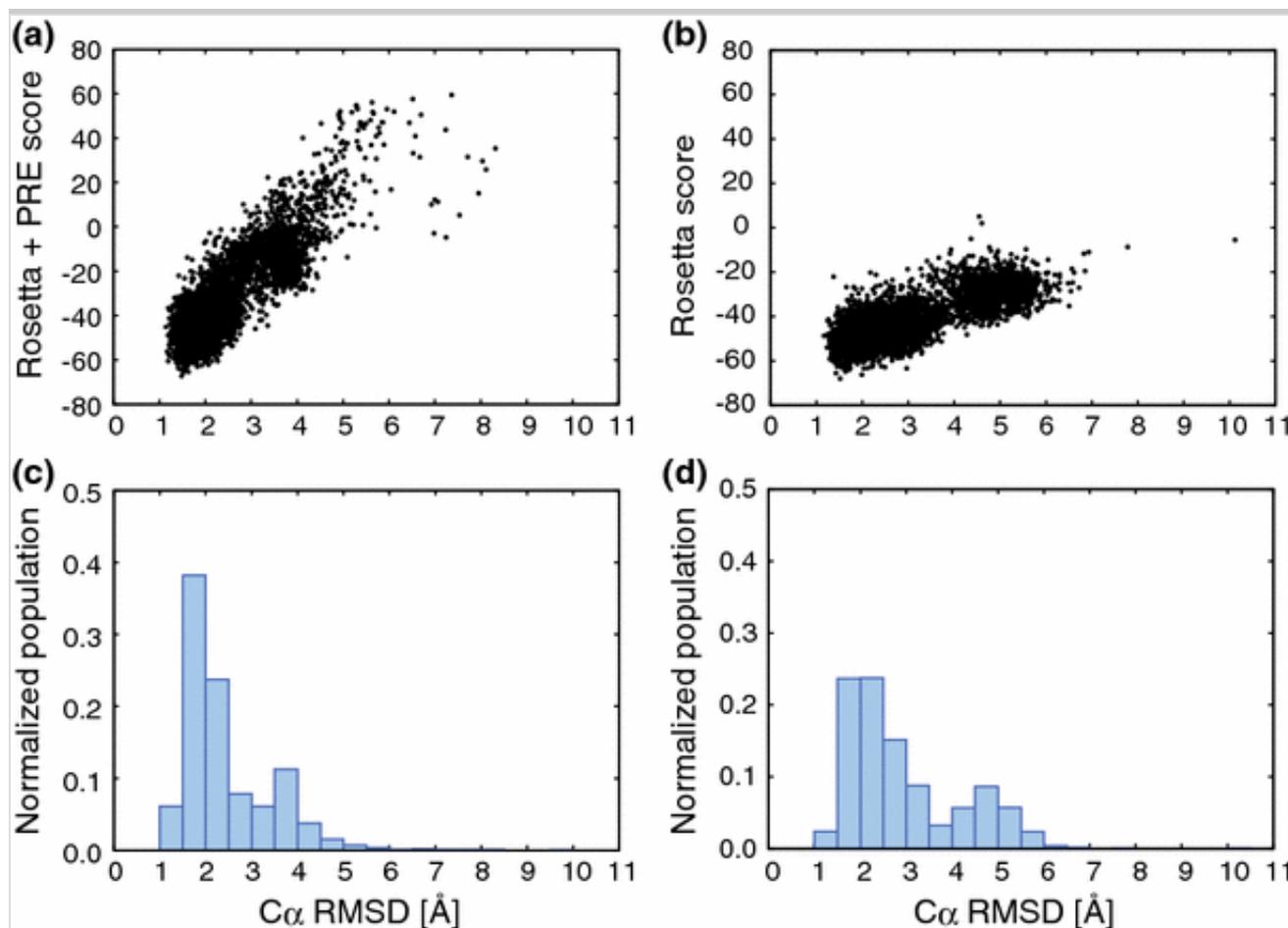
The largest RMSD without any PRE violation and its structure with different combinations of restraints from three EDTA-M tags. The *color map* of the structure indicates the residue number



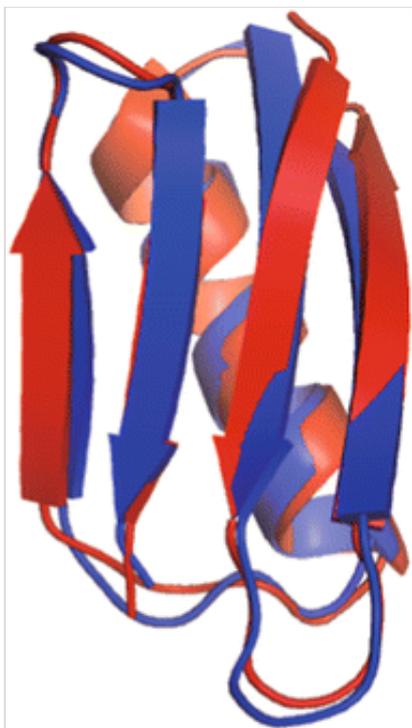
We performed low-resolution CS-Rosetta calculations with the PRE restraints. Figure 9a presents the energy profiles of the CS-Rosetta calculations. Figure 10 shows that the lowest scored structure has the native fold. Additionally, the energy plot funnels towards native-like structures, suggesting that the CS-Rosetta calculations were performed successfully. The PRE restraints improved slightly the $C\alpha$ RMSD of the lowest score structure without PREs from 1.52 to 1.49 Å. Moreover, the convergence (average $C\alpha$ RMSD calculated between the lowest-score and the next four lowest-scoring structures) also improved from 1.74 to 1.22 Å (Table 1). Figure S13 shows the $C\alpha$ RMSD versus PRE violation plot. PRE violations were dramatically reduced when compared with the non-restrained CS-Rosetta calculations. These show clearly that the conformational sampling space was reduced by the PRE restraints. Figure 9c, d show the frequencies of $C\alpha$ RMSD values to the crystal structure along the structure calculation with and without PRE restraints. Using PRE structural restraints, ~44 % of the structures have $C\alpha$ RMSD values of 1–2 Å. However, only 26 % of the structures obtained without PRE structural restraints have $C\alpha$ RMSD values of 1–2 Å. To evaluate the effect of the upper-bound distance restraints value, we performed CS-Rosetta calculations with $r_{ub} = 15$ –18 Å. In all cases, the convergence and population of the low-RMSD structures were improved (Table S2, Fig. S14). This observation illustrates the ability of the PRE restraints to bias sampling towards the native structure.

Fig. 9

a, b Energy landscapes generated by CS-Rosetta with and without PRE restraints, respectively. **c, d** The distribution of the $C\alpha$ RMSD with and without PRE restraints

**Fig. 10**

Result of the GB1 structure calculation. The X-ray structure is presented in *red* and the CS-Rosetta structure with PRE restraints is presented in *blue*



We performed all-atom refinement of the Rosetta protocol with and without the PRE restraints using 10 % of the structures from the lowest energy structure of the low-resolution structures with and without PRE restraints (the details are described in the SI text). The energy profiles and RMSD histograms show that the conformational sampling space of the refinement calculation is also reduced by the PRE restraints (Fig. S15).

Discussion

To use PREs for protein structure determination in solid-state NMR factors such as the dense molecular packing and interference from numerous nuclear dipolar couplings must be considered. Dense molecular packing gives rise to intermolecular PREs that can severely hamper the effective use of PRE data in structure calculations. Numerical simulations using the X-ray structure of GB1 and its crystal packing information indicated that a paramagnetic molar ratio of 0.25 and a combination of purely attractive and repulsive restraints was effective in suppressing the affect of intermolecular PREs when Mn^{2+} was used. This result is very similar to the report of Jaroniec and co-workers using Cu^{2+} (Nadaud et al. 2011). In addition, our results also suggest that the influence is nearly independent of the paramagnetic labeling site. Since GB1 is a small globular protein, the influence of intermolecular PREs for other globular proteins is also expected to be suppressed by the same paramagnetic molar ratio

and using a combination of purely attractive and repulsive restraints when Mn^{2+} , Cu^{2+} , or a nitroxide radical is used. The influence of intermolecular PREs in non-globular molecular systems (e.g., amyloid fibrils and bacterial secretion systems) remains unclear. However, some of those structures have been solved recently and it should be possible to estimate intermolecular PREs using a combination of those structures and numerical calculations.

Following the results of the numerical calculation and low-precision PRE information, we applied purely attractive and repulsive restraints defined by Eqs. (11) and (12) for $\text{C}\alpha$ atoms where strong and weak PREs, respectively, were observed. The PRE restraints used herein were less quantitative than PREs used in previous structural studies by solid-state NMR (Wang et al. 2013; Jaroniec 2015). We found that combinations of qualitative PRE restraints from paramagnetic Mn^{2+} and CS-Rosetta fragment assembly can effectively determine protein folds. Additionally, by combining those restraints and CS-Rosetta centroid modeling, we successfully obtained the native protein fold. We postulate that these successes were provided from the fragment assembly and score function based on the knowledge of protein folding. The fragment assembly method significantly reduces the conformational sampling space when compared with conventional structure calculation methods that are based on restrained molecular dynamics. Insertion of 3- and 9-residue fragments generated from the similarity of backbone chemical shifts and sequential homology, and a score function derived from the knowledge of protein folding reduces the chances of obtaining a structure that has improper protein structural features (Rohl et al. 2004; Shen et al. 2008). This feature is preferable to structure calculations using limited datasets of qualitative restraints. Indeed, the native-like structure of GB1 can be extracted using PRE restraints from the structure pool of the CS-Rosetta calculation without PRE restraints; however, this native structure cannot be obtained by conventional simulated annealing calculations using Xplor-NIH.

In the case of the CS-Rosetta structure calculations with PRE restraints, Rosetta scores of the high $\text{C}\alpha$ RMSD structure were shifted upward and $\text{C}\alpha$ RMSDs of the lowest score structure decreased when compared with non-restrained calculations (Fig. 9). Similar changes have been reported to occur in CS-Rosetta calculations with RDC restraints (Raman et al. 2010). The reason for these changes is because of the concordance of the structural information

defined in the experimental restraints and the detailed physical chemistry of protein folding implicit in the Rosetta score function. This concordance provides two favorable effects for structural optimization. First, optimization far from the native structure is impeded, resulting in an upward shift of the score of non-native structures, and second, optimization near the minimum is improved as the restraints guide the search towards the global minimum. Impeded optimization far from the native structure is also present in an all-atom refinement (see SI text for details). We conclude that qualitative PRE restraints obtained from solid-state NMR are a good conformational sampling guide for CS-Rosetta fragment assembly calculations.

Jaroniec and co-workers have proposed using Cu^{2+} - ^{15}N longitudinal PREs to improve structure determination quantitatively (Nadaud et al. 2009). In addition, they successfully determined the backbone structure of GB1 from those PREs and dihedral angle constraints calculated by TALOS+ (Sengupta et al. 2012). Our approach differs to their approach because qualitative PREs are used, and has a number of strengths. Obtaining qualitative PRE profiles of transverse relaxation and longitudinal relaxation in the rotating frame is more facile than longitudinal PREs. The qualitative PRE profiles can be obtained from peak intensity ratios between paramagnetic and diamagnetic spectra. However, to determine ^{15}N longitudinal relaxation rates, the inversion-recovery method is mandatory for gaining sensitivity by CP. The ^{15}N longitudinal rates of EDTA- Zn^{2+} attached GB1 are 0.03 – 0.2 s^{-1} (Nadaud et al. 2009). Therefore, a few seconds are required per scan and the measurements are time consuming. Our approach can be easily extended to collect PRE profiles of side chains. However, it is difficult to obtain correct longitudinal PREs of side chains because the influence of the spin diffusion is severe for ^1H and ^{13}C . Low precision of the restraints is the obvious drawback of our approach. In general, increasing the number of PRE restraints is the simplest way to overcome this problem; however, this is time consuming because it is necessary to prepare and measure more paramagnetic- and diamagnetic-labeled samples. Our results clearly show that the native fold can be identified from CS-Rosetta structural pools using only two or three site-specific paramagnetic labeled samples. In the study of GB1 backbone structure determination by Cu^{2+} - ^{15}N longitudinal PREs, six labeling samples were used (Sengupta et al. 2012). The drawback of low precision can be overcome by advanced computational techniques.

Additionally, deuteration and/or ^{13}C spin dilution of proteins suppresses the nuclear dipolar couplings and likely improves the precision of PRE restraints. The combination of sophisticated isotope labeling and advanced computational techniques represents an intriguing approach.

The key question remains as to how general and transferable our approach is to other proteins. The size of the target protein is very important because of two reasons: (1) the conformational sampling limitation of the CS-Rosetta protocol (Raman et al. 2010), and (2) increasing the molecular size raises the likely problem of severe peak overlap. The former problem can be solved by the RASREC approach (Lange and Baker 2011). RASREC CS-Rosetta can determine solution structures of proteins up to 40 kDa using limited NOE and RDC restraints (Lange et al. 2012). To consider the latter problem, we performed structure calculations using simulated PRE restraints (the details are described in SI). Our results suggest that the backbone structure of a protein whose size is ca. 110 residues is feasible to determine. For larger proteins or α -helical membrane proteins whose resonance dispersion is lower, three-dimensional or higher dimensional measurements are a promising way to avoid this problem. However, it is not feasible to obtain PRE restraints from transverse relaxation enhancement because the efficiency of magnetization transfers is reduced by PRE. In addition, the sensitivity decreases because of dilution of the paramagnetic labeled sample. The sensitivity enhancement by dynamic nuclear polarization probably compensates these issues and facilitates the acquisition of higher dimensional experiments (Maly et al. 2008). The spectral editing by PRE with long-time spin-locking can probably simplify the crowded spectra at the cost of precision in the distance restraints. The spin-locking time can control the distance that the effective PRE signal decay can be observed, and local line broadened signals can be removed. As an alternative approach, amino acid selective labeling (unlabeling) is also a powerful approach to avoid peak overlap issues. By selecting appropriate labeling schemes, peak overlap should be dramatically reduced even in cases of heptahelical membrane proteins (Etzkorn et al. 2007; Shi et al. 2009). Amino acid selective labeling is not a cost-efficient method for solid-state NMR because the method requires large amounts (10–20 mg) of sample. Nonetheless, recent progress of ultra-fast MAS probes and proton detection should enable the determination of protein structures from sub-milligram samples (Agarwal et al.

2014). It is likely that cell-free expression systems will emerge as a choice for sample preparation. High-throughput structure determination will be achieved by combining the introduction of the metal-chelating unnatural amino acids (Park et al. 2015), selective labeling by cell-free expression systems and proton detection under the ultra-fast MAS condition.

Conclusions

In this paper, we have demonstrated backbone structure determination of GB1 using a combination of the CS-Rosetta protocol and qualitative PRE restraints derived from solid-state NMR data. We revealed that a reduction of the conformational sampling space by fragment assembly in the CS-Rosetta calculation effectively redeems the quality of the PRE restraints. Using this protocol, protein backbone structures can be obtained rapidly from backbone chemical shifts and pairs of paramagnetic and diamagnetic spectra. The backbone structure should aid assignment and validation of complicated through-space correlation spectra. Additionally, the derived structure is a good starting point to refine structures using powder patterns of dipolar couplings and chemical shift anisotropies (Wylie et al. 2011; Das et al. 2012; Park et al. 2012). We believe that our protocol should offer an approach to alleviate the bottleneck of structure determination in solid-state NMR.

Acknowledgments

This work was supported by a Grant-in-Aid for JSPS Fellows (No. 26-4060), and a Grant-in-Aid for Scientific Research on Innovative Areas (No. 26120501) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was performed under the Collaborative Research Program of the Institute for Protein Research, Osaka University.

Electronic supplementary material

Below is the link to the electronic supplementary material.

Supplementary material 1 (PDF 12936 kb)

References

Agarwal V, Penzel S, Székely K et al (2014) De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. *Angew Chem Int Ed Engl* 53:12253–12256. doi:10.1002/anie.201405730

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39:5355–5365. doi:10.1021/bi000060h

Bennett AE, Rienstra CM, Auger M et al (1995) Heteronuclear decoupling in rotating solids. *J Chem Phys* 103:6951–6958. doi:10.1063/1.470372

Bertini I, Luchinat C, Parigi G (2001) *Solution NMR of paramagnetic molecules: applications to metalloproteins and models*. Elsevier, Amsterdam

Bloembergen N (1950) Fine structure of the magnetic resonance line of protons in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. *Physica* 16:95–112. doi:10.1016/0031-8914(50)90067-X

Bradley P (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309:1868–1871. doi:10.1126/science.1113801

Castellani F, van Rossum B, Diehl A et al (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 420:98–102. doi:10.1038/nature01070

Das BB, Nothnagel HJ, Lu GJ et al (2012) Structure determination of a membrane protein in proteoliposomes. *J Am Chem Soc* 134:2047–2056. doi:10.1021/ja209464f

De Paëpe G, Lewandowski JR, Loquet A et al (2008) Proton assisted recoupling and protein structure determination. *J Chem Phys* 129:245101. doi:10.1063/1.3036928

Delaglio F, Grzesiek S, Vuister G et al (1995) NMRPipe: A

- multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:1–17. doi:10.1007/BF00197809
- Demers J-P, Habenstein B, Loquet A et al (2014) High-resolution structure of the *Shigella* type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat Commun* 5:1–12. doi:10.1038/ncomms5976
- Etzkorn M, Martell S, Andronesi OC et al (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. doi:10.1002/anie.200602139
- Franks WT, Zhou DH, Wylie BJ et al (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. doi:10.1021/ja044497e
- Furuita K, Kataoka S, Sugiki T et al (2015) Utilization of paramagnetic relaxation enhancements for high-resolution NMR structure determination of a soluble loop-rich protein with sparse NOE distance restraints. *J Biomol NMR* 61:55–64. doi:10.1007/s10858-014-9882-7
- Goddard TD, Kneller DG (2015) SPARKY 3. <https://www.cgl.ucsf.edu/home/sparky/>. Accessed 17 Jun 2015
- Gottstein D, Reckel S, Dötsch V, Güntert P (2012) Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* 20:1019–1027. doi:10.1016/j.str.2012.03.010
- Grommek A, Meier BH, Ernst M (2006) Distance information from proton-driven spin diffusion under MAS. *Chem Phys Lett* 427:404–409. doi:10.1016/j.cplett.2006.07.005
- Jaroniec CP (2015) Structural studies of proteins by paramagnetic solid-state NMR spectroscopy. *J Magn Reson* 253:50–59.

doi:10.1016/j.jmr.2014.12.017

Koenig SH (1982) A classical description of relaxation of interacting pairs of unlike spins: extension to $T_{1\rho}$, T_2 , and $T_{1\rho\text{off}}$, including contact interactions. *J Magn Reson* 47:441–453. doi:10.1016/0022-2364(82)90211-6

Kosen PA (1989) Spin labeling of proteins. *Meth Enzymol* 177:86–121. doi:10.1016/0076-6879(89)77007-5

Lange OF (2014) Automatic NOESY assignment in CS-RASREC-Rosetta. *J Biomol NMR* 59:147–159. doi:10.1007/s10858-014-9833-3

Lange OF, Baker D (2011) Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation. *Proteins* 80:884–895. doi:10.1002/prot.23245

Lange A, Luca S, Baldus M (2002) Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J Am Chem Soc* 124:9704–9705. doi:10.1021/ja026691b

Lange OF, Rossi P, Sgourakis NG et al (2012) Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci USA* 109:10873–10878. doi:10.1073/pnas.1203013109

Leaver-Fay A, Tyka M, Lewis SM et al (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth Enzymol* 487:545–574. doi:10.1016/B978-0-12-381270-4.00019-6

Lewandowski JR (2013) Advances in solid-state relaxation methodology for probing site-specific protein dynamics. *Acc Chem Res* 46:2018–2027. doi:10.1021/ar300334g

Li J, Pilla KB, Li Q et al (2013) Magic angle spinning NMR structure determination of proteins from pseudocontact shifts. *J Am Chem Soc* 135:8294–8303. doi:10.1021/ja4021149

Liang B, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* 128:4389–4397. doi:10.1021/ja0574825

Loquet A, Bardiaux B, Gardiennet C et al (2008) 3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints. *J Am Chem Soc* 130:3579–3589. doi:10.1021/ja078014t

Loquet A, Sgourakis NG, Gupta R et al (2013) Atomic model of the type III secretion system needle. *Nature* 486:276–279. doi:10.1038/nature11079

Maly T, Debelouchina GT, Bajaj VS et al (2008) Dynamic nuclear polarization at high magnetic fields. *J Chem Phys* 128:052211–052219. doi:10.1063/1.2833582

Manolikas T, Herrmann T, Meier BH (2008) Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. *J Am Chem Soc* 130:3959–3966. doi:10.1021/ja078039s

Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. *J Magn Reson* 162:479–486

Nadaud PS, Helmus JJ, Kall SL, Jaroniec CP (2009) Paramagnetic ions enable tuning of nuclear relaxation rates and provide long-range structural restraints in solid-state NMR of proteins. *J Am Chem Soc* 131:8108–8120. doi:10.1021/ja900224z

Nadaud PS, Sengupta I, Helmus JJ, Jaroniec CP (2011) Evaluation of the influence of intermolecular electron-nucleus couplings and intrinsic metal binding sites on the measurement of ^{15}N longitudinal paramagnetic relaxation enhancements in proteins by solid-state NMR. *J Biomol NMR* 51:293–302. doi:10.1007/s10858-011-9536-y

Nieuwkoop AJ, Rienstra CM (2010) Supramolecular protein structure determination by site-specific long-range intermolecular solid state NMR spectroscopy. *J Am Chem Soc* 132:7570–7571. doi:10.1021/ja100992y

Park SH, Das BB, Casagrande F et al (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491:779–783. doi:10.1038/nature11580

Park SH, Wang VS, Radoicic J et al (2015) Paramagnetic relaxation enhancement of membrane proteins by incorporation of the metal-chelating unnatural amino acid 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid (HQA). *J Biomol NMR* 61:185–196. doi:10.1007/s10858-014-9884-5

Pauli J, van Rossum B, Förster H et al (2000) Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the α -spectrin SH3 domain. *J Magn Reson* 143:411–416. doi:10.1006/jmre.2000.2029

Pines A, Gibby MG, Waugh JS (1973) Proton-enhanced NMR of dilute spins in solids. *J Chem Phys* 59:569–590. doi:10.1063/1.1680061

Raman S, Lange OF, Rossi P et al (2010) NMR structure determination for larger proteins using backbone-only data. *Science* 327:1014–1018. doi:10.1126/science.1183649

Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. *Meth Enzymol* 383:66–93. doi:10.1016/S0076-6879(04)83004-0

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65–73. doi:10.1016/S1090-7807(02)00014-9

Sengupta I, Nadaud PS, Helmus JJ et al (2012) Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy. *Nat Chem* 4:410–417. doi:10.1038/nchem.1299

Shen Y, Lange O, Delaglio F et al (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685–4690. doi:10.1073/pnas.0800256105

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223. doi:10.1007/s10858-009-9333-z

Shi L, Lake EMR, Ahmed MAM et al (2009) Solid-state NMR study of proteorhodopsin in the lipid environment: secondary structure and dynamics. *BBA-Biomembranes* 1788:2563–2574. doi:10.1016/j.bbamem.2009.09.011

Solomon I (1955) Relaxation processes in a system of two spins. *Phys Rev* 99:559–565. doi:10.1103/PhysRev.99.559

Takegoshi K, Nakamura S, Terao T (2001) ^{13}C – ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem Phys Lett* 344:631–637. doi:10.1016/S0009-2614(01)00791-6

Wang S, Munro RA, Kim SY et al (2012) Paramagnetic relaxation enhancement reveals oligomerization interface of a membrane protein. *J Am Chem Soc* 134:16995–16998. doi:10.1021/ja308310z

Wang S, Munro RA, Shi L et al (2013) Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat Meth* 10:1007–1012. doi:10.1038/nmeth.2635

Wickramasinghe NP, Ishii Y (2006) Sensitivity enhancement, assignment, and distance measurement in ^{13}C solid-state NMR spectroscopy for paramagnetic systems under fast magic angle spinning. *J Magn Reson* 181:233–243. doi:10.1016/j.jmr.2006.05.008

Wishart DS, Bigam CG, Holm A et al (1995) ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5:67–81. doi:10.1007/BF00227471

Wylie BJ, Sperling LJ, Nieuwkoop AJ et al (2011) Ultrahigh resolution protein structures using NMR chemical shift tensors. *Proc Natl Acad Sci USA* 108:16974–16979. doi:10.1073/pnas.1103728108