



Title	Study on Protein Structure Determination using Paramagnetic Magic-Angle Spinning Solid-state NMR
Author(s)	田巻, 初
Citation	北海道大学. 博士(生命科学) 甲第12264号
Issue Date	2016-03-24
DOI	10.14943/doctoral.k12264
Doc URL	http://hdl.handle.net/2115/61820
Type	theses (doctoral)
File Information	Hajime_Tamaki.pdf



[Instructions for use](#)

Ph.D. Thesis

Study on Protein Structure Determination using
Paramagnetic Magic-Angle Spinning Solid-state NMR
(常磁性マジック角試料回転固体 NMR 法による
タンパク質立体構造決定に関する研究)

Graduate School of Life Science

Hokkaido University

2016

Hajime Tamaki

Abbreviations	4
Abstract	5
General introduction	7
References	11
Part I Distance restraints from paramagnetic MAS solid-state NMR	18
Abstract	19
Introduction	19
Paramagnetic signal decay of rotating samples	20
Solomon mechanism	20
Paramagnetic anisotropic shift dispersion	26
Materials and Methods	28
Numerical simulation of intra-/intermolecular PREs	28
Protein expression and purification	30
Ligation of EDTA-metal complex to single cysteine mutants	30
Microcrystallization of GB1 samples	30
MAS Solid-state NMR spectroscopy	31
Molecular dynamics simulation	31
Results and Discussion	33
Influence of the intermolecular PREs	33
Structural restrains from PREs	35
Evaluation of error from the pseudo paramagnetic center approximation and EDTA-tag flexibility	38
References	54
Part II Protein structure determination using qualitative PRE restraints	59
Abstract	60
Introduction	60
Methods	61
Structure calculation of GB1	61
CS-Rosetta calculation for simulated data of proteins	62
Results	64
Structure calculation of GB1	64
CS-Rosetta calculation for simulated data of proteins	66

Discussion	67
References	84
Conclusions	87
References	87
Acknowledgements	88

Abbreviations

BMRB	Biological Magnetic Resonance Bank
CP	cross polarization
CSA	CSA
CW	continuous wave
DARR	dipolar assisted rotational resonance
EDTA	2,2',2'',2'''-(Ethane-1,2-diylidinitrilo)tetraacetic acid
GB1	B1 domain of streptococcal protein G
MAS	magic angle spinning
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PCS	pseudocontact shift
PDB	Protein Data Bank
PRE	paramagnetic relaxation enhancement
RDC	residual dipolar coupling
RF	radio-frequency
RMSD	root-mean-square deviation
TPPM	two pulse phase modulation

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 thesis, an approach for structure determination of proteins that combines qualitative transverse PRE restraints and CS-Rosetta is described.

In the part I, the paramagnetic signal decay of the rotating samples is considered by numerical simulations, and the theoretical values are compared to the experimental ones derived from uniformly ^{13}C , ^{15}N labeled and paramagnetic EDTA- Mn^{2+} tag attached B1 domain of streptococcal protein G (GB1). Then, I found the purely attractive and repulsive restraints expressed by the rectified quadratic function is suitable to handle the low precision PRE restraints derived from solid-state NMR.

In the part II, I demonstrated the structure calculations combining the qualitative transverse PRE restraints and CS-Rosetta to redeem the quality of the PRE restraints derived from solid-state NMR. The derived structure of GB1 has a $\text{C}\alpha$ 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- Mn^{2+} mutants because this number of PRE sites is insufficient when using a conventional structure calculation method based on restrained molecular dynamics and simulated annealing. In addition, transferability of this approach is discussed by the structure calculations using simulated PRE restraints, then I concluded the structure of a protein whose size is at least ca. 110 residues can be determined.

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.

General introduction

Solid-state NMR is a powerful method to analyze structure of insoluble molecules. In the field of structural biology, membrane proteins, protein fibrils and large molecular assemblies are the major targets of solid-state NMR measurements. The solid-state NMR signal lines of "raw" samples are extremely broad (typically >10 kHz), because the effect of various anisotropic interaction (e.g. chemical shift anisotropy and dipolar coupling) cannot be averaged out by fast molecular tumbling. To overcome this problem, two principal approaches are generally used. Preparation of mechanically, magnetically and/or motionally aligned samples is widely used approach to investigate membrane protein structures and their membrane orientation [1-4]. The second approach is using magic angle spinning (MAS) technique to remove the anisotropic interactions by fast sample spinning (>5 kHz) around the magic angle (54.74° to the external magnetic field) [5-8]. This approach has three significant advantages: (i) The sample orientation is not necessary; then various samples can be measured. (ii) The isotropic chemical shifts those are observable in solution NMR can be measured. (iii) The removed anisotropic interactions can be reintroduced by radio-frequency (RF) pulse schemes (so-called recoupling); then coherence transfer and collection of structural information via anisotropic interactions can be achieved. These features allow us to design measurement and analysis strategies analogous with those of solution NMR. The feasibility of protein structure determination by MAS solid-state NMR has been shown by structure determination of α -spectrin SH3 domain[9]. Recently, technologies included high-field magnets[10], spectrometers and probes[11-14], and RF schemes/techniques [15-18] are rapidly improved. At last, MAS solid-state NMR achieved a significant milestone for protein structure determination with studies describing the structure of heptahelical membrane proteins [4,19].

In general, protein structure determination by MAS solid-state NMR is performed by following three steps; (1) chemical shift assignment, (2) unambiguous space through correlation assignment and initial structure calculation and (3) model refinement by ambiguous distance restraints. The backbone and side-chain assignment protocols using e.g. 3D-NCACX, NCOCX,

CANCO and CCC measurements are well established [20,21]. Recently, rapid assignment strategies based on proton detection under the ultra fast MAS condition (>60 kHz) are rapidly developed [22-27]. Using 4D measurements are also proposed to resolve the chemical shift ambiguity [28-31]. In addition, automatic assignment algorithms specialized for solid-state NMR are designed [32-36]. Therefore, difficulty of the chemical shift assignment is gradually decreased. However, even today, collection of space through correlations is still a challenging task and a major bottleneck of protein structure determination.

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 [37]. 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 [9,19,38,39]. 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 [18]. Proton-assisted recoupling (PAR) avoids homonuclear dipolar truncation by the third spin-assisted recoupling mechanism [16]. 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 [40,41] and CANDID [42]) have been introduced for solid-state NMR data analysis using methods that were originally developed for solution-state NMR [43,44]. 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 [45]. 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 utilization of paramagnetic species is a powerful method to obtain long-range distance restraints without increasing the number of cross peaks. The unpaired electron changes the NMR observables, chemical shifts and relaxation rates, depending on the nuclei-unpaired electron geometry. The changes of the chemical shift and relaxation rate by the paramagnetic species are called pseudocontact shift (PCS) [46] and paramagnetic relaxation enhancement (PRE) [47], respectively. The PCSs and PREs can edit the NMR spectra based on the molecular shape (Figure. 1). In general, proteins do not have the paramagnetic center. Then, the paramagnetic probes can be introduced to some arbitrary position on the protein surface [48-54], and the geometry between the specific site and nuclei of the protein can be obtained easily from simple chemical shift correlation spectra once the backbone assignment process is complete. Addition, the electron gyromagnetic ratio is 2-3 orders of magnitude larger than that of nuclei, providing $>10 \text{ \AA}$ distance restraints. For *de-novo* protein structure determination, measurements of PREs are more suitable than those of PCSs, because values of magnetic susceptibility anisotropy tensor ($\Delta\chi$ -tensor) determined using known protein 3D structure are required to convert the PCS values to geometric restraints. PRE techniques are widely used for protein structure determination from limited distance restraints in solution NMR [55-59]. In solid-state NMR, the initial study describing introduction of PREs to MAS solid-state NMR protein structure analyses were reported around 2007 [60]. Recently, structure determination using PRE distance restraints have been reported [19,61-63].

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 [64]. Thus, a method that use low precision distance restraints is attractive for protein structure analysis by solid-state NMR.

In this thesis, an approach for structure determination of proteins that combines qualitative transverse PRE restraints and an advanced computational method is described. In part I of this thesis, I investigated the property of paramagnetic signal decay by transverse PRE in rotating solid-state samples using numerical calculations and measurements of a model protein, B1 domain of streptococcal protein G (GB1). In part II, I demonstrated the structure calculations combining the qualitative transverse PRE restraints and CS-Rosetta, a method of protein 3D structure prediction from the backbone chemical shifts [65].

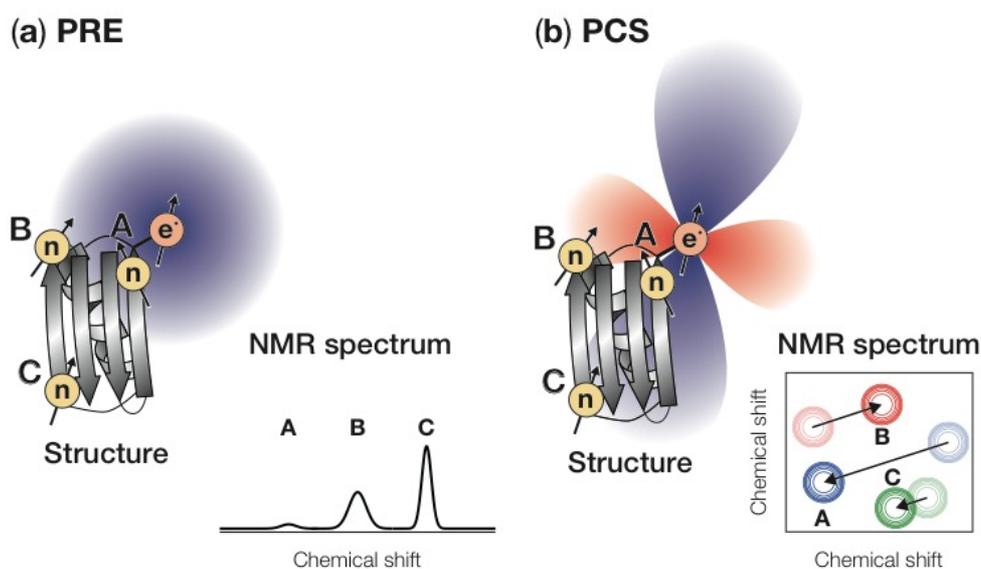


Figure 1. Schematic paramagnetic NMR spectra. (a) and (b) illustrate the effects of PRE and PCS, respectively.

References

- [1] N.J. Traaseth, L. Shi, R. Verardi, D.G. Mullen, G. Barany, G. Veglia, Structure and topology of monomeric phospholamban in lipid membranes determined by a hybrid solution and solid-state NMR approach, *Proc. Natl. Acad. Sci. U.S.a.* 106 (2009) 10165–10170. doi:10.1073/pnas.0904290106.
- [2] M. Sharma, M. Yi, H. Dong, H. Qin, E. Peterson, D.D. Busath, et al., Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer, *Science*. 330 (2010) 509–512. doi:10.1126/science.1191750.
- [3] B.B. Das, H.J. Nothnagel, G.J. Lu, W.S. Son, Y. Tian, F.M. Marassi, et al., Structure Determination of a Membrane Protein in Proteoliposomes, *J. Am. Chem. Soc.* 134 (2012) 2047–2056. doi:10.1021/ja209464f.
- [4] S.H. Park, B.B. Das, F. Casagrande, Y. Tian, H.J. Nothnagel, M. Chu, et al., Structure of the chemokine receptor CXCR1 in phospholipid bilayers, *Nature*. 491 (2012) 779–783. doi:10.1038/nature11580.
- [5] E.R. Andrew, R.A. Newing, The Narrowing of Nuclear Magnetic Resonance Spectra by Molecular Rotation in Solids, *Proc. Phys. Soc.* 72 (1958) 959–972. doi:10.1088/0370-1328/72/6/304.
- [6] E.R. Andrew, A. Bradbury, R.G. Eades, Nuclear magnetic resonance spectra from a crystal rotated at high speed, *Nature*. (1958). doi:10.1038/1821659a0.
- [7] E.R. Andrew, A. Bradbury, R.G. Eades, Removal of dipolar broadening of nuclear magnetic resonance spectra of solids by specimen rotation, *Nature*. 183 (1959) 1802–1803. doi:10.1038/1831802a0.
- [8] I.J. Lowe, Free induction decays of rotating solids, *Phys. Rev. Lett.* 2 (1959) 285–287. doi:10.1103/PhysRevLett.2.285.
- [9] F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy, *Nature*. 420 (2002) 98–102. doi:10.1038/nature01070.
- [10] K. Hashi, S. Ohki, S. Matsumoto, G. Nishijima, A. Goto, K. Deguchi, et al., Achievement of 1020MHz NMR, *J. Magn. Reson.* 256 (2015) 30–33. doi:10.1016/j.jmr.2015.04.009.

- [11] V. Agarwal, S. Penzel, K. Székely, R. Cadalbert, E. Testori, A. Oss, et al., De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy, *Angew. Chem. Int. Ed. Engl.* 53 (2014) 12253–12256. doi:10.1002/anie.201405730.
- [12] C.H. Wu, C.V. Grant, G.A. Cook, S.H. Park, S.J. Opella, *Journal of Magnetic Resonance*, *J Magn Reson.* 200 (2009) 74–80. doi:10.1016/j.jmr.2009.06.004.
- [13] S.A. McNeill, P.L. Gor'kov, K. Shetty, W.W. Brey, J.R. Long, A low-E magic angle spinning probe for biological solid state NMR at 750MHz, *J Magn Reson.* 197 (2009) 135–144. doi:10.1016/j.jmr.2008.12.008.
- [14] J.A. Stringer, C.E. Bronnimann, C.G. Mullen, D.H. Zhou, S.A. Stellfox, Y. Li, et al., Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes, *J Magn Reson.* 173 (2005) 40–48. doi:10.1016/j.jmr.2004.11.015.
- [15] V. Chevelkov, K. Giller, S. Becker, A. Lange, Efficient CO-CA transfer in highly deuterated proteins by band-selective homonuclear cross-polarization, *J. Magn. Reson.* 230 (2013) 205–211. doi:10.1016/j.jmr.2013.02.021.
- [16] G. De Paëpe, J.R. Lewandowski, A. Loquet, A. Böckmann, R.G. Griffin, Proton assisted recoupling and protein structure determination, *J. Chem. Phys.* 129 (2008) 245101. doi:10.1063/1.3036928.
- [17] J.R. Lewandowski, G. De Paëpe, R.G. Griffin, Proton assisted insensitive nuclei cross polarization, *J. Am. Chem. Soc.* 129 (2007) 728–729. doi:10.1021/ja0650394.
- [18] A. Lange, S. Luca, M. Baldus, Structural Constraints from Proton-Mediated Rare-Spin Correlation Spectroscopy in Rotating Solids, *J. Am. Chem. Soc.* 124 (2002) 9704–9705. doi:10.1021/ja026691b.
- [19] S. Wang, R.A. Munro, L. Shi, I. Kawamura, T. Okitsu, A. Wada, et al., Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein, *Nat Meth.* 10 (2013) 1007–1012. doi:10.1038/nmeth.2635.
- [20] A. Schuetz, C. Wasmer, B. Habenstein, R. Verel, J. Greenwald, R. Riek, et al., Protocols for the Sequential Solid-State NMR Spectroscopic Assignment of a Uniformly Labeled 25 kDa Protein: HET-s(1-227), *Chem. Eur. J. of Chem. Bio.* 11 (2010) 1543–1551. doi:10.1002/cbic.201000124.
- [21] L. Shi, M.A.M. Ahmed, W. Zhang, G. Whited, L.S. Brown, V. Ladizhansky, Three-Dimensional Solid-State NMR Study of a Seven-Helical Integral Membrane Protein

- Pump—Structural Insights, *Journal of Molecular Biology*. 386 (2009) 1078–1093. doi:10.1016/j.jmb.2009.01.011.
- [22] N. Kulminkaya, S.K. Vasa, K. Giller, S. Becker, A. Kwan, M. Sunde, et al., Access to side-chain carbon information in deuterated solids under fast MAS through non-rotor-synchronized mixing, *Chem. Commun.* (2015) 1–4. doi:10.1039/C5CC07345F.
- [23] D. Mance, T. Sinnige, M. Kaplan, An Efficient Labelling Approach to Harness Backbone and Side-Chain Protons in ^1H -Detected Solid-State NMR Spectroscopy, *Angewandte ...* 127 (2015) 16025–16029. doi:10.1002/ange.201509170.
- [24] S. Penzel, A.A. Smith, V. Agarwal, A. Hunkeler, M.-L. Org, A. Samoson, et al., Protein resonance assignment at MAS frequencies approaching 100 kHz: a quantitative comparison of J-coupling and dipolar-coupling-based transfer methods, *J Biomol NMR*. 63 (2015) 165–186. doi:10.1007/s10858-015-9975-y.
- [25] L.B. Andreas, J. Stanek, T. Marchand, A. Bertarello, D.C.-D. Paepe, D. Lalli, et al., Protein residue linking in a single spectrum for magic-angle spinning NMR assignment, *J Biomol NMR*. (2015) 1–9. doi:10.1007/s10858-015-9956-1.
- [26] E. Barbet-Massin, A.J. Pell, J.S. Retel, L.B. Andreas, K. Jaudzems, W.T. Franks, et al., Rapid Proton-Detected NMR Assignment for Proteins with Fast Magic Angle Spinning, *J. Am. Chem. Soc.* (2014) 140818104742001. doi:10.1021/ja507382j.
- [27] A. Marchetti, S. Jehle, M. Felletti, M.J. Knight, Y. Wang, Z.-Q. Xu, et al., Backbone assignment of fully protonated solid proteins by ^1H detection and ultrafast magic-angle-spinning NMR spectroscopy, *Angew. Chem. Int. Ed. Engl.* 51 (2012) 10756–10759. doi:10.1002/anie.201203124.
- [28] S. Xiang, V. Chevelkov, S. Becker, A. Lange, Towards automatic protein backbone assignment using proton-detected 4D solid-state NMR data, *J Biomol NMR*. 60 (2014) 85–90. doi:10.1007/s10858-014-9859-6.
- [29] W. Trent Franks, H.S. Atreya, T. Szyperski, C.M. Rienstra, GFT projection NMR spectroscopy for proteins in the solid state, *J Biomol NMR*. 48 (2010) 213–223. doi:10.1007/s10858-010-9451-7.
- [30] L. Shi, E.M.R. Lake, M.A.M. Ahmed, L.S. Brown, V. Ladizhansky, Solid-state NMR study of proteorhodopsin in the lipid environment: Secondary structure and dynamics, *BBA - Biomembranes*. 1788 (2009) 2563–2574. doi:10.1016/j.bbamem.2009.09.011.

- [31] W.T. Franks, K.D. Kloepper, B.J. Wylie, C.M. Rienstra, Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins, *J Biomol NMR*. 39 (2007) 107–131. doi:10.1007/s10858-007-9179-1.
- [32] J.T. Nielsen, N. Kulminskaya, M. Bjerring, N.C. Nielsen, Automated robust and accurate assignment of protein resonances for solid state NMR, *J Biomol NMR*. 59 (2014) 119–134. doi:10.1007/s10858-014-9835-1.
- [33] E. Schmidt, J. Gath, B. Habenstein, F. Ravotti, K. Székely, M. Huber, et al., Automated solid-state NMR resonance assignment of protein microcrystals and amyloids, *J Biomol NMR*. (2013). doi:10.1007/s10858-013-9742-x.
- [34] K.-N. Hu, W. Qiang, R. Tycko, A general Monte Carlo/simulated annealing algorithm for resonance assignment in NMR of uniformly labeled biopolymers, *J Biomol NMR*. 50 (2011) 267–276. doi:10.1007/s10858-011-9517-1.
- [35] R. Tycko, K.-N. Hu, A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning, *J Magn Reson*. 205 (2010) 304–314. doi:10.1016/j.jmr.2010.05.013.
- [36] H.N.B. Moseley, L.J. Sperling, C.M. Rienstra, Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of β 1 immunoglobulin binding domain of protein G (GB1), *J Biomol NMR*. 48 (2010) 123–128. doi:10.1007/s10858-010-9448-2.
- [37] A. Grommek, B.H. Meier, M. Ernst, Distance information from proton-driven spin diffusion under MAS, *Chem Phys Lett*. 427 (2006) 404–409. doi:10.1016/j.cplett.2006.07.005.
- [38] A. Loquet, N.G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, et al., Atomic model of the type III secretion system needle, *Nature*. 486 (2013) 276–279. doi:10.1038/nature11079.
- [39] J.-P. Demers, B. Habenstein, A. Loquet, S.K. Vasa, K. Giller, S. Becker, et al., High-resolution structure of the *Shigella* type-III secretion needle by solid-state NMR and cryo-electron microscopy, *Nat. Commun*. 5 (2014) 1–12. doi:10.1038/ncomms5976.
- [40] M. Nilges, Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities, *Journal of Molecular Biology*. 245 (1995) 645–660. doi:10.1006/jmbi.1994.0053.

- [41] M. Nilges, M.J. Macias, S.I. O'Donoghue, H. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β -spectrin, *Journal of Molecular Biology*. 269 (1997) 408–422.
- [42] T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *Journal of Molecular Biology*. 319 (2002) 209–227. doi:10.1016/S0022-2836(02)00241-3.
- [43] T. Manolikas, T. Herrmann, B.H. Meier, Protein Structure Determination from ^{13}C Spin-Diffusion Solid-State NMR Spectroscopy, *J. Am. Chem. Soc.* 130 (2008) 3959–3966. doi:10.1021/ja078039s.
- [44] A. Loquet, B. Bardiaux, C. Gardiennet, C. Blanchet, M. Baldus, M. Nilges, et al., 3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints, *J. Am. Chem. Soc.* 130 (2008) 3579–3589. doi:10.1021/ja078014t.
- [45] J. Pauli, B. van Rossum, H. Förster, H.J.M. de Groot, H. Oschkinat, 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 (2000) 411–416. doi:10.1006/jmre.2000.2029.
- [46] H.M. McConnell, R.E. Robertson, Isotropic Nuclear Resonance Shifts, *J. Chem. Phys.* 29 (1958) 1361–1365. doi:10.1063/1.1744723.
- [47] I. Solomon, Relaxation processes in a system of two spins, *Phys Rev.* 99 (1955) 559–565. doi:10.1103/PhysRev.99.559.
- [48] I. Sengupta, M. Gao, R.J. Arachchige, P.S. Nadaud, T.F. Cunningham, S. Saxena, et al., Protein structural studies by paramagnetic solid-state NMR spectroscopy aided by a compact cyclen-type Cu(II) binding tag, *J Biomol NMR.* 61 (2015) 1–6. doi:10.1007/s10858-014-9880-9.
- [49] C.T. Loh, K. Ozawa, K.L. Tuck, N. Barlow, T. Huber, G. Otting, et al., Lanthanide Tags for Site-Specific Ligation to an Unnatural Amino Acid and Generation of Pseudocontact Shifts in Proteins, *Bioconjugate Chem.* 24 (2013) 260–268. doi:10.1021/bc300631z.
- [50] N.L. Fawzi, M.R. Fleissner, N.J. Anthis, T. Kálai, K. Hideg, W.L. Hubbell, et al., A rigid disulfide-linked nitroxide side chain simplifies the quantitative analysis of PRE data, *J Biomol NMR.* 51 (2011) 105–114. doi:10.1007/s10858-011-9545-x.

- [51] X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, et al., A Dipicolinic Acid Tag for Rigid Lanthanide Tagging of Proteins and Paramagnetic NMR Spectroscopy, *J. Am. Chem. Soc.* 130 (2008) 10486–10487. doi:10.1021/ja803741f.
- [52] Y.W. Ebright, Y. Chen, P.S. Pendergrast, R.H. Ebright, Incorporation of an EDTA-metal complex at a rationally selected site within a protein: application to EDTA-iron DNA affinity cleaving with catabolite gene activator ..., *Biochemistry*. 31 (1992) 10664–10670. doi:10.1021/bi00159a004.
- [53] M.R. Ermácora, J.M. Delfino, B. Cuenoud, A. Schepartz, R.O. Fox, Conformation-dependent cleavage of staphylococcal nuclease with a disulfide-linked iron chelate, *Proc. Natl. Acad. Sci. U.S.a.* 89 (1992) 6383–6387.
- [54] L.J. Berliner, J. Grunwald, H.O. Hankovszky, K. Hideg, A novel reversible thiol-specific spin label: papain active site labeling and inhibition, *Anal. Biochem.* 119 (1982) 450–455. doi:10.1016/0003-2697(82)90612-1.
- [55] K. Furuita, S. Kataoka, T. Sugiki, Y. Hattori, N. Kobayashi, T. Ikegami, et al., 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 (2015) 55–64. doi:10.1007/s10858-014-9882-7.
- [56] D. Gottstein, S. Reckel, V. Dötsch, P. Güntert, Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR, *Structure*. 20 (2012) 1019–1027. doi:10.1016/j.str.2012.03.010.
- [57] S. Reckel, D. Gottstein, J. Stehle, F. Löhr, M.-K. Verhoefen, M. Takeda, et al., Solution NMR Structure of Proteorhodopsin, *Angew. Chem. Int. Ed.* 50 (2011) 11942–11946. doi:10.1002/anie.201105648.
- [58] B. Liang, J.H. Bushweller, L.K. Tamm, Site-Directed Parallel Spin-Labeling and Paramagnetic Relaxation Enhancement in Structure Determination of Membrane Proteins by Solution NMR Spectroscopy, *J. Am. Chem. Soc.* 128 (2006) 4389–4397. doi:10.1021/ja0574825.
- [59] J.L. Battiste, G. Wagner, 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 (2000) 5355–5365. doi:10.1021/bi000060h.

- [60] P.S. Nadaud, J.J. Helmus, N. Höfer, C.P. Jaroniec, Long-Range Structural Restraints in Spin-Labeled Proteins Probed by Solid-State Nuclear Magnetic Resonance Spectroscopy, *J. Am. Chem. Soc.* 129 (2007) 7502–7503. doi:10.1021/ja072349t.
- [61] P. Rovó, K. Grohe, K. Giller, S. Becker, R. Linser, Proton transverse relaxation as a sensitive probe for long-distance paramagnetic structural information, *Chemphyschem.* (2015). doi:10.1002/cphc.201500799.
- [62] I. Sengupta, P.S. Nadaud, J.J. Helmus, C.D. Schwieters, C.P. Jaroniec, Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy, *Nat Chem.* 4 (2012) 410–417. doi:10.1038/nchem.1299.
- [63] S. Wang, R.A. Munro, S.Y. Kim, K.-H. Jung, L.S. Brown, V. Ladizhansky, Paramagnetic Relaxation Enhancement Reveals Oligomerization Interface of a Membrane Protein, *J. Am. Chem. Soc.* 134 (2012) 16995–16998. doi:10.1021/ja308310z.
- [64] J.R. Lewandowski, Advances in solid-state relaxation methodology for probing site-specific protein dynamics, *Acc. Chem. Res.* 46 (2013) 2018–2027. doi:10.1021/ar300334g.
- [65] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G. Liu, et al., Consistent blind protein structure generation from NMR chemical shift data, *Proc Natl Acad Sci USA.* 105 (2008) 4685–4690. doi:10.1073/pnas.0800256105.

Part I

Distance restraints from paramagnetic MAS solid-state NMR

Abstract

Paramagnetic relaxation enhancement brings unambiguous long-range distance restraints those have crucial roles for protein structure determination by NMR. In solid-state, the PRE process is interfered by the incompletely averaged anisotropic interactions originated from immobility of the molecules. Paramagnetic anisotropic shift dispersion in immobile solid-state samples also decreases the signal intensities in addition to PREs. Furthermore, dense molecular packing causes non-negligible intermolecular PREs those are unwanted in the protein structure determination. Then, consideration of the influence of those factors are necessary to handle PRE distance restraints in solid-state NMR. In this part, the paramagnetic signal decay of the rotating samples is considered by numerical simulations, and the theoretical values are compared to the experimental ones derived from uniformly ^{13}C , ^{15}N labeled and paramagnetic EDTA- Mn^{2+} tag attached GB1. Spin dynamics simulations of cross polarization under the paramagnetic center is presence condition illustrate the influence of paramagnetic anisotropic interaction is included in the PRE signal decay model. Intermolecular PREs derived from the crystal packing model of GB1 suggest 1:3 dilution of paramagnetic molecules by diamagnetic ones are required for the effective intermolecular PRE suppression. Despite of those considerations, some experimentally derived distance restraints have large error. Finally, we found the purely attractive and repulsive restraints expressed by the rectified quadratic function is suitable to handle the low precision PRE restraints derived from solid-state NMR.

Introduction

The measurement of PRE is a promising method to obtain long-range ($\sim 20\text{--}25$ Å) distance restraints. In solution NMR, it is well known that the nucleus-unpaired electron distance can be quantitatively obtained from peak intensity ratio of the paramagnetic and diamagnetic samples and Solomon-Bloembergen equation [1-3]. However, in solid-state NMR, it is not an obvious manner. The PRE process is originated from fluctuation of the dipolar coupling between nucleus and unpaired electron, and this fluctuation is caused by random (incoherent) molecular motions.

Under the condition that the anisotropic interactions are incompletely averaged, coherent process (e.g. dipolar dephasing by homo- and heteronuclear dipolar couplings between ^1H and ^{13}C spins) also interferes the nuclear relaxation. The coherent processes are difficult to distinguish from the incoherent PRE process [4]. In addition, some paramagnetic center causes paramagnetic anisotropic shift dispersion and decreases signal intensities because the cross polarization (CP: a fundamental RF technique in solid-state NMR [5]) efficiency is reduced in addition to PRE [6]. Thus, the distance information derived from PREs has low precision. To avoid this precision problem, Jaroniec and co-workers used Cu^{2+} - ^{15}N longitudinal PREs to improve precision [7]. 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, investigation of the influences of the coherent processes, paramagnetic anisotropic shift dispersion and intermolecular PREs is important to handle the PRE profiles derived from transverse PREs those can readily obtained from the peak intensity ratio of the paramagnetic and diamagnetic samples. In this part, the paramagnetic signal decay of the rotating samples is considered. the influences of the coherent processes, paramagnetic anisotropic shift dispersion and intermolecular PREs are estimated by numerical simulations. Then, the theoretical values are compared to the experimental ones derived from uniformly ^{13}C , ^{15}N labeled and paramagnetic EDTA- Mn^{2+} tag attached GB1. Finally, we propose a PRE restraints handling manner.

Paramagnetic signal decay of rotating samples

Solomon mechanism

PRE restraints were obtained using the pulse sequence shown in Figure 1-1(a). The ^{13}C - ^{13}C

spectra using short DARR mixing [8] 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.

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 CP [5] period was estimated by a kinetic model as follow. The spin temperature changes during CP can be described in the following differential equation.

$$\frac{d}{dt} \begin{pmatrix} \beta_H \\ \beta_C \end{pmatrix} = \begin{pmatrix} -\frac{1}{T_{IS}} - \frac{1}{T_{1\rho,H}} & \frac{1}{T_{IS}} \\ \frac{\epsilon'}{T_{IS}} & -\frac{\epsilon'}{T_{IS}} - \frac{1}{T_{1\rho,C}} \end{pmatrix} \begin{pmatrix} \beta_H \\ \beta_C \end{pmatrix} \quad (1-1)$$

β_H and β_C are the inverse spin temperatures of ^1H and ^{13}C spins, respectively. $1/T_{IS}$ is the transfer rate between ^1H and ^{13}C spins, and $1/T_{1\rho,H}$ and $1/T_{1\rho,C}$ are longitudinal relaxation rates in the rotating frame of ^1H and ^{13}C , respectively. ϵ' is the ratio of heat capacities of the ^{13}C and ^1H spins [9].

β_H and β_C are proportional to the ^1H and ^{13}C magnetization, respectively ($\beta_H \propto I_H, \beta_C \propto I_C$). $1/T_{1\rho,H}$ and $1/T_{1\rho,C}$ can be expressed as

$$\frac{1}{T_{1\rho,H}} = R_{1\rho,H} + \Gamma_{1\rho,H} \quad (1-2)$$

$$\frac{1}{T_{1\rho,C}} = R_{1\rho,C} + \Gamma_{1\rho,C} \quad (1-3)$$

$R_{1\rho,H}$ and $R_{1\rho,C}$ are the intrinsic relaxation rates in the rotating frame of ^1H and ^{13}C , respectively. $\Gamma_{1\rho,H}$ and $\Gamma_{1\rho,C}$ are the paramagnetic induced relaxation rates. The intrinsic contribution parts can be neglected ($R_{1\rho,H}, R_{1\rho,C} \sim 0$) because the contact time used in our experiments is short.

In the Hartmann-Hahn condition, ϵ' is given as:

$$\epsilon' = \frac{N_S S(S+1)}{N_I I(I+1)} \quad (1-4)$$

N_I and N_S are the numbers of ^1H and ^{13}C spins, respectively, and I and S are the spin quantum numbers of ^1H and ^{13}C , respectively. For the ^{13}C enriched ^1H - ^{13}C spin pairs, the ratio of heat capacities is deemed to be ~ 1 .

Using the above approximations, equation (1-1) can be described as:

$$\frac{d}{dt} \begin{pmatrix} I_H \\ I_C \end{pmatrix} = \begin{pmatrix} -R_{IS} - \Gamma_{1\rho,H} & R_{IS} \\ R_{IS} & -R_{IS} - \Gamma_{1\rho,C} \end{pmatrix} \begin{pmatrix} I_H \\ I_C \end{pmatrix} \quad (1-5)$$

where, $R_{IS} = 1/T_{IS}$.

The ^{13}C magnetization, I_C given by equation (1-5), is expressed as:

$$I_C = \exp(\lambda_1 t) + \exp(\lambda_2 t) \quad (1-6)$$

where,

$$\lambda_1 = \frac{-(2R_{IS} + \Gamma_{1\rho,H} + \Gamma_{1\rho,C}) + \sqrt{4R_{IS}^2 + (\Gamma_{1\rho,H} - \Gamma_{1\rho,C})^2}}{2}, \quad (1-7)$$

$$\lambda_2 = \frac{-(2R_{IS} + \Gamma_{1\rho,H} + \Gamma_{1\rho,C}) - \sqrt{4R_{IS}^2 + (\Gamma_{1\rho,H} - \Gamma_{1\rho,C})^2}}{2} \quad (1-8)$$

The rate constant, R_{IS} was experimentally determined to be 6.63 ms^{-1} by the build-up curve of contact time dependent $\text{C}\alpha$ signal intensities in CPMAS spectra of N-[1- ^{13}C]acetyl-[U- ^{13}C , ^{15}N]L-valine (NAV). The $\text{C}\alpha$ signal intensities and the curve fitting result are shown in Figure 1-2(a,b). The ^{13}C magnetization growth in some Mn^{2+} - $^1\text{H} / ^{13}\text{C}$ distances are shown in Figure 1-2(c).

The signal reduction of ^{13}C , I_{CP} in the CP period, is given by:

$$I_{CP}(t) = \frac{\exp[\lambda_1(\Gamma_{1\rho,H}, \Gamma_{1\rho,C})t] + \exp[\lambda_2(\Gamma_{1\rho,H}, \Gamma_{1\rho,C})t]}{1 - \exp(-2R_{IS}t)} \quad (1-9)$$

where t is the contact time. The denominator, $1 - \exp(-2R_{IS}t)$ is a form of I_C under the condition of $\Gamma_{1\rho,H} = \Gamma_{1\rho,C} = 0$.

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:

$$I_{SL}(t) = \exp(-\Gamma_{1\rho,C}t) \quad (1-10)$$

$$I_{DARR}(t) = \exp(-\Gamma_{1,C}t) \quad (1-11)$$

$\Gamma_{1,C}$ 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 ($\Gamma_{2,C}$), $R_{2,para} = R_{2,dia} + \Gamma_{2,C}$ [10]. Thus, the peak intensity reduction during the evolution and detection periods, I_{t_1} and I_{t_2} are expressed as:

$$I_{t_1} = I_{t_2} = \frac{H_{para}}{H_{dia}} = \frac{1/(R_{2,dia} + \Gamma_{2,C})}{1/R_{2,dia}} = \frac{\pi w_{dia}}{\pi w_{dia} + \Gamma_{2,C}} \quad (1-12)$$

Finally, the total signal decay, $I = I_{para}/I_{dia}$ is expressed using equation (1-9), (1-10), (1-11) and (1-12) as:

$$I = \frac{I_{para}}{I_{dia}} = I_{CP}(\tau_{CP}) \cdot I_{SL}(\tau_{SL}) \cdot I_{t_1} \cdot I_{DARR}(\tau_{DARR}) \cdot I_{t_2} \quad (1-13)$$

The paramagnetic relaxation rates, Γ_1 , $\Gamma_{1\rho}$, and Γ_2 are given by the Solomon-Bloembergen equation [1-3].

$$\Gamma_1 = \frac{2}{15} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_n^2 g_e^2 \beta^2 s(s+1)}{r^6} \left(\frac{3\tau_{c1}}{1 + \omega_n^2 \tau_{c1}^2} + \frac{7\tau_{c2}}{1 + \omega_e^2 \tau_{c2}^2} \right) \quad (1-14)$$

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

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} [7].

The peak intensity decay profile is dependent on the unpaired electron-nucleus distance as shown in Figure 1-1(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 ([11], 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 (Figure 1-3). 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 RF 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.

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 $T_{1e} \approx 100 \text{ ns}$ [11,12]) 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.

Paramagnetic anisotropic shift dispersion

Mn^{2+} causes paramagnetic anisotropic shift dispersion in the immobile solid-state owing to its large electron spin quantum number and decreases signal intensities because the CP efficiency is reduced in addition to PRE. The line shape of paramagnetic anisotropic shift dispersion is similar to the line shape of chemical shift anisotropy (CSA). Thus, the line shape scales linearly with the B_0 field and can be removed by MAS [13]. In addition, the CP efficiency is dependent on the B_1 field.

To evaluate the effects from paramagnetic anisotropic shift dispersions under the experimental conditions used in this report, we performed spin dynamics simulations of ^1H - ^{13}C CP in the presence of Mn^{2+} at $B_0 = 14.1$ T, a sample temperature of 273 K and a MAS condition of 12.5 kHz. Spin dynamics simulations were performed on a ^1H - ^{13}C two-spin system (I-S spin-system) with thermally averaged Mn^{2+} electron spins using an in-house program written in C++. The ^1H - ^{13}C distance was set at 1.08 Å (the $\text{H}\alpha$ - $\text{C}\alpha$ bond length in CHARMM22 force field [14,15]) and the Mn^{2+} - $^1\text{H}/^{13}\text{C}$ distance was set to 1.0–25.0 Å (variable). The coordinates of ^1H and ^{13}C viewed from the Mn^{2+} were approximated to be the same. The coupling vectors of Mn^{2+} - $^1\text{H}/^{13}\text{C}$ and ^1H - ^{13}C in the principal axis frame were aligned to the B_0 field. The B_1 field of ^1H and ^{13}C were set to 50.0 kHz and 37.5 kHz, respectively. The Hamiltonian in the Zeeman interaction frame was defined by:

$$H = H_{RF,H} + H_{RF,C} + H_{D,Mn-H} + H_{D,Mn-C} + H_{D,H-C} \quad (1-16)$$

where $H_{RF,H}$ and $H_{RF,C}$ are RF Hamiltonians of ^1H and ^{13}C (the phases are set to x), and $H_{D,H-C}$ is the dipolar coupling Hamiltonians of ^1H - ^{13}C . $H_{D,Mn-H}$ and $H_{D,Mn-C}$ are the dipolar coupling Hamiltonians of $^1\text{H}/^{13}\text{C}$ and thermally averaged Mn^{2+} electron spins. The mathematical

expressions of the Hamiltonians are described in [16]. The thermally averaged dipolar coupling Hamiltonians are treated in the same manner as the axially symmetric CSA. The dipolar coupling constants between Mn^{2+} - $^1\text{H}/^{13}\text{C}$ are approximately given by:

$$\frac{\gamma\mu_0}{4\pi r^3} \left\{ \frac{2s+1}{2s} \coth \frac{(2s+1)\beta g B_0}{2k_B T} - \frac{1}{2s} \coth \frac{g\beta B_0}{2k_B T} \right\} \quad (1-17)$$

where γ is the gyromagnetic ratio of nuclei, μ_0 is the permeability of free space, s is the electron spin number of Mn^{2+} , β is the Bohr magneton, g is the g-factor of Mn^{2+} , k_B is the Boltzmann constant, T is the sample temperature and r is the distance between Mn^{2+} and nuclei [2,17]. The g-factor of Mn^{2+} is anisotropic and, g_{xx} , g_{yy} and g_{zz} are 2.00305, 2.00339 and 2.00953, respectively [18]. However, the anisotropy is small, and thus the isotropic g-factor ($g = (g_{xx} + g_{yy} + g_{zz})/3$) was used.

Time evolution of the density operator was calculated by the integrated Liouville-von Neumann equation:

$$\rho(t) = U(t, 0)\rho(0)U^\dagger(t, 0) \quad (1-18)$$

where $\rho(t)$ is the density operator, and $U(t, 0)$ is the unitary propagator for the time period from 0 to t . The initial state, $\rho(0)$ is set to I_x and the transverse magnetization of the ^{13}C spin (S) at time, t , was obtained by $\text{tr}\{S_x\rho(t)\}$. The relaxation of nuclei was omitted. The integration of the unitary operator derived from the total Hamiltonian, H , coordinates the transformation of the principal axis into the laboratory frame and powder averaging was performed following a previous reference [16]. The powder averaging was performed by 28,656 poses of the planar ZCW method [19-22].

Figure 1-4(a) shows the CP buildup curves under various Mn^{2+} - $^1\text{H}/^{13}\text{C}$ distance

conditions. It is obvious that the efficiencies of the polarization transfer increase as the Mn^{2+} - $^1\text{H}/^{13}\text{C}$ distance is incremented. At 10 Å, maximum efficiency relative to a condition with an infinite distance is ~100%. Indeed, the line width ($\delta_{\parallel} - \delta_{\perp}$) of a static ^1H in an Mn^{2+} coupled ^1H single spin-system is ca. 6.9 kHz (11.5 ppm at 14.1 T). The value is lower than the MAS speed and the interaction can therefore be removed. Figure 1-4(b) shows that the polarization transfer efficiencies are dependent on the Mn^{2+} - $^1\text{H}/^{13}\text{C}$ distance under paramagnetic anisotropic shift dispersion and the Solomon mechanism. At 2.0 Å, the transfer efficiency of paramagnetic shift dispersion is higher than the transfer efficiency at either 1.0 or 3.0 Å. This result was probably caused by interference of the RF irradiation and dipolar couplings, because the peak was weakened by other Hartmann-Hahn matching conditions (data not shown). The efficiency from paramagnetic anisotropic shift dispersion reaches a plateau before that from Solomon-mechanism is raised. From these results, we conclude that paramagnetic anisotropic shift dispersions can be included in our experimental condition.

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 (Figure 1-1(a)), I_{whole} , is given by:

$$I_{whole,i} = \frac{1}{N} \sum_{\mathbf{p} \in \text{pattern}} I(\Gamma'_{1,C}(i, \mathbf{p}), \Gamma'_{1\rho,H}(i, \mathbf{p}), \Gamma'_{1\rho,C}(i, \mathbf{p}), \Gamma'_{2,C}(i, \mathbf{p})) \quad (1-19)$$

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

$$p_k = \begin{cases} 1, & q \geq s \\ 0, & \text{otherwise} \end{cases} \quad (1-20)$$

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 equation (1-13). $\Gamma'_{1,X}$, $\Gamma'_{1\rho,X}$ 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:

$$\Gamma'(i, \mathbf{p}) = \sum_{j \in A, \dots, O} p_j \Gamma(r_{ij}) \quad (1-21)$$

where j is the chain index of the PDB data ($j \in A, \dots, O$), Γ is the paramagnetic relaxation rates given by equations (1-14) and (1-15), 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-N8Mn²⁺, GB1-E19Mn²⁺, and GB1-T53Mn²⁺). 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 [23]. 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 [7]. 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 [7,24]. 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 d at 18 °C. Finally, the precipitant was collected by centrifugation at 2000 g for 10 min and packed into 3.2 mm Varian standard-wall and JEOL RESONANCE zirconia rotors.

MAS 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 [25]. The pulse sequence used in the experiments is shown in Figure 1-1(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. 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 [8]. TPPM decoupling [26] 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 [27] and analyzed in Sparky [28].

Molecular dynamics simulation

The initial coordinates of GB1-N8, E19 and T53EDTA-Mn²⁺ (Figure 1-5) were generated based on the X-ray structure (PDB ID: 2GI9) using tleap included in AmberTools 15 [29]. The charge of the EDTA tag was assigned by the RESP method [30]. Force field parameters of AMBER ff12SB were applied [31] to the protein and EDTA-based tag. The Mn²⁺ ion was described using

the parameter developed by Bradbrook [32]. A cuboidal solvent box whose edges are 12 Å from the closest GB1 and EDTA-tag atoms was added to each structure. 4611, 4928 and 4295 molecules of TIP3P water were added to each solvent box, respectively [33]. Then, Na⁺ ions were added to the solvent boxes to neutralize each system [34].

MD simulations were performed using GROMACS 5.0.2 [35]. The files of the initial coordinates and force field parameters were converted into GROMACS format using ACPYPE [36]. The initial structures were first energy minimized by steepest descent without any constraints. The 100 ps of NVT (constant particle number, volume and temperature) dynamics run was initially run and followed by the NPT (constant particle number, pressure and temperature) dynamics run for equilibration. During the equilibration, position restraints were applied to the GB1 and EDTA-tag atoms using a force constant of 1000 kJ mol⁻¹ nm⁻². After equilibration, a 50 ns NPT dynamics run was performed. To avoid dissociation of the Mn²⁺-EDTA-tag complex, distance restraints expressed in following equation between Mn²⁺ and the tag were applied:

$$E(r_{ij}) = \begin{cases} \frac{1}{2}k(r_{ij} - r_0)^2 & \text{for } r_{ij} < r_0 \\ 0 & \text{for } r_0 \leq r_{ij} < r_1 \\ \frac{1}{2}k(r_{ij} - r_1)^2 & \text{for } r_1 \leq r_{ij} < r_2 \\ \frac{1}{2}k(r_2 - r_1)(2r_{ij} - r_2 - r_1) & \text{for } r_2 \leq r_{ij} \end{cases} \quad (1-23)$$

where, r_{ij} is the distance between Mn²⁺ and the other atoms and k is a force constant. For Mn²⁺-O2, O4 and O6 restraints, r_0 , r_1 and r_2 were set to 2.0, 2.5 and 3.0 Å, respectively. For Mn²⁺-N2 and N3, those were set to 3.0, 4.0 and 5.0 Å, respectively (Figure 1-5). k was set to 2.5×10^5 kJ mol⁻¹ nm⁻². All dynamics simulations were performed using the following protocol. We used 2 fs time steps with constraints to all bond lengths by the LINCS algorithm [37]. The cut-off of

van der Waals and electrostatic interactions were set to 10.0 Å. The long-range electrostatic interactions were calculated using the particle-mesh Ewald method [38]. The temperature was controlled by the velocity-rescaling thermostat with 0.1 ps of the temperature coupling time constant [39]. The pressure was controlled by the Parrinello-Rahman approach with 2.0 ps of the pressure time constant [40]. The above simulation sequence was performed twice for each initial structure using different random seeds and it was confirmed that the results did not depend on the initial velocities.

Results and Discussion

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 [41,42]. 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 (Figure 1-61-6(a)). 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 1-1. 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.

Figure 1-6(b) 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 (data not shown). Figure 1-7 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 (Figure 1-1(b))) 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 **Evaluation of error from the pseudo paramagnetic center approximation and EDTA-tag flexibility**, Figure 1-13 and Table 1-2) 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 Å. 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 ($>0.6\sim 0.7$) 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⁺ (data not shown). 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.

Our result is very similar to the report of Jaroniec and co-workers using Cu²⁺ [41]. 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 Mn²⁺, Cu²⁺, 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.

Structural restraints from PREs

2D ¹³C-¹³C spectra of GB1-N8, E19 and T53EDTA-M are shown in Figure 1-8. 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 (Figure 1-9). Spectra of these labeled mutants clearly showed that Mn²⁺ bleached out signals arising from some residues. This is because Mn²⁺ 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 Mn²⁺ attached samples.

The absence of particular peaks reflects the GB1 fold. For example, the resonance representing M1 of E19-EDTA-Mn²⁺ was absent, whereas in the N8 and T53-EDTA-Mn²⁺

samples, the resonance of M1 was observed. The M1C α -N8C β , -E19C β and -T53C β distances in the X-ray structure (PDB ID: 2GI9) are 23.3, 5.7 and 17.9 Å, respectively (Figure 1-10(a)). The resonance of G9 in the N8 and T53-EDTA-Mn²⁺ spectra was missing, whereas in the spectrum of the E19-EDTA-Mn²⁺ sample, the resonance of G9 was observed. The G9C α -N8C β , -E10C β and -T53C β distances in the X-ray structure are 4.9, 23.4 and 9.0 Å, respectively (Figure 1-10(b)).

The effect of intermolecular PREs was assessed by examining the peak intensities of G38 and G41 of the GB1-T53EDTA-Mn²⁺ 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 G41C α and the tagged Mn²⁺ 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 [43]) 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 (Figure 1-10(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 1-11 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 Figure 1-12, we compared the observed peak intensities with the calculated values using the X-ray structure and equation (1-13). 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 equation (11) for the strong PREs observed and repulsive restraints defined by equation (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 equation (1-13) 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 expressed by following rectified quadratic functions can suitably handle the low-precision and qualitative PRE information:

$$E_{ub} = \begin{cases} 0 & \text{if } r \leq r_{ub} \\ (r - r_{ub})^2 & \text{if } r_{ub} < r \end{cases} \quad (1-24)$$

$$E_{lb} = \begin{cases} (r - r_{lb})^2 & \text{if } r \leq r_{lb} \\ 0 & \text{if } r_{lb} < r \end{cases} \quad (1-25)$$

E_{ub} and E_{lb} are the upper bound (attractive) and lower bound (repulsive) PRE restraint function, respectively. r is the distance between the C α atom of a PRE affected residue and the C β atom of the EDTA-M modified residue, and r_{ub} and r_{lb} are set to 16.0 and 10.0 Å, respectively.

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 **Evaluation of error from the pseudo paramagnetic center approximation and EDTA-tag flexibility**). A total of 40 upper and 65 lower bound restraints were obtained from N8, E19 and T53EDTA-M samples. Comparisons of the restraints and distances between C α atoms to the pseudo-paramagnetic center on the X-ray structure are shown in Figure 1-13. The restraints agree fully with the distances between the C α atoms to the pseudo-paramagnetic center.

Evaluation of error from the pseudo paramagnetic center approximation and EDTA-tag flexibility

We applied the PRE distance restraints between $C\alpha$ atoms and the $C\beta$ atom of paramagnetic labeled residues (pseudo paramagnetic center), because CS-Rosetta cannot handle the EDTA-M molecule. To evaluate the influence of the approximation, we performed molecular dynamics (MD) simulations of EDTA- Mn^{2+} attached to GB1.

Most of the PRE restraints agreed with the distances between $C\alpha$ atoms and the Mn^{2+} -EDTA-tag, after taking into consideration the error associated with tag flexibility (Figure 1-13 and Table 1-2). However, a minor number of the restraints did not agree with the distances between $C\alpha$ atoms and Mn^{2+} in the EDTA-tag obtained from trajectories of molecular dynamics simulations of EDTA- Mn^{2+} attached GB1. The errors have the potential to cause violations in the calculation. To overcome this problem, an iterative protocol was considered to be useful. For example: (i) CS-Rosetta calculation with PRE restraints; (ii) MD simulations of the lowest scored structure with the EDTA-M tag; and (iii) exclusion of clearly wrong restraints. Recently, a cyclen-type metal ion binding tag that is more compact than the EDTA-tag has been developed [44]. As another approach, introduction of metal-chelating unnatural amino acids whose molecular size is similar to tryptophan is also an attractive option [45]. The options of different tags should also help alleviate the problem.

The average of the standard deviation of $C\alpha$ atoms to Mn^{2+} distances of GB1-N8, E19 and T53EDTA- Mn^{2+} in the MD simulation trajectories reflects the tag flexibility with average standard deviations of 2.9, 2.6 and 0.8 Å, respectively (Table 1-2). In our experiments, PRE restraints were determined to have errors whose widths were > 3 Å. Thus, the impact of the tag flexibilities is negligible.

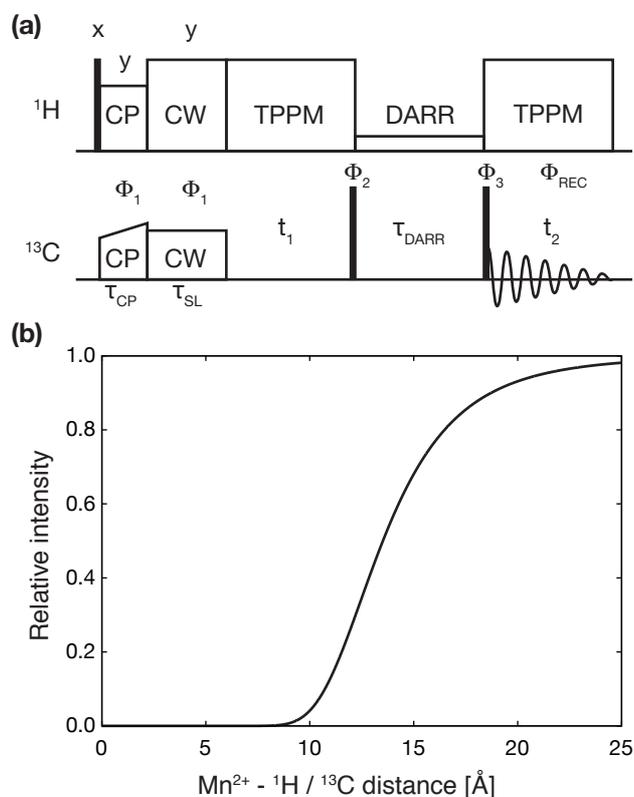


Figure 1-1 (a) The ^{13}C - ^{13}C correlation pulse sequence. The filled rectangles represent $\pi/2$ pulses. The following phase cycling was used: $\Phi_1 = (-y, y)$, $\Phi_2 = (-x, -x, x, x)$, $\Phi_3 = (x, x, x, x, -x, -x, -x, -x)$ and $\Phi_{REC} = (-x, x, x, -x, x, -x, -x, x)$. Quadrature phase 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 equation (1-13). τ_{CP} , τ_{SL} and τ_{DARR} in equation (1-13) were set to 0.32, 1.04 and 20.0 ms, respectively. The intrinsic line width was set to 75 Hz.

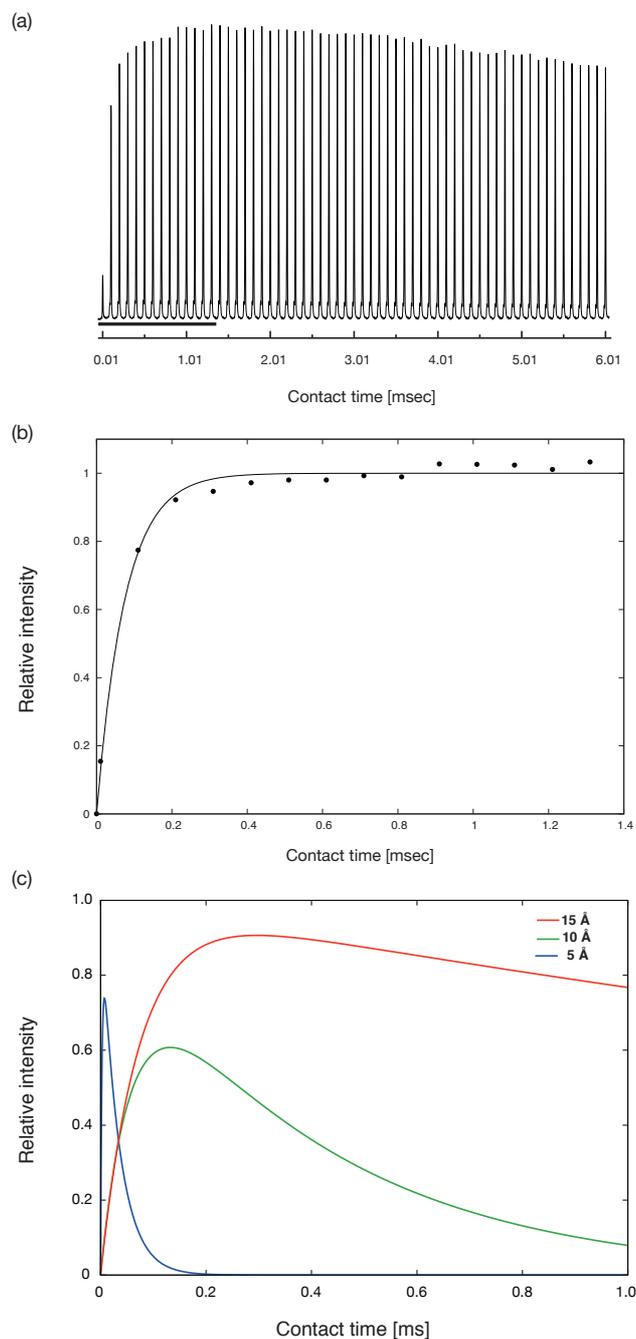


Figure 1-2 (a) C α signals of CPMAS spectra of NAV using various contact times. (b) The C α signal intensities and curve fitting results. The plotted intensities are from the spectra indicated by the solid line in (a). The curve function represents the expression, $I(t) = 1 - \exp(-2R_{IS}t)$, where $R_{IS} = 6.63 \text{ ms}^{-1}$. (c) The magnetization change depends on the nuclei-Mn²⁺ distances represented by equation (1-6).

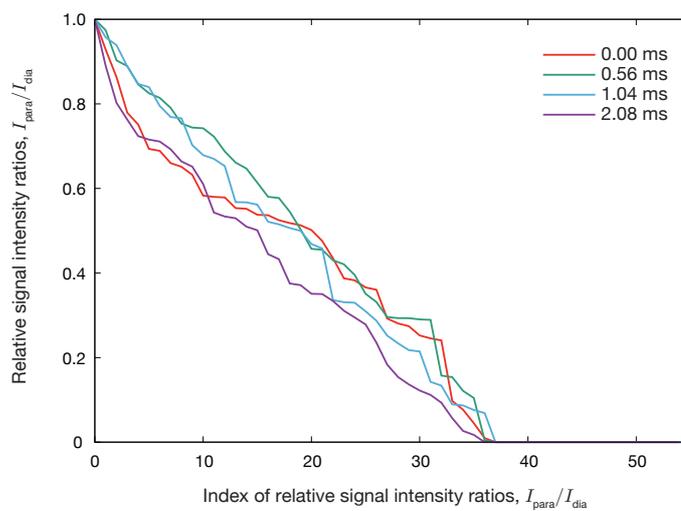


Figure 1-3 ^{13}C spin-locking time dependency on relative peak intensity. The peak intensity ratios of GB1-N8EDTA- Mn^{2+} to Zn^{2+} in each spin-locking time are sorted in descending order and plotted. The ratios are normalized by the maximal values.

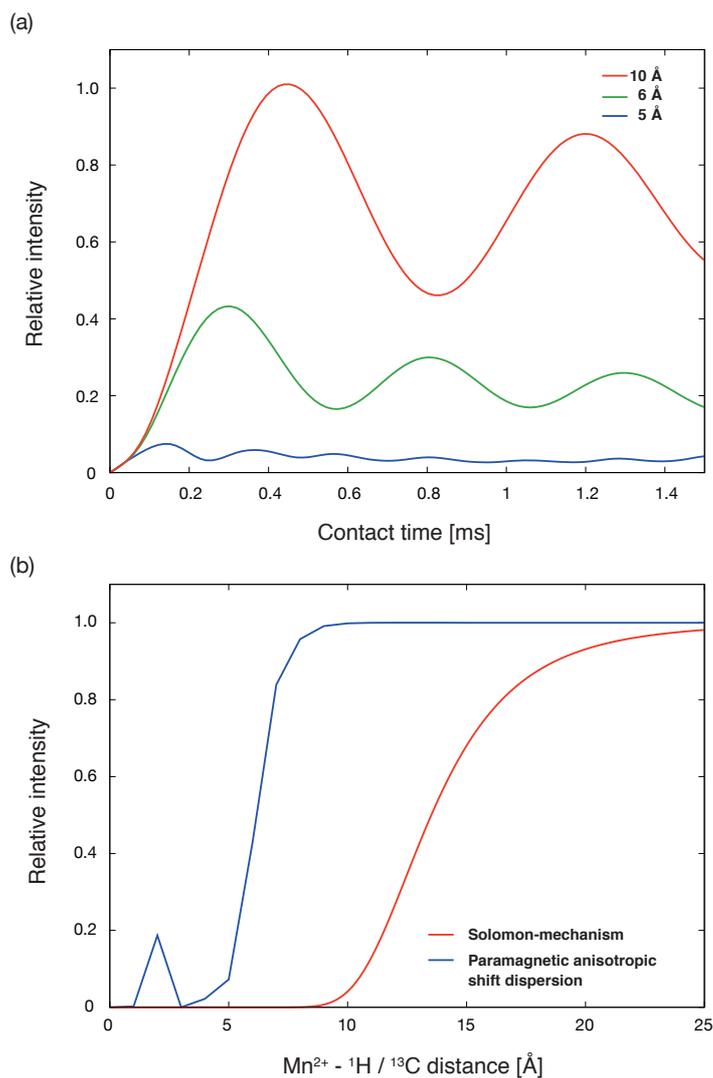


Figure 1-4 (a) Numerical calculation of the ¹H-¹³C CP transfer for an Mn²⁺-¹H-¹³C spin system with 5.0, 6.0 and 10.0 Å distances between the Mn²⁺ and ¹H-¹³C. (b) Relative intensities estimated from spin dynamics calculations of the CP transfer and Solomon mechanism represented by equation (1-13) for the various Mn²⁺-¹H/¹³C distances. The intensities of the CP transfer were normalized by the intensity for an infinite Mn²⁺-¹H/¹³C distance. The plot of the Solomon mechanism is the same as Figure 1-1(b).

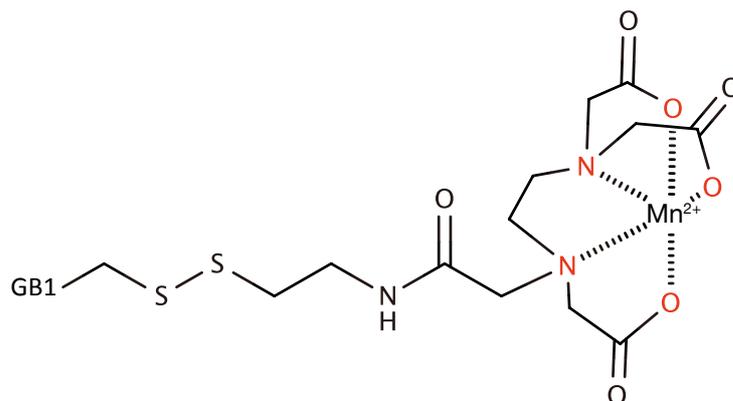


Figure 1-5 Schematic structure of the GB1 attached Mn^{2+} -EDTA-tag complex. The red atoms correspond to O2, O4, O6, N2 and N3 atoms.

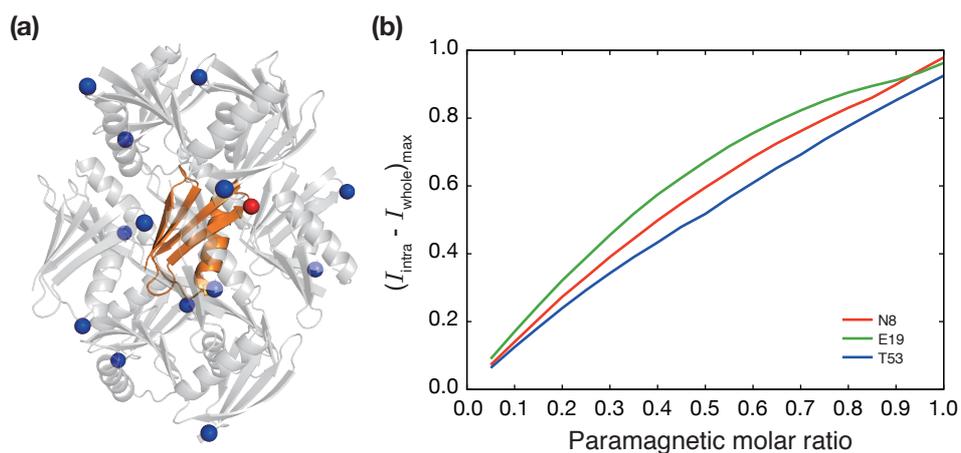


Figure 1-6 (a) Crystal packing of GB1 used in the intermolecular PRE simulation (PDB ID: 2G19). The central monomer used in the intramolecular PRE calculation is colored orange. The red- and blue-colored spheres indicate $\text{C}\beta$ atoms of E19 used in the intramolecular and intermolecular PREs, 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.

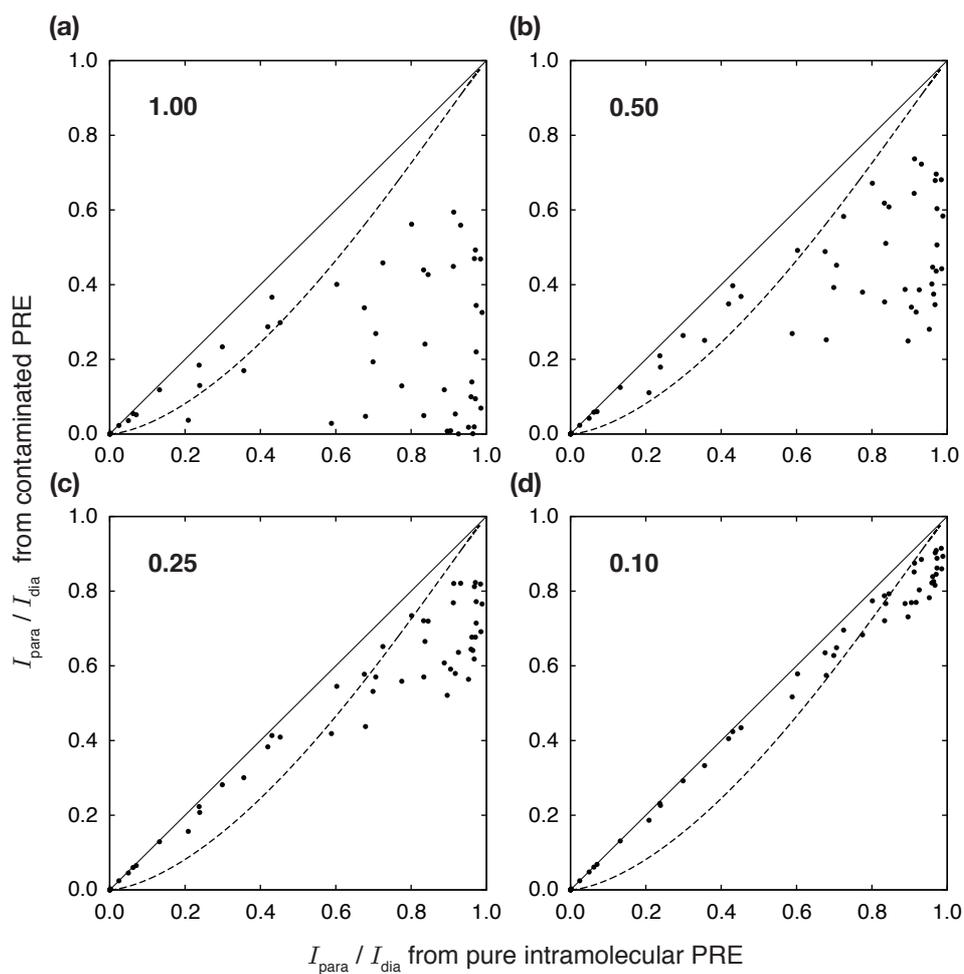


Figure 1-7 Comparison of $I_{\text{para}}/I_{\text{dia}}$ of GB1-E19Mn²⁺ in the absence and presence of intermolecular PREs for paramagnetic molar ratios of (a) 1.0, (b) 0.5, (c) 0.25 and (d) 0.1. The dashed lines indicate an error of 1 Å between the pure intramolecular and contaminated PREs.

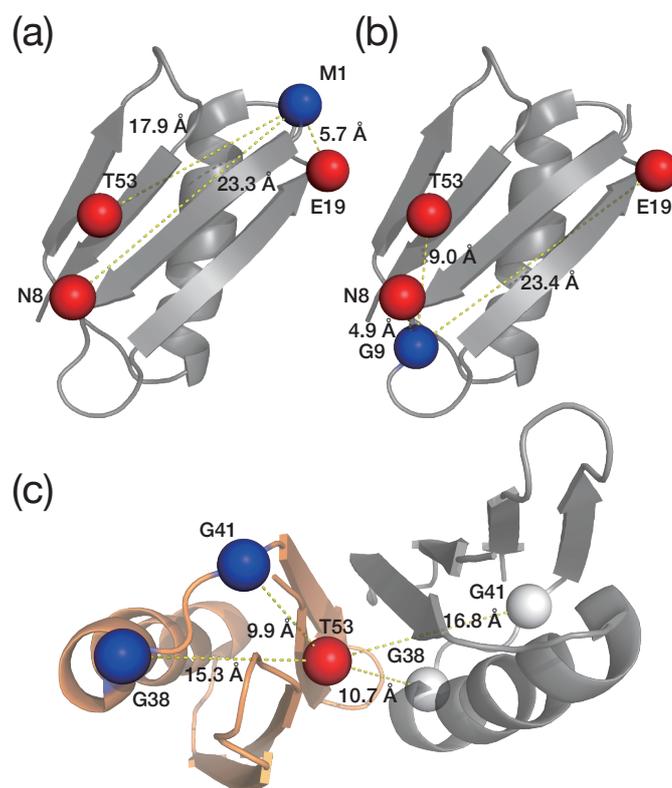


Figure 1-10 (a), (b) X-ray structures of GB1 monomer (PDB ID: 2GI9) and (c) its solid-state NMR quaternary structure (2KWD, only chain A and B are shown and presented in orange and gray, respectively). Red spheres indicate C β atoms of EDTA-M attached residues. Blue and white spheres indicate the positions of the C α atoms that were used to describe intra- and intermolecular PREs in the text, respectively.

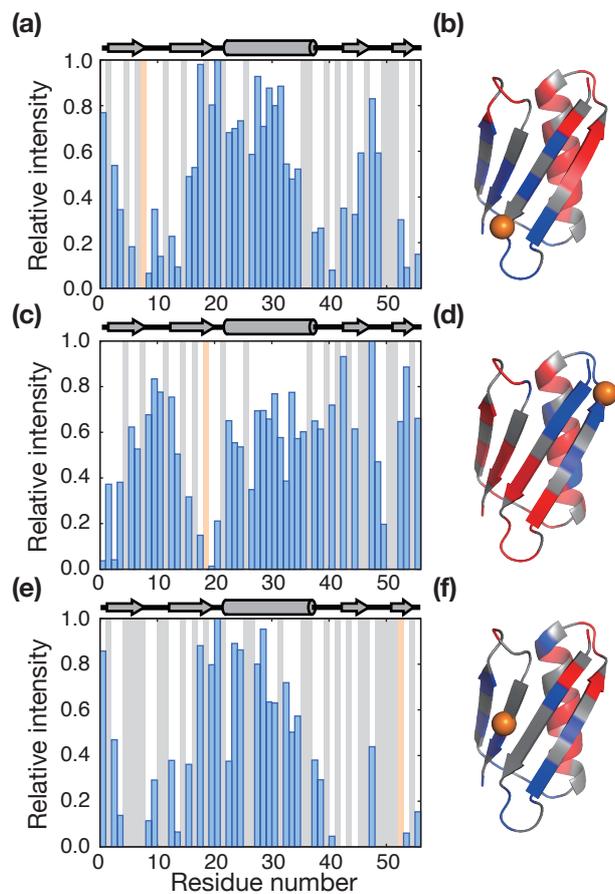


Figure 1-11 (a), (c), (e) C α -C' peak intensity ratios of N8, E19 and T53EDTA-Mn²⁺ to Zn²⁺, respectively. The orange bars indicate EDTA-M attached residues. The gray bars indicate overlapped peaks. (b), (d), (e) Crystal structure of GB1 (PDB ID: 2G19). The peak intensity ratios of red colored residues are > 0.4, whereas those of blue colored residues are < 0.4

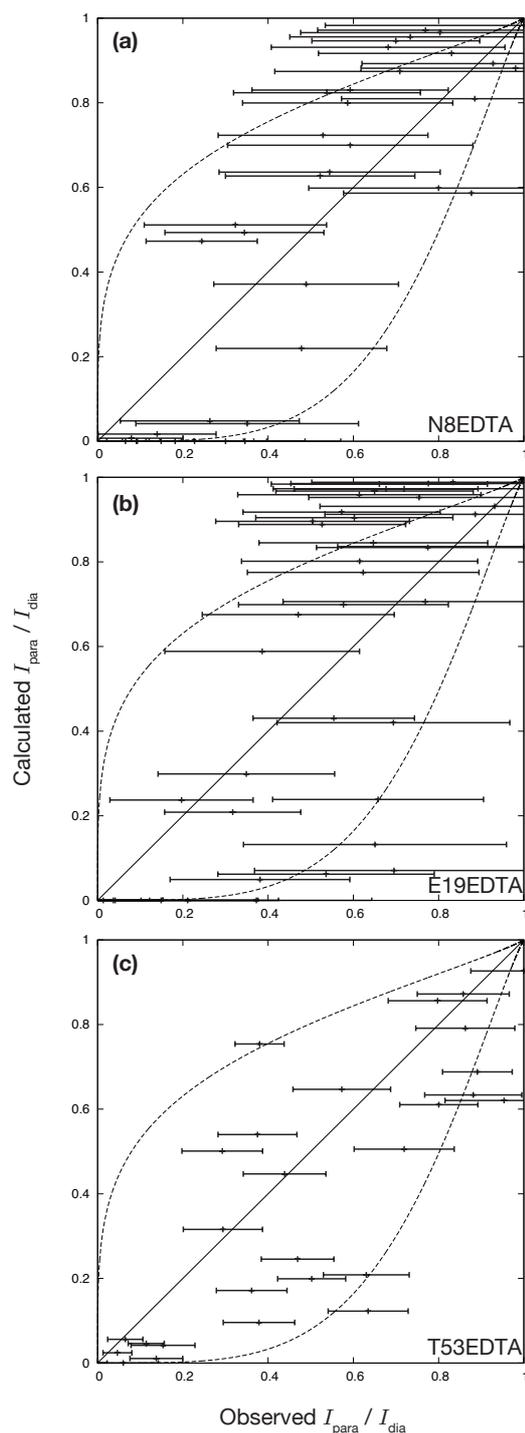


Figure 1-12 Comparison of $I_{\text{para}}/I_{\text{dia}}$ of (a) GB1-N8EDTAMn²⁺, (b) GB1-E19EDTAMn²⁺ and (c) GB1-T53EDTAMn²⁺ to the calculated values from the C α -pseudo paramagnetic center distances in the X-ray structure (PDB: 2GI9) using equation (1-13). The dashed line indicates the $\pm 3 \text{ \AA}$ error threshold between observed and calculated signal decays. The error bars were calculated from the spectral noise.

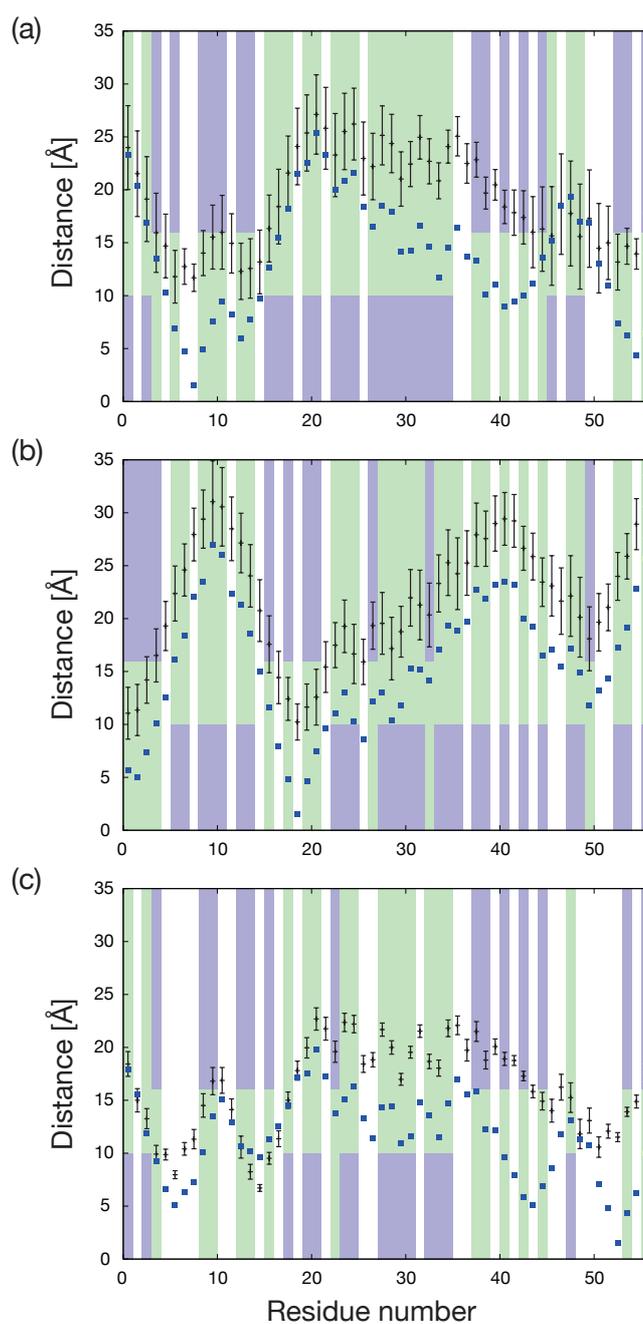


Figure 1-13 Summary of PRE restraints, C α to pseudo-paramagnetic center distances in the X-ray structure (PDB ID: 2GI9) and C α to Mn²⁺ distances in the trajectory of the MD simulations of (a) GB1-N8, (b) E19 and (c) T53-EDTA-Mn²⁺. The green and violet bars indicate the restraint function of zero and quadratic regions, respectively. The blue squared dots indicate the distance between C α and the pseudo-paramagnetic center. The black colored crosses with error-bar indicate the average and standard deviation of the distances between C α and Mn²⁺ on the MD simulation trajectory.

Table 1-1. Summary of the distance distributions of GB1 (PDB ID: 2GI9) and its crystal neighbors. The distances between C α atoms in the central monomer and the pseudo-paramagnetic center in crystal neighbors are summarized.

Chain	N8Mn ²⁺		E19Mn ²⁺		T53Mn ²⁺	
	Min. ¹	Max. ²	Min.	Max.	Min.	Max.
B	11.7	27.6	19.3	36.0	17.1	32.7
C	23.2	45.0	23.9	39.3	18.1	38.9
D	5.7	27.8	22.7	49.7	8.4	33.8
E	19.0	47.7	18.9	45.3	18.8	46.8
F	19.9	45.6	28.8	55.3	19.2	46.0
G	19.0	41.5	6.7	27.9	18.0	39.7
H	26.7	46.9	11.9	31.8	20.8	42.2
I	11.8	35.6	29.0	50.2	17.2	39.0
J	11.6	29.4	23.5	44.8	9.5	28.4
K	20.8	48.2	12.4	36.1	18.9	45.5
L	15.5	40.7	16.3	44.0	12.1	38.5
M	24.7	49.8	9.7	29.5	22.0	44.5
N	25.3	40.3	24.7	48.6	27.2	44.1
O	21.7	40.6	16.9	33.3	15.4	35.6

^{1,2}The minimal and maximal distance between C α atoms of in the central monomer and the pseudo-paramagnetic center of the crystal neighbor, respectively.

Table 1-2 Summary of the chemical shifts, peak intensity ratios and C α to paramagnetic center distances of GB1-N8, E19 and T53EDTA-M.

GB1-N8EDTA-M

Residue	Chemical Shifts		I_{para}/I_{dia}^1	Distance	
	C α	C'		X-ray ²	MD ³
M1	54.1	171.2	0.769	23.3	24.0 \pm 3.96
Y3	56.9	174.6	0.538	16.9	19.1 \pm 4.01
K4	54.8	173.1	0.344	13.5	15.9 \pm 3.74
I6	59.9	174.8	0.182	6.91	11.8 \pm 2.49
G9	44.8	173.3	0.065	4.91	14.0 \pm 2.13
K10	58.7	179.3	0.346	7.52	15.5 \pm 3.02
T11	61.8	173.6	0.140	9.45	16.0 \pm 3.49
K13	53.3	175.8	0.227	6.01	12.3 \pm 2.66
G14	44.7	171.4	0.093	7.71	12.6 \pm 2.81
T16	60.1	171.8	0.489	12.7	16.3 \pm 3.16
T17	60.0	174.2	0.529	15.5	18.4 \pm 3.53
T18	61.2	170.9	0.981	18.2	21.6 \pm 3.50
A20	50.6	177.7	0.803	22.5	25.3 \pm 3.60
V21	63.2	174.8	1.000	25.3	27.1 \pm 3.75
A23	54.6	179.6	0.682	20.0	23.3 \pm 3.93
A24	54.6	180.9	0.700	20.9	25.5 \pm 3.61
T25	67.2	175.6	0.734	21.6	26.2 \pm 3.39
E27	59.2	177.8	0.587	16.5	22.2 \pm 3.14
K28	60.2	178.9	0.928	18.5	25.1 \pm 2.79
V29	66.3	178.8	0.709	18.0	24.3 \pm 2.76
F30	57.3	179.0	0.878	14.2	21.0 \pm 2.57
K31	60.1	179.6	0.800	14.3	22.4 \pm 2.13
Q32	58.9	177.4	0.885	16.6	25.0 \pm 2.06
Y33	61.5	178.7	0.544	14.6	22.7 \pm 2.13
A34	56.0	179.5	0.478	11.7	20.9 \pm 1.70
N35	57.0	179.4	0.522	14.5	24.1 \pm 1.56

G38	46.9	173.9	0.244	13.3	22.8±1.65
V39	61.7	175.1	0.263	10.1	19.7±1.51
G41	45.1	172.8	0.079	9.01	18.4±1.58
W43	57.5	177.1	0.351	10.0	17.4±2.52
Y45	57.5	172.2	0.323	13.6	16.3±3.99
D46	50.8	176.3	0.593	15.2	15.6±4.66
A48	54.1	179.5	0.830	19.3	17.8±4.96
T49	60.1	175.6	0.592	17.0	15.6±4.96
T53	60.3	172.1	0.300	7.36	13.2±2.64
V54	58.1	172.4	0.091	6.23	14.7±1.70
E56	56.9	180.3	0.149	6.39	15.9±1.88

GB1-E19EDTA-M

Residue	Chemical Shifts		I_{para}/I_{dia}^1	Distance	
	C α	C'		X-ray ²	MD ³
M1	53.8	171.2	0.037	5.68	11.1±2.45
Q2	55.8	174.7	0.373	5.04	11.4±2.42
Y3	57.0	175.0	0.041	7.36	14.2±2.20
K4	54.7	173.4	0.381	10.1	16.5±2.51
I6	59.8	175.1	0.623	16.1	22.4±2.60
L7	54.8	175.0	0.527	18.4	24.6±2.47
G9	44.6	172.9	0.678	23.4	29.4±2.75
K10	59.3	178.9	0.833	27.0	31.0±3.87
T11	61.9	173.4	0.776	26.0	30.6±3.71
K13	53.3	175.7	0.734	21.3	27.1±2.81
G14	44.7	171.3	0.504	18.6	24.0±2.95
T16	59.9	171.6	0.317	11.6	17.6±2.69
T18	61.1	172.0	0.149	4.82	12.4±2.03
A20	50.5	177.6	0.013	4.59	11.6±2.18
V21	63.4	174.7	0.211	7.50	12.6±2.64
A23	54.6	179.3	0.651	11.0	17.5±2.13
A24	54.6	181.1	0.554	13.0	19.3±2.48

T25	67.2	175.6	0.536	10.3	16.7±2.80
E27	59.2	177.8	0.349	12.2	19.3±2.23
K28	60.2	178.8	0.694	13.0	19.6±2.92
V29	66.4	178.4	0.695	10.4	17.2±2.96
F30	57.3	179.1	0.658	11.8	18.8±2.40
K31	60.1	179.8	0.769	15.3	22.0±2.67
Q32	58.8	177.5	0.577	15.2	21.3±3.29
Y33	61.7	178.5	0.386	14.2	20.3±3.01
A34	56.0	179.5	0.775	17.1	23.3±2.70
N35	57.3	179.5	0.573	19.4	25.3±3.10
D36	55.6	175.6	0.602	18.9	24.2±3.39
G38	46.8	173.7	0.650	22.8	27.9±2.99
V39	61.7	175.0	0.615	21.8	27.6±2.61
G41	45.1	172.5	0.719	23.5	29.4±2.49
W43	57.5	177.0	0.932	20.0	26.6±2.08
Y45	57.5	172.2	0.615	16.5	23.4±2.29
A48	53.8	179.0	1.000	17.1	22.1±3.81
T49	60.1	175.5	0.471	15.0	20.1±3.77
T53	60.3	171.9	0.647	19.2	24.0±2.26
V54	58.4	172.3	0.886	19.2	28.9±2.41
E56	57.3	179.9	0.661	25.6	30.6±3.01

GB1-T53EDTA-M

Residue	Chemical shift		I_{para}/I_{dia}^1	Distance	
	C α	C'		X-ray ²	MD ³
M1	54.2	171.4	0.858	17.9	18.4±1.17
Y3	57.0	175.0	0.469	11.8	13.3±0.95
K4	54.6	173.1	0.138	9.21	9.92±0.82
G9	44.6	173.0	0.114	10.1	14.5±1.11
K10	59.3	179.0	0.292	13.5	16.8±1.29
K13	53.5	175.7	0.379	10.7	10.6±1.04
G14	44.7	171.5	0.065	10.2	8.25±0.70

T16	60.1	171.9	0.361	11.3	9.52±0.56
T18	61.4	171.1	0.881	14.6	15.0±0.78
A20	50.6	177.7	0.798	17.5	20.0±0.96
V21	63.4	175.0	1.000	19.8	22.7±1.05
A23	54.6	179.6	0.375	16.3	19.6±1.00
A24	54.5	181.0	0.891	15.1	22.4±0.87
T25	67.3	175.7	0.863	16.3	22.2±0.82
K28	60.2	178.7	0.800	14.4	21.7±0.62
V29	66.3	178.7	0.954	14.4	20.0±0.63
F30	57.2	179.1	0.635	10.9	17.0±0.56
K31	60.1	179.4	0.630	11.6	19.5±0.56
Y33	61.6	178.6	0.719	13.4	18.7±0.68
A34	56.1	179.5	0.502	11.5	18.0±0.76
N35	57.0	179.3	0.573	14.7	21.8±0.79
G38	46.8	174.0	0.380	15.8	21.5±0.94
V39	62.0	175.0	0.294	12.3	18.8±0.84
G41	45.0	172.6	0.046	9.65	18.9±0.84
W43	57.6	176.6	0.000	5.83	17.3±0.45
Y45	56.9	172.2	0.000	6.86	14.9±0.83
A48	54.0	179.5	0.439	13.2	15.3±1.40
V54	57.4	172.5	0.060	4.38	13.9±0.43
E56	57.5	180.4	0.153	10.0	17.6±0.84

¹ C α -C' peak intensity ratios of N8, E19 and T53EDTA-Mn²⁺ to Zn²⁺. The ratios were normalized using the maximal value.

² C α to pseudo-paramagnetic center distances on X-ray structure (PDB ID: 2GI9)

³ C α to Mn²⁺ distances on trajectory of MD simulation

References

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

- [2] N. Bloembergen, Fine structure of the magnetic resonance line of protons in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, *Physica*. 16 (1950) 95–112. doi:10.1016/0031-8914(50)90067-X.
- [3] S.H. Koenig, 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 (1982) 441–453. doi:10.1016/0022-2364(82)90211-6.
- [4] J.R. Lewandowski, Advances in solid-state relaxation methodology for probing site-specific protein dynamics, *Acc. Chem. Res*. 46 (2013) 2018–2027. doi:10.1021/ar300334g.
- [5] A. Pines, M.G. Gibby, J.S. Waugh, Proton-enhanced NMR of dilute spins in solids, *J. Chem. Phys*. 59 (1973) 569–590. doi:10.1063/1.1680061.
- [6] N.P. Wickramasinghe, Y. Ishii, 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 (2006) 233–243. doi:10.1016/j.jmr.2006.05.008.
- [7] P.S. Nadaud, J.J. Helmus, S.L. Kall, C.P. Jaroniec, 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 (2009) 8108–8120. doi:10.1021/ja900224z.
- [8] K. Takegoshi, S. Nakamura, T. Terao, ^{13}C – ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR, *Chem Phys Lett*. 344 (2001) 631–637. doi:10.1016/S0009-2614(01)00791-6.
- [9] M. Mehring, Principles of high resolution NMR in solids, Second, Revised and Enlarged Edition, Springer-Verlag Berlin Heidelberg New York, 1983. doi:10.1007/978-3-642-68756-3.
- [10] J.L. Battiste, G. Wagner, 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 (2000) 5355–5365. doi:10.1021/bi000060h.
- [11] I. Bertini, C. Luchinat, G. Parigi, Solution NMR of paramagnetic molecules: applications to metalloproteins and models, Elsevier, Amsterdam, 2001.
- [12] P.A. Kosen, Spin labeling of proteins, *Meth. Enzymol*. 177 (1989) 86–121. doi:10.1016/0076-6879(89)77007-5.

- [13] C.P. Grey, N. Dupré, NMR Studies of Cathode Materials for Lithium-Ion Rechargeable Batteries, *Chem. Rev.* 104 (2004) 4493–4512. doi:10.1021/cr020734p.
- [14] A.D. Mackerell Jr., M. Feig, C.L. Brooks III, Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J. Comput. Chem.* 25 (2004) 1400–1415. doi:10.1002/jcc.20065.
- [15] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B.* 102 (1998) 3586–3616. doi:10.1021/jp973084f.
- [16] M. Bak, J.T. Rasmussen, N.C. Nielsen, SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy, *J Magn Reson.* 147 (2000) 296–330. doi:10.1006/jmre.2000.2179.
- [17] A. Nayeem, J.P. Yesinowski, Calculation of magic-angle spinning nuclear magnetic resonance spectra of paramagnetic solids, *J. Chem. Phys.* 89 (1988) 4600–4608. doi:10.1063/1.454800.
- [18] T.H. Yeom, C. Rudowicz, S.H. Choh, Comparative electron paramagnetic resonance studies of the s-state ions in BiVO₄ crystals, *J. Korean Phys. Soc.* 41 (2002) 756–762.
- [19] M. Bak, N.C. Nielsen, REPULSION, a novel approach to efficient powder averaging in solid-state NMR, *J Magn Reson.* 125 (1997) 132–139. doi:10.1006/jmre.1996.1087.
- [20] V.B. Cheng, H.H. Suzukawa Jr, M. Wolfsberg, Investigations of a nonrandom numerical method for multidimensional integration, *J. Chem. Phys.* 59 (1973) 3992–3999. doi:10.1063/1.1680590.
- [21] H. Conroy, Molecular Schrödinger Equation. VIII. A New Method for the Evaluation of Multidimensional Integrals, *J. Chem. Phys.* 47 (1967) 5307–5318. doi:10.1063/1.1701795.
- [22] S.K. Zaremba, Good lattice points, discrepancy, and numerical integration, *Ann Mat Pup Appl.* 73 (1966) 293–317. doi:10.1007/BF02415091.
- [23] J. Li, K.B. Pilla, Q. Li, Z. Zhang, X. Su, T. Huber, et al., Magic angle spinning NMR structure determination of proteins from pseudocontact shifts, *J. Am. Chem. Soc.* 135 (2013) 8294–8303. doi:10.1021/ja4021149.
- [24] W.T. Franks, D.H. Zhou, B.J. Wylie, B.G. Money, D.T. Graesser, H.L. Frericks, et al., Magic-Angle Spinning Solid-State NMR Spectroscopy of the β 1 Immunoglobulin Binding Domain of Protein G (GB1): ¹⁵N and ¹³C Chemical Shift Assignments and

- Conformational Analysis, *J. Am. Chem. Soc.* 127 (2005) 12291–12305.
doi:10.1021/ja044497e.
- [25] C.R. Morcombe, K.W. Zilm, Chemical shift referencing in MAS solid state NMR, *J Magn Reson.* 162 (2003) 479–486.
- [26] A.E. Bennett, C.M. Rienstra, M. Auger, K.V. Lakshmi, R.G. Griffin, Heteronuclear decoupling in rotating solids, *J. Chem. Phys.* 103 (1995) 6951–6958.
doi:10.1063/1.470372.
- [27] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR.* 6 (1995) 1–17. doi:10.1007/BF00197809.
- [28] T.D. Goddard, D.G. Kneller, SPARKY 3, (n.d.).
<https://www.cgl.ucsf.edu/home/sparky/> (accessed June 17, 2015).
- [29] D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, I. T E Cheatham, T.A. Darden, et al., AmberTools 15, [Httpambermd.org](http://ambermd.org). (n.d.).
- [30] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J Phys Chem.* 97 (1993) 10269–10280. doi:10.1021/j100142a004.
- [31] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins.* 65 (2006) 712–725. doi:10.1002/prot.21123.
- [32] G.M. Bradbrook, T. Gleichmann, S.J. Harrop, J. Habash, J. Raftery, J. Kalb Gilboa, et al., X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding, *J Chem Soc Faraday Trans.* 94 (1998) 1603–1611. doi:10.1039/a800429c.
- [33] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935. doi:10.1063/1.445869.
- [34] J. Aqvist, Ion-water interaction potentials derived from free energy perturbation simulations, *J Phys Chem.* 94 (1990) 8021–8024. doi:10.1021/j100384a009.
- [35] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics.* 29 (2013) 845–854. doi:10.1093/bioinformatics/btt055.

- [36] A.W. Sousa da Silva, W.F. Vranken, ACPYPE - AnteChamber PYthon Parser interfacE, *BMC Res Notes*. 5 (2012) 367. doi:10.1186/1756-0500-5-367.
- [37] B. Hess, H. Bekker, H. Berendsen, F. JGEM, LINCS: A Linear Constraint Solver for Molecular Simulations, *J Comput Chem*. 18 (1997) 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
- [38] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092. doi:10.1063/1.464397.
- [39] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007) 014101–8. doi:10.1063/1.2408420.
- [40] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J Appl Phys*. 52 (1981) 7182–7190. doi:10.1063/1.328693.
- [41] P.S. Nadaud, I. Sengupta, J.J. Helmus, C.P. Jaroniec, 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 (2011) 293–302. doi:10.1007/s10858-011-9536-y.
- [42] I. Sengupta, P.S. Nadaud, J.J. Helmus, C.D. Schwieters, C.P. Jaroniec, Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy, *Nat Chem*. 4 (2012) 410–417. doi:10.1038/nchem.1299.
- [43] A.J. Nieuwkoop, C.M. Rienstra, Supramolecular Protein Structure Determination by Site-Specific Long-Range Intermolecular Solid State NMR Spectroscopy, *J. Am. Chem. Soc.* 132 (2010) 7570–7571. doi:10.1021/ja100992y.
- [44] I. Sengupta, M. Gao, R.J. Arachchige, P.S. Nadaud, T.F. Cunningham, S. Saxena, et al., Protein structural studies by paramagnetic solid-state NMR spectroscopy aided by a compact cyclen-type Cu(II) binding tag, *J Biomol NMR*. 61 (2015) 1–6. doi:10.1007/s10858-014-9880-9.
- [45] S.H. Park, V.S. Wang, J. Radoicic, A.A. De Angelis, S. Berkamp, S.J. Opella, 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 (2015) 185–196. doi:10.1007/s10858-014-9884-5.

Part II
Protein structure determination using qualitative PRE restraints

Abstract

It is trivial that the precision of structural restraints is highly important to determine precise protein structure. However, the precision of distance restraints provided by PRE is limited in solid-state NMR because of incomplete averaged interactions and intermolecular PREs. This is a large barrier for the application of PRE restraints for the structure determination. In this part, the backbone structure of 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. In addition, the structure calculations using simulated PRE restraints also suggest this approach is transferable to the backbone structure determination of a protein whose size is at least ca. 110 residues.

Introduction

It is trivial that the precision of structural restraints is highly important to determine precise protein structure [1]. The low precision of PRE distance restraints derived from solid-state NMR is a large barrier to application of those restraints for the structure calculation. 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. Cu²⁺-¹⁵N longitudinal PREs bring the high precision distance restraints [2], however this approach is also time consuming owing to the slow ¹⁵N longitudinal relaxation rate. Thus, a method that can use low precision distance restraints is attractive for protein structure analysis by solid-state NMR.

In this part, structure determination protocol from the low precision PRE distance restraints

combining CS-Rosetta is described. Using the CS-Rosetta fragment assembly [3], 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 [4,5]. NMR researchers have been extending the Rosetta protocol, for example, fragment picking restrained by backbone chemical shifts [3], conformational sampling combined with NOE restraints [6] and automated NOE assignment [7]. 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 [8]. Here, I demonstrate the backbone structure determination of the EDTA-Mn²⁺ complex attached B1 domain of streptococcal protein G (GB1) by Rosetta fragment assembly using backbone chemical shifts and relaxation profiles of ¹³C-¹³C correlation spectra by short DARR mixing [9]. In addition, I discuss how general and transferable this protocol by structure calculations of three proteins having 76-108 amino acid residues from simulated data set.

Methods

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 (Table 2-1). 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 ¹⁵N chemical shifts were ignored. The C α , C β and C' chemical shifts of Y52 were calculated

from chemical shifts of F52 (BMRB entry) and random coil shifts of phenylalanine and tyrosine [10]. 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 [11]. 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 equation (1-24) and (1-25). r_{ub} and r_{lb} of those equations 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 500 lowest-energy structures from the low-resolution CS-Rosetta structure pool that was derived from structures solved with and without PRE restraints were used for all-atom refinement calculations as the initial structure. The refinement protocol followed the default protocol of *Relax* in Rosetta 3.5 [11]. The weight factor of PRE restraints was set to 2.0. Ten refined structures were generated from one centroid model. The re-scoring by chemical shifts [3] was not performed.

Conventional structural calculations based on restrained molecular dynamics and simulated annealing were performed using Xplor-NIH [12]. 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+ [13] 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 [14]. Dihedral angles that were not classified as *good* were ignored. A thousand structures were calculated.

CS-Rosetta calculation for simulated data of proteins

We produced simulated distance restraints of human ubiquitin, the N-terminal domain of human N60D calmodulin and *E. coli* thioredoxin (referred to as ubiquitin, N-calmodulin and thioredoxin, respectively) from entries in the BMRB database or published data, and the protein structures from the PDB. The number of tag molecules was assumed to be four for ubiquitin and N-calmodulin, and six for thioredoxin. This number is derived from the GB1 result that three tags

were required for the determining correctly the 56-residue GB1 fold (i.e., 0.054 tags per residue are required). The labeling scheme is illustrated in Figure 2-9(a,b,c). To make the situation more realistic, simulations of spectral resolution anticipated from 2D-NCA spectra that provide intra-residue correlations were performed using logic similar to that described previously [15]. Briefly, the peak i in N or C α axes is considered *resolved* from peak j ($j \neq i$), if $|\Omega_{i,m} - \Omega_{j,m}| > lw_m$, where m is an N or C α atom, lw_m is the line width of all peaks in the m axis, and $\Omega_{i,m}$ and $\Omega_{j,m}$ are the chemical shifts of peaks i and j in the m axis. Further, if peak i is resolved in any axis from all peaks $j \neq i$, then peak i is considered to be resolved. lw_m was set to 1.0 ppm for both N and C α . The unassigned residues were ignored. The resolved peaks were used for production of the simulated PRE distance restraints. The simulated PRE distances were measured between the C β atom of potentially labeled residues and the backbone C α atoms, and classified into the upper and lower bound restraints by:

$$E = \begin{cases} E_{ub} & \text{if } p > P(r) \\ E_{lb} & \text{if } p \leq P(r) \end{cases} \quad (2-1)$$

where, p is a uniformly distributed random number on the interval [0, 1], r is distance between the C β atom of potentially EDTA-M attached residues and the backbone C α atom, and $P(r)$ is the cumulative distribution function of normal distribution whose average and standard deviation are 12.5 and 1.0 Å, respectively. $P(r)$ is expressed as:

$$P(r) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{r - 12.5}{\sqrt{2}} \right) \right] \quad (2-2)$$

For example, in the case of $r = 10.0, 12.5$ and 15.0 Å, the PRE restraint is classified into the upper bound restraint by the probability of 99.3, 50.0 and 0.7%, respectively. The details of the data set for structure calculation are summarized in Table 2-4. The structure calculations were performed in identical manner to the GB1 calculation. The all-atom refinement calculations were not

performed.

Results

Structure calculation of GB1

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 (Figure 2-1). 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 2-4(b) 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 (Figure 2-2(a)) 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 (Figure 2-2(b)). 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.

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 2-3 presents the largest C α 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.

We performed low-resolution CS-Rosetta calculations with the PRE restraints. Figure 2-4(a) presents the energy profiles of the CS-Rosetta calculations. Figure 2-5 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 2-1). Figure 2-6 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 2-4(c,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 2-2, Figure 2-7). Moreover, the Rosetta scores of the high $C\alpha$ RMSD structures derived from the structure calculations with PRE restraints were shifted upward when compared with non-restrained calculations, suggesting the high $C\alpha$ RMSD structures were destabilized by the PRE restraints. These observations illustrate the ability of the PRE restraints to bias sampling towards the native structure.

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 results are summarized in Table 2-3. Compared

with the results of the structure refinements with and without PRE restraints, the C α -RMSD of the lowest scored structure was improved when the PRE restrained initial structures were used. On the other hand, when using the non-restrained centroid modeling structures as the initial structure, the C α -RMSD of the lowest scored structure was slightly deteriorated, however the difference was quite small. The energy profiles and RMSD histograms show that the conformational sampling space of the refinement calculation is also reduced by the PRE restraints (Figure 2-8). Thus, the PRE restraints impede sampling of high C α -RMSD structures; however, they do not affect the low C α -RMSD structures in the refinement calculations. This is probably because the shapes of PRE scoring terms have large flat bottoms. Convergences were degraded by the PRE restraints. We speculate that the restraints prevent the Monte-Carlo structure refinement around the native fold.

CS-Rosetta calculation for simulated data of proteins

The simulated paramagnetic labeling schemes are illustrated in Figure 2-9(a, b, c). PRE violation profiles of CS-Rosetta calculations without PRE restraints are shown in Figure 2-10. The funnel shapes are similar to GB1. The maximal C α RMSDs of ubiquitin, N-calmodulin and thioredoxin structures with no PRE violations are 3.6, 7.1 and 3.7 Å, respectively. The value of N-calmodulin is relatively higher than the others and is due to the different arrangement of the N-terminal helix when compared with the native fold. The C α RMSD without the N-terminal helix and loop (sequence 29 to 79) is 3.6 Å. Thus, the overall structures are similar to their native forms. These results suggest that the purely upper and lower bound restraints can sufficiently identify the native structure from the CS-Rosetta structure pools up to a protein size of *ca.* 110 residues.

The results of the CS-Rosetta calculations with PRE restraints are shown in Figure 2-9(d,e,f) and Table S4. In all cases, the C α RMSD of the lowest score, the lowest RMSD from the structure pool and convergence are improved from the calculations without PRE restraints. In addition, bias sampling toward the native structure was performed (Figure 2-11, 2-12 and 2-13). In the cases of ubiquitin and thioredoxin, the Rosetta scores of the high C α RMSD structure were

shifted upward and the C α RMSDs of the lowest score structure were reduced when compared with the non-restrained calculations. These are similar to the results obtained with GB1. However, for N-calmodulin, obvious upward shifting of the Rosetta score was not observed. This is because the Rosetta score function unduly assessed its native structure (Rosetta scores using *score3* weights set [4] of the native and lowest score structure of non-restrained CS-Rosetta calculation were 26.859 and -10.994, respectively). However, the C α RMSD of the lowest score structure and the conformational sampling bias toward the native structure are improved. These effects indicate that PRE restraints correct the unduly assessment of the Rosetta score function.

Our results of the GB1 structure calculation using experimentally obtained restraints and those simulations clearly show that PRE restraints from solid-state NMR can guide sampling towards the native structure and help Rosetta fragment assembly calculations.

Discussion

We found that combinations of qualitative PRE restraints from paramagnetic Mn²⁺ 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 [3,4]. 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 C α RMSD structure were shifted upward and C α RMSDs of the lowest score structure decreased when compared with non-restrained calculations (Figure 2-4). Similar changes have been reported to occur in CS-Rosetta calculations with RDC restraints [16]. 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. 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 Cu²⁺-¹⁵N longitudinal PREs to improve structure determination quantitatively [2]. In addition, they successfully determined the backbone structure of GB1 from those PREs and dihedral angle constraints calculated by TALOS+ [14]. 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 ¹⁵N longitudinal relaxation rates, the inversion-recovery method is mandatory for gaining sensitivity by CP. The ¹⁵N longitudinal rates of EDTA-Zn²⁺ attached GB1 are 0.03–0.2 s⁻¹ [2]. 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 ¹³C. 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 [14]. 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: (i) the conformational sampling limitation of the CS-Rosetta protocol [16], and (ii) increasing the molecular size raises the likely problem of severe peak overlap. The former problem can be solved by the RASREC approach [6]. RASREC CS-Rosetta can determine solution structures of proteins up to 40 kDa using limited NOE and RDC restraints [8]. To consider the latter problem, we performed structure calculations using simulated PRE restraints. Our results suggest that the backbone structure of a protein whose size is at least *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 [17]. 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 [18,19]. 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 [20]. 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 [21], selective labeling by cell-free expression systems and proton detection under the ultra-fast MAS condition.

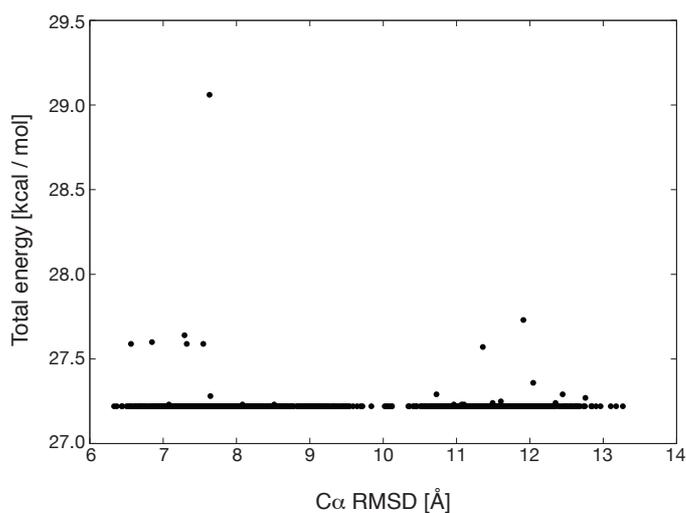


Figure 2-1 Energy landscape generated by restrained molecular dynamics using Xplor-NIH. The 1,000 structures were calculated with PRE and dihedral restraints. The number of Φ and Ψ dihedral angle restraints are both 49.

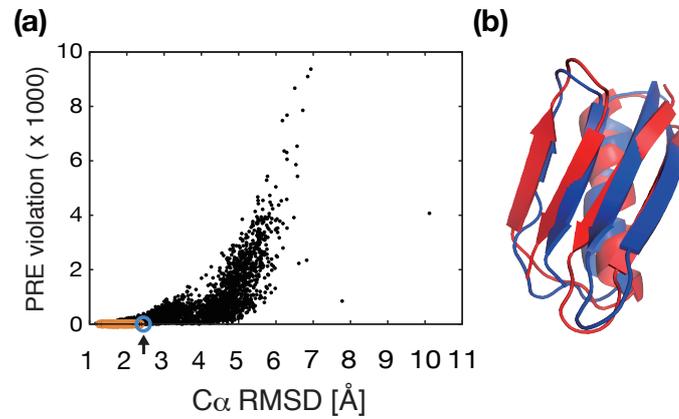


Figure 2-2 (a) PRE violation score from three EDTA-M tags versus the C α RMSD to the X-ray structure of GB1 (PDB ID: 2GI9). The structures without any violations are marked with orange and blue colored circles. The blue circle (indicated by an arrow) indicates the largest RMSD structure. (b) The X-ray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

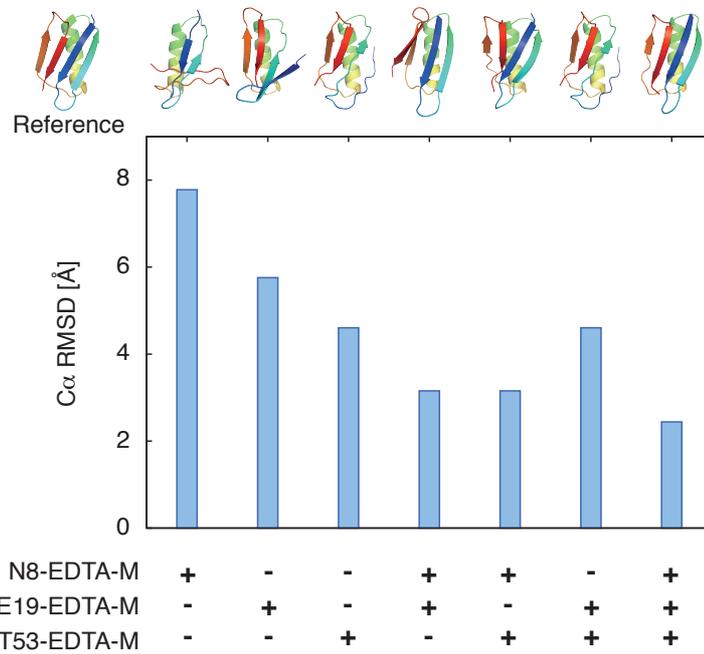


Figure 2-3 The largest RMSD without any PRE violation and its structure with different combinations of restraints from three EDTA-M tags. The color map of the structure indicates the residue number.

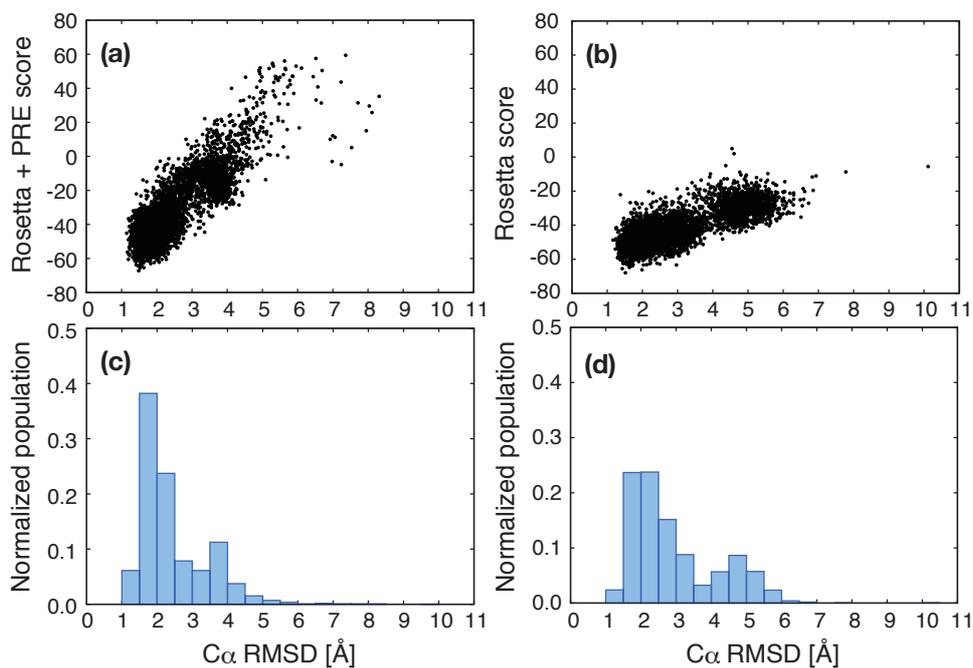


Figure 2-4 (a), (b) Energy landscapes generated by CS-Rosetta with and without PRE restraints, respectively. (c), (d) The distribution of the C α RMSD with and without PRE restraints.



Figure 2-5 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.

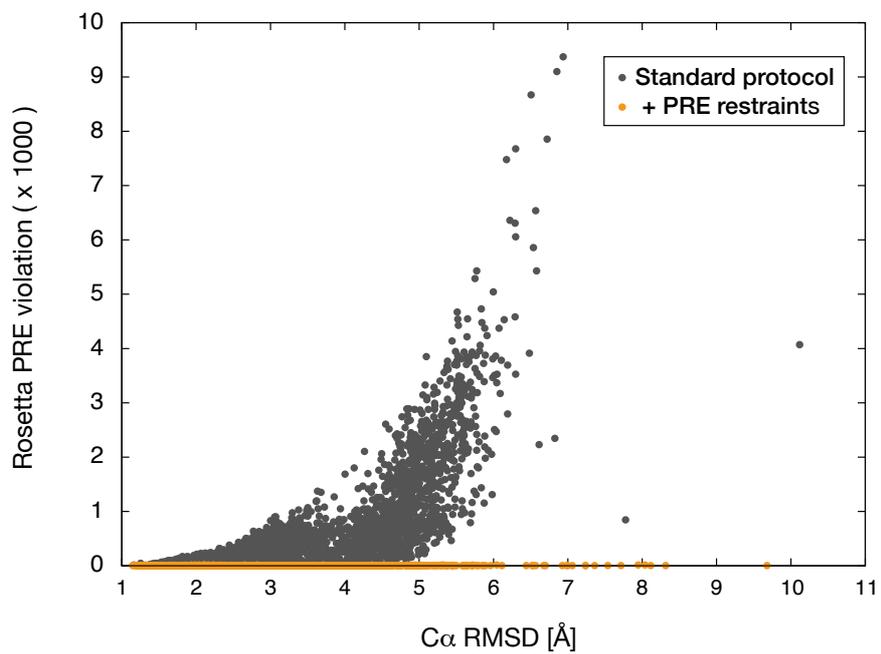


Figure 2-6 The PRE violation profile of the CS-Rosetta structure calculation of GB1. The gray and orange dots indicate the profile with and without PRE restraints.

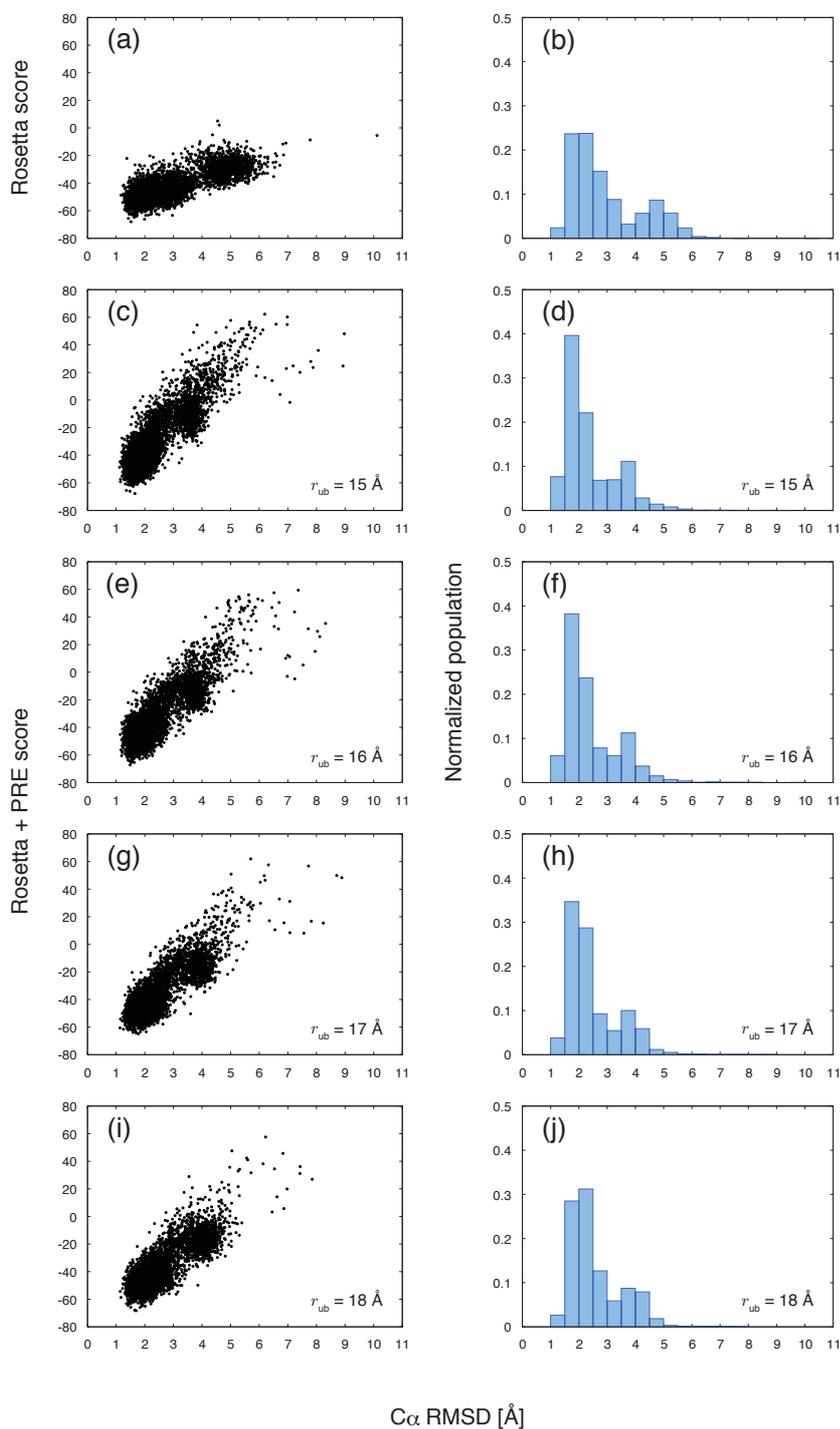


Figure 2-7 (a), (c), (e), (g), (i) Energy landscapes generated by CS-Rosetta without PRE restraints, with $r_{ub} = 15, 16, 17$ and 18 Å for the PRE restraints, respectively. (b), (d), (f), (h), (j) The distribution of the C α RMSD without PRE restraints, with $r_{ub} = 15, 16, 17$ and 18 Å for the PRE restraints, respectively.

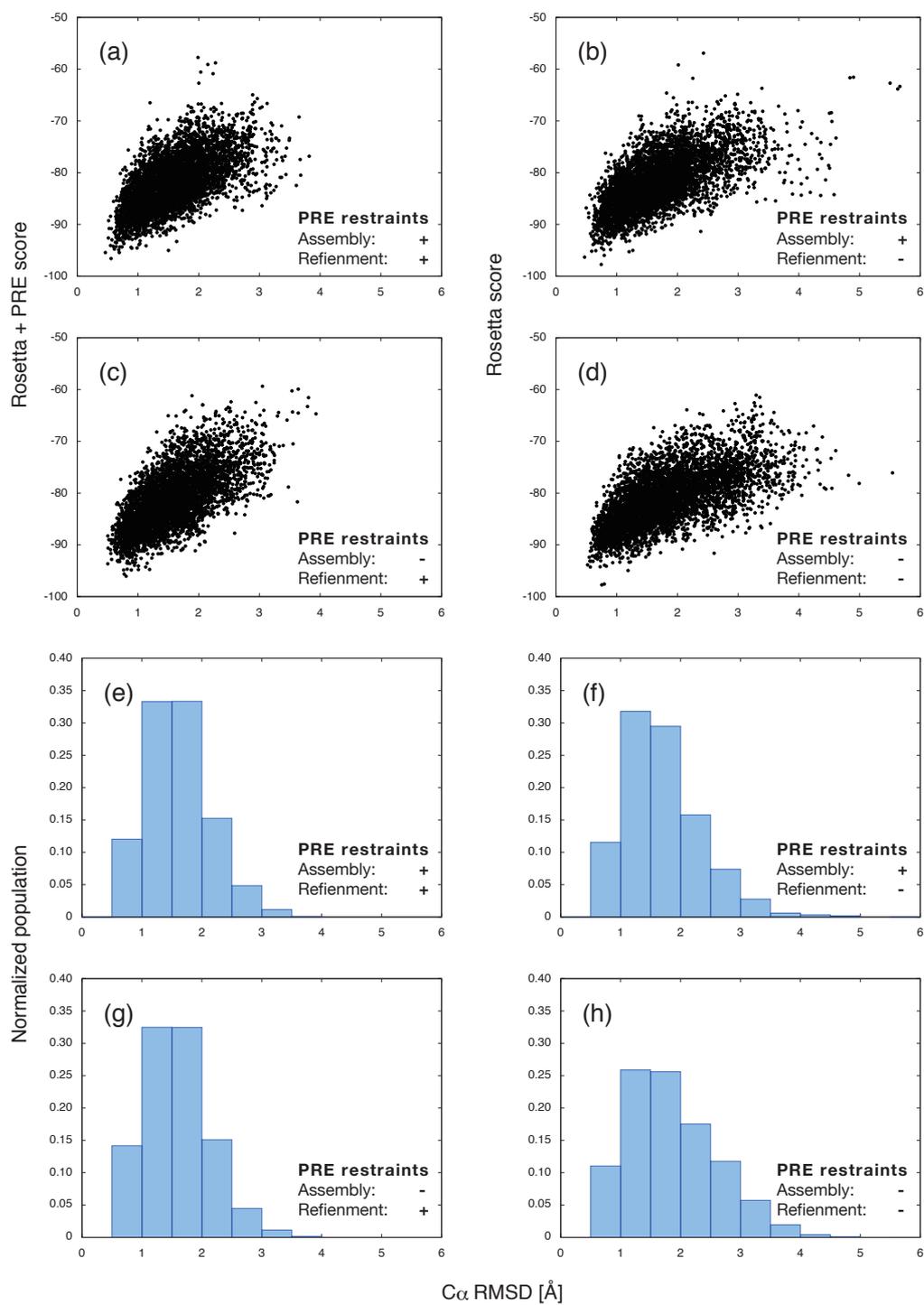


Figure 2-8 (a), (b), (c), (d) Energy landscapes and (e), (f), (g), (h) The distribution of the $C\alpha$ RMSD generated by Rosetta all-atom refinement with and without PRE restraints in each step.

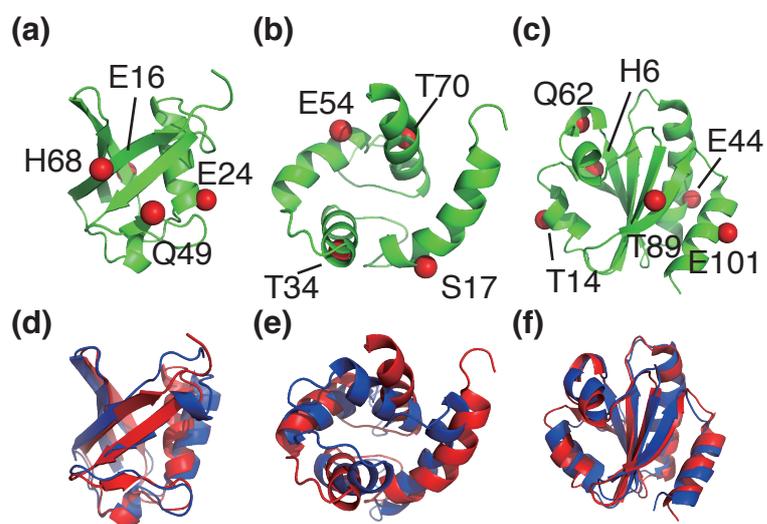


Figure 2-9 (a), (b), (c) Labeling scheme used for the structure calculations of ubiquitin, N-calmodulin and thioredoxin, respectively. Labeling sites are indicated by the red spheres. (d), (e), (f) Results of the structure calculations. The reference structures are illustrated in red. The CS-Rosetta structures with PRE restraints are illustrated in blue.

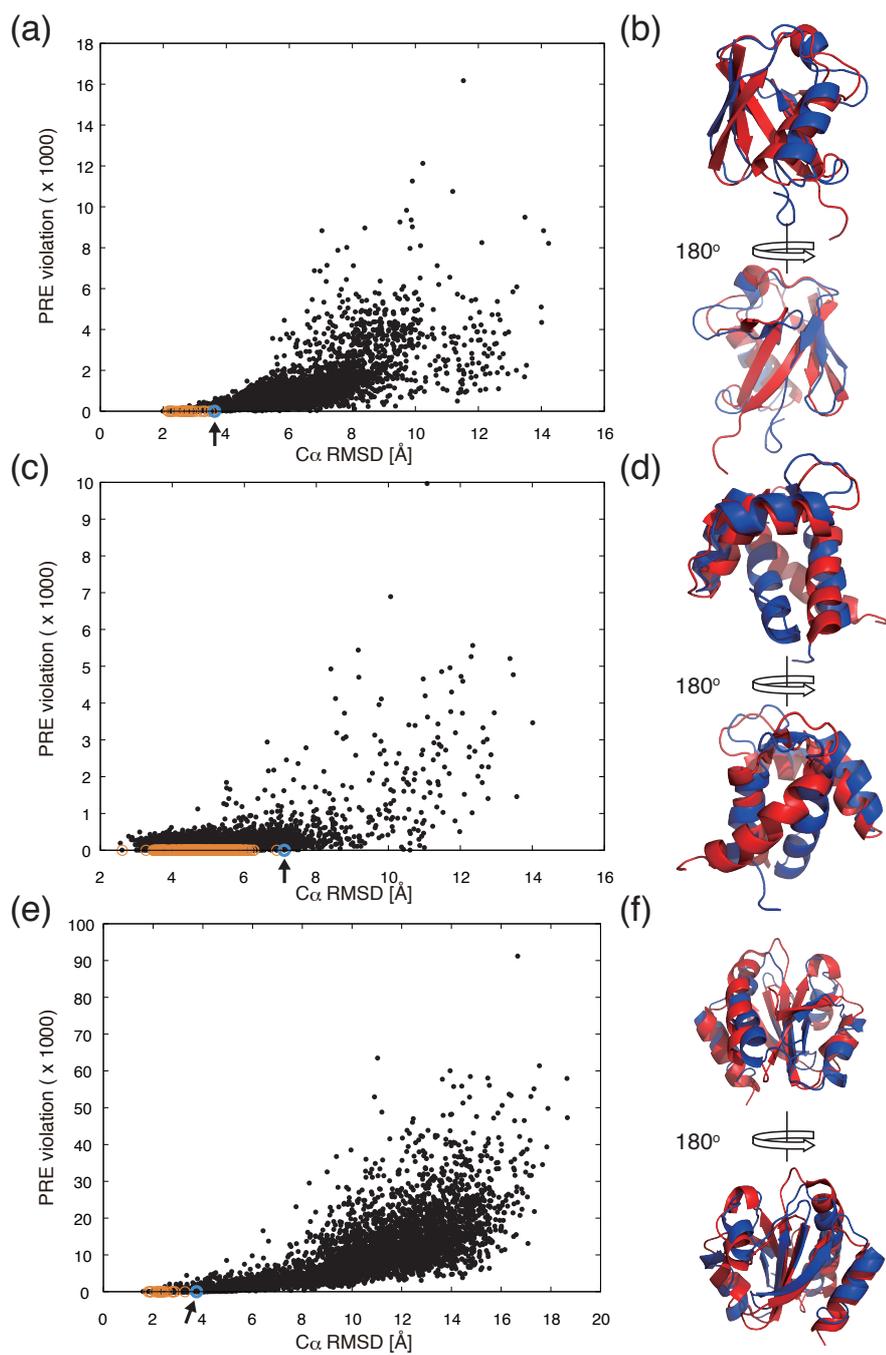


Figure 2-10 (a), (c), (e) PRE violation score from EDTA-M tags versus the C α -RMSD to the reference structure of ubiquitin, N-calmodulin and thioredoxin. The structures without any violations are marked with orange and blue circles. The blue circles (indicated by arrows) indicate the largest RMSD structures. (b), (d), (f) The reference structures of ubiquitin, N-calmodulin and thioredoxin are shown in red. The structures indicated by the arrows in (a), (c), (e) are shown in blue.

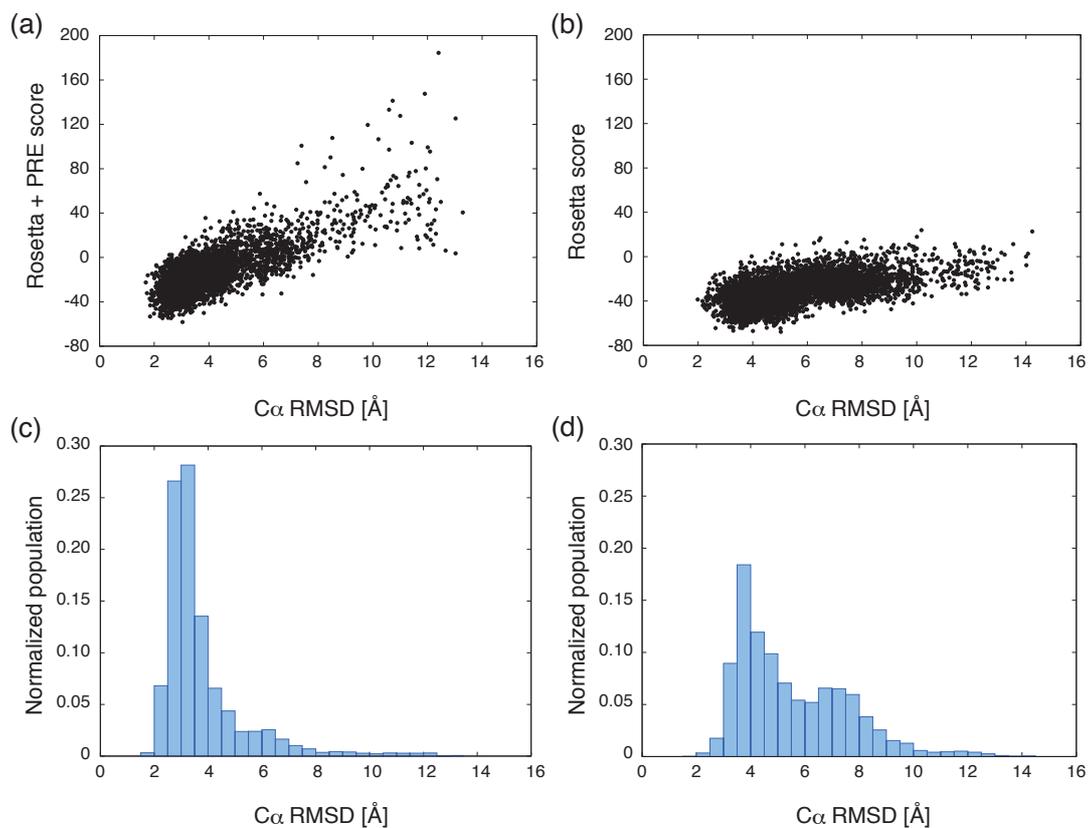


Figure 2-11 (a), (b) Energy landscapes generated by the CS-Rosetta calculation of ubiquitin with and without PRE restraints, respectively. (c), (d) The distribution of the C α RMSD to the X-ray structure with and without PRE restraints, respectively.

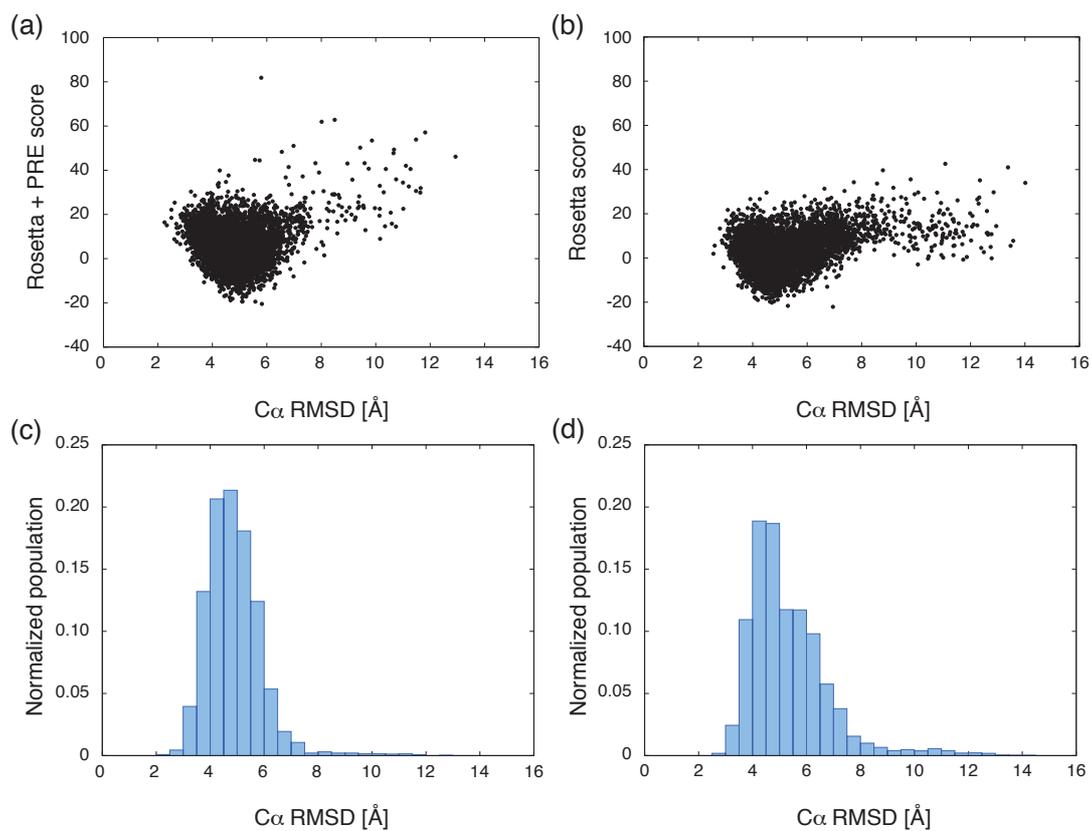


Figure 2-12 (a), (b) Energy landscapes generated by the CS-Rosetta calculation of N-calmodulin with and without PRE restraints, respectively. (c), (d) The distribution of the C α RMSD to the solution NMR structure with and without PRE restraints, respectively.

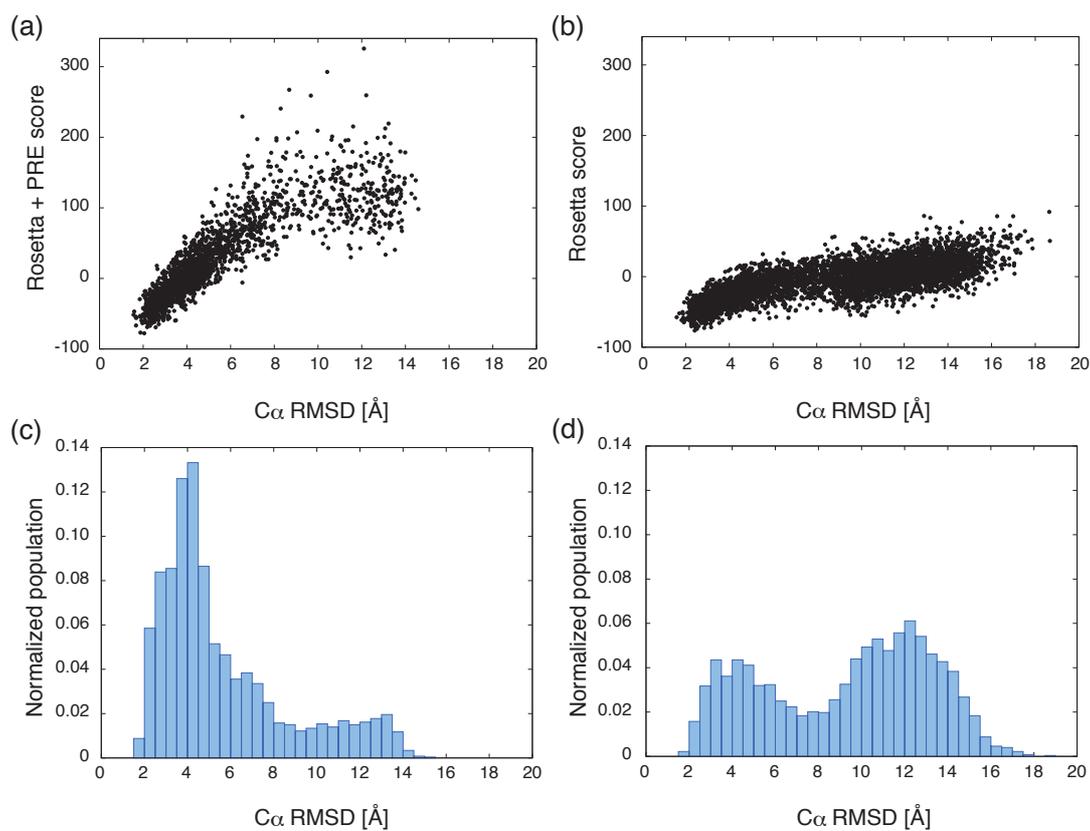


Figure 2-13 (a), (b) Energy landscapes generated by the CS-Rosetta calculation of thioredoxin with and without PRE restraints, respectively. (c), (d) The distribution of the C α RMSD to the X-ray structure with and without PRE restraints, respectively.

Table 2-1 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

Table 2-2 CS-Rosetta calculation with various r_{ub} value

r_{ub} ¹	N/A ⁵	15	16	17	18
RMSD ²	1.52	1.66	1.49	1.78	1.70
Lowest ³	1.17	1.12	1.14	1.14	1.14
Convergence ⁴	1.74	1.15	1.22	1.52	1.17

¹Distance of upper-bound restraints in angstroms

²The Ca RMSD of the lowest structure in angstroms

³The lowest Ca RMSD in the structure pool in angstroms

⁴Average Ca RMSD calculated between the lowest score structure and the next four lowest score structure in angstroms

⁵Results without PRE restraints.

Table 2-3 Summaries of Rosetta refinement results

Restrains¹	+/+	+/-	-/+	-/-
RMSD²	0.558	0.747	0.806	0.756
Convergence³	0.801	0.833	0.646	0.592

¹Combination of whether PRE restraints were used in the fragment assembly and refinement steps. For example, +/- indicates that PRE restraints were used in fragment assembly and not used in refinement steps.

²The Ca RMSD of the lowest score structure in angstroms

³Average Ca RMSD calculated between the lowest score structure and the next four lowest score structure in angstroms

Table 2-4 Summaries of CS-Rosetta calculation of simulated data.

Targets	PDB	N _{res} ¹	N _{peak} ²	N _{resolv} ³	N _{tag} ⁴	N _{ub} ⁵	N _{lb} ⁶	SS ⁷	With PREs			Without PREs			CS ¹¹
									RMSD ⁸	Lowest ⁹	Convergence ¹⁰	RMSD	Lowest	Convergence	
Ubiquitin	1ubq	76	70	26	4	37	67	-	3.04	1.69	2.78	5.03	1.99	3.28	BMRB5387
N-Calmodulin	1sw8 ¹²	79	77	20	4	29	49	-	5.82	2.25	3.69	6.95	2.56	5.27	Reference*
Thioredoxin	2trx ¹³	108	106	44	6	58	205	³²⁻³⁵	2.05	1.56	1.98	2.42	1.59	3.19	BMRB17700

¹The number of residues of the target protein

²The number of peaks those were used in simulation of spectral resolution (see Materials and Methods)

³The number of resolved peaks in simulation of spectral resolution (see Materials and Methods)

⁴The number of tag molecules

⁵The number of the upper bound PRE restraints

⁶The number of the lower bound PRE restraints

⁷The pairs of disulfide bond restraints.

⁸The Ca RMSD of the lowest score structure in angstroms

⁹The lowest Ca RMSD in the structure pool in angstroms

¹⁰Average Ca RMSD calculated between the lowest score structure and the next four lowest score structure in angstroms

¹¹Reference to source of chemical shifts

¹² The structure of model 1 was used for generation of simulated PRE restraints.

¹³ The structure of chain A was used for generation of simulated PRE restraints.

* The backbone chemical shifts of N, C α and C β were obtained from ref. [22]. C' chemical shifts were not used for the fragment picking

References

- [1] B. Vögeli, S. Kazemi, P. Güntert, R. Riek, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, *Nat Struct Mol Biol.* 19 (2012) 1053–1057. doi:10.1038/nsmb.2355.
- [2] P.S. Nadaud, J.J. Helmus, S.L. Kall, C.P. Jaroniec, 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 (2009) 8108–8120. doi:10.1021/ja900224z.
- [3] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G. Liu, et al., Consistent blind protein structure generation from NMR chemical shift data, *Proc Natl Acad Sci USA.* 105 (2008) 4685–4690. doi:10.1073/pnas.0800256105.
- [4] C.A. Rohl, C.E.M. Strauss, K.M.S. Misura, D. Baker, Protein structure prediction using Rosetta, *Meth. Enzymol.* 383 (2004) 66–93. doi:10.1016/S0076-6879(04)83004-0.
- [5] P. Bradley, Toward High-Resolution de Novo Structure Prediction for Small Proteins, *Science.* 309 (2005) 1868–1871. doi:10.1126/science.1113801.
- [6] O.F. Lange, D. Baker, Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation, *Proteins.* 80 (2011) 884–895. doi:10.1002/prot.23245.
- [7] O.F. Lange, Automatic NOESY assignment in CS-RASREC-Rosetta, *J Biomol NMR.* 59 (2014) 147–159. doi:10.1007/s10858-014-9833-3.
- [8] O.F. Lange, P. Rossi, N.G. Sgourakis, Y. Song, H.-W. Lee, J.M. Aramini, et al., Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples, *Proc. Natl. Acad. Sci. U.S.a.* 109 (2012) 10873–10878. doi:10.1073/pnas.1203013109.
- [9] K. Takegoshi, S. Nakamura, T. Terao, ^{13}C - ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR, *Chem Phys Lett.* 344 (2001) 631–637. doi:10.1016/S0009-2614(01)00791-6.

- [10] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, ^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 (1995) 67–81. doi:10.1007/BF00227471.
- [11] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompson, R. Jacak, et al., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules, *Meth. Enzymol.* 487 (2011) 545–574. doi:10.1016/B978-0-12-381270-4.00019-6.
- [12] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, *J Magn Reson.* 160 (2003) 65–73. doi:10.1016/S1090-7807(02)00014-9.
- [13] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol NMR*. 44 (2009) 213–223. doi:10.1007/s10858-009-9333-z.
- [14] I. Sengupta, P.S. Nadaud, J.J. Helmus, C.D. Schwieters, C.P. Jaroniec, Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy, *Nat Chem.* 4 (2012) 410–417. doi:10.1038/nchem.1299.
- [15] W.T. Franks, K.D. Kloepper, B.J. Wylie, C.M. Rienstra, Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins, *J Biomol NMR*. 39 (2007) 107–131. doi:10.1007/s10858-007-9179-1.
- [16] S. Raman, O.F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, et al., NMR Structure Determination for Larger Proteins Using Backbone-Only Data, *Science*. 327 (2010) 1014–1018. doi:10.1126/science.1183649.
- [17] T. Maly, G.T. Debelouchina, V.S. Bajaj, K.-N. Hu, C.-G. Joo, M.L. Mak Jurkauskas, et al., Dynamic nuclear polarization at high magnetic fields, *J. Chem. Phys.* 128 (2008) 052211–19. doi:10.1063/1.2833582.
- [18] L. Shi, E.M.R. Lake, M.A.M. Ahmed, L.S. Brown, V. Ladizhansky, Solid-state NMR study of proteorhodopsin in the lipid environment: Secondary structure and dynamics, *BBA - Biomembranes*. 1788 (2009) 2563–2574. doi:10.1016/j.bbamem.2009.09.011.
- [19] M. Eitzkorn, S. Martell, O.C. Andronesi, K. Seidel, M. Engelhard, M. Baldus, Secondary Structure, Dynamics, and Topology of a Seven-Helix Receptor in Native Membranes, Studied by Solid-State NMR Spectroscopy, *Angew. Chem. Int. Ed.* 46 (2007) 459–462. doi:10.1002/anie.200602139.

- [20] V. Agarwal, S. Penzel, K. Székely, R. Cadalbert, E. Testori, A. Oss, et al., De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy, *Angew. Chem. Int. Ed. Engl.* 53 (2014) 12253–12256. doi:10.1002/anie.201405730.
- [21] S.H. Park, V.S. Wang, J. Radoicic, A.A. De Angelis, S. Berkamp, S.J. Opella, 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 (2015) 185–196. doi:10.1007/s10858-014-9884-5.
- [22] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, et al., Experimentally exploring the conformational space sampled by domain reorientation in calmodulin, *Proc. Natl. Acad. Sci. U.S.a.* 101 (2004) 6841–6846. doi:10.1073/pnas.0308641101.

Conclusions

In this thesis, I 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. I 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 [1-3]. I believe that my protocol should offer an approach to alleviate the bottleneck of structure determination in solid-state NMR.

References

- [1] B.B. Das, H.J. Nothnagel, G.J. Lu, W.S. Son, Y. Tian, F.M. Marassi, et al., Structure Determination of a Membrane Protein in Proteoliposomes, *J. Am. Chem. Soc.* 134 (2012) 2047–2056. doi:10.1021/ja209464f.
- [2] S.H. Park, B.B. Das, F. Casagrande, Y. Tian, H.J. Nothnagel, M. Chu, et al., Structure of the chemokine receptor CXCR1 in phospholipid bilayers, *Nature*. 491 (2012) 779–783. doi:10.1038/nature11580.
- [3] B.J. Wylie, L.J. Sperling, A.J. Nieuwkoop, W.T. Franks, E. Oldfield, C.M. Rienstra, Ultrahigh resolution protein structures using NMR chemical shift tensors, *Proc. Natl. Acad. Sci. U.S.a.* 108 (2011) 16974–16979. doi:10.1073/pnas.1103728108.

Acknowledgements

I gratefully acknowledge giving the chance to study solid-state NMR, invaluable suggestions and support of Professor Makoto Demura, Professor Toshimichi Fujiwara (Institute for Protein Research, Osaka University) and Dr. Ayako Egawa (Institute for Protein Research, Osaka University).

Advice and comments given by Dr. Tomoshi Kameda (Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology) has been a great help in the molecular dynamics simulation.

I would also like to express my gratitude to Professor Keiichi Kawano, Associate Professor Tomoyasu Aizawa, Lecturer Takashi Kikukawa, Assistant Professor Masakatsu Kamiya, Dr. Yasuhiro Kumaki, Dr. Yuki Onishi, Assistant Professor Yoh Matsuki (Institute for Protein Research, Osaka University) and Assistant Professor Keisuke Ikeda (Graduate School of Medicine and Pharmaceutical Science, University of Toyama) for their fruitful suggestions and support.

I also deeply appreciate the efforts of Professor Min Yao in reviewing this thesis.

I am deeply grateful to Kouki Kido for his help to start-up this project.

This work was supported by a JSPS KAKENHI Grant Number 26-4060 and performed under the Collaborative Research Program of the Institute for Protein Research, Osaka University.

I also deeply thank my colleagues for their many advice, discussion, support and kindness in my daily life.

Finally, I would like to express grateful appreciation to my family for many support and encouragement during my campus life in Hokkaido University.