Osaka University Knowledge Archive

Title	Solution NMR using paramagnetic relaxation enhancement for the structure determination of loop-rich protein prepared with cold-shock vector containing GST tag
Author(s)	片岡, 沙織
Citation	大阪大学, 2014, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/50468
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

Solution NMR using paramagnetic relaxation enhancement for the structure determination of loop-rich protein prepared with cold-shock vector containing GST tag

常磁性緩和効果を用いた溶液 NMR による、GST タグ付きコールドショック・ベクターで調製したループリッチ蛋白質の立体構造決定法

大阪大学大学院　理学研究科化学専攻　博士後期

蛋白質研究所　機能構造計測学研究室

片岡　沙織

Table of contents

1. General introduction 3

Biological context of sample 10
G protein-coupled receptor belonging to class B 10
Sp Sin1$_{CRIM}$ 16

2. Development of sample preparation methods for protein NMR study 19

2-1. Refinement of pCold-GST expression system 19
2-1-1. Introduction 19
2-1-2. Materials and methods 25
2-1-3. Results 29
2-1-4. Discussion 38

2-2. Sample preparation for PRE measurements 40
2-2-1. ntroduction 40
2-2-2. Materials and methods 43
2-2-3. Results 45
2-2-4. Discussion 46

3. Development of structural analysis method using PRE information for loop rich protein 48

3-1. Iintroduction 48

3-2. Materials and Methods 52
Purification for solution NMR anyalysis 52
NMR measurements and data analyses 53
Structure calculation 54

3-3. Results 59
^{1}H, ^{15}N and ^{13}C resonanaceresonance assignment of sp Sin1$_{CRIM}$ 59
Structure calculation with PRE-derived distance restraints 67

Evaluation of effective minimum number of PRE restraints 70

3-4. Discussion 82

4. General conclusion 83

References 86

1. General introduction

Protein is playing important roles, when an organism performs a life activity. The protein which consists of combination of 20 kinds of amino acid has an original structure and a function, they are working within our cell. And now, many of the protein is the target of drugs. In order to draw a series of flows for a structure determination, functional structure elucidation and drug development, structure determination is important first.

The various approach to elucidate the structure determination of the protein has been established already. The first step in the general protocol of study, is construction of protein expression systems. And next, it is the structural analysis of proteins using variety measurements.

First, in the expression system of the protein, the recombinant DNA is generally used. Protein is produced by the host, for example *E.coli*, yeast, an insect, an animal, and a non-cell. The production method of such protein is chosen according to the purpose, cost, efficiency, and the means of measurement for structural analysis. Among them, the expression system

using E. coli has been widely used for the following reasons: low cost, the rate of expression is faster, and large scale culture is possible. However, producing in large quantities of proteins that is a very small amount in naturally is not easy.

Next, in the analysis of protein, X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy is used. X-ray crystallography is a method of measuring static electron density. The advantage is that there is no limit on the molecular weight of measurable protein and has a high resolution. On the other hand, the concentration of protein required for the measurement was 10 mg/ml, and therefore, the construction of a large-scale expression system and protein large-scale culture of the protein is time-consuming and costly. In addition, disadvantage is the need to crystallize the sample, the investigation of optimum conditions of crystallization is very difficult. NMR is a method of measuring for analyzing the dynamic bond. The advantage is that measurement is possible regardless of the sample conditions such as solution or solid. Then, it is possible to measure the interaction states with the ligand and the dynamic information under various conditions. Especially, the greatest advantage of solution

NMR is that the measurement is possible under conditions close to in vivo. The protein concentration for measurement is about 0.1mM (1 mg/ml at 10 kD, 4 mg/ml at 40 kDa). However, there is a limit to the molecular weight that can be measured, and it is suitable only for the protein of about 40 kDa or less in solution NMR. The electron microscopy is a method of directly measuring the molecules using an electron beam. The advantage is that it is effective for membrane proteins and necessary amount of sample is little. On the other hand, the disadvantage is noise and low resolution.

However, it is difficult to determine the structure although the technique has been established, for the following reasons: diversity of amino acid sequences, isoelectric point (PI), the existence place of in vivo, function, etc. Therefore, it is necessary to consider the expression systems and analysis conditions for individual proteins.

In this research, we established the structure determining method for loop-rich protein using solution NMR. In order to utilization in drug development, we used a solution NMR in this study for determine the structure under conditions similar to in vivo.

Loop-rich protein consists of long loop or has many loop region. The structure of protein is composed of combination of α-helix and β-strand. The α-helix and β-strand are linked by an amino acid sequence called loop, which have irregular shapes and various lengths. Advantage of the loop region is that the loop plays an important role in the formation of the protein structure, for example, the length or conformation of loop participate in folding and the determination of the running direction of the main chain. In addition, since the surface of the loop has usually projected toward the solvent (solution), loop region has become a convenient site for protein recognition, ligand binding and membrane interactions. The loop plays a prominent role in the functions of the protein. However, the structural analysis of soluble loop-rich protein is very difficult from the following problems: (1) The production of soluble protein, (2) The assignment rate of NOE peaks is low when using a solution NMR.

The first matter is the production of soluble protein. The loop region exposing to the solvent is rich in polar hydrophilic residues and charged residues, and loop region frequently is composed of similar amino acids. Therefore expression of the protein using recombinant protein has been

difficult, and even if protein expressed, it is insoluble protein. In the case of insoluble protein, it is necessary to refolding after the denatured with guanidine hydrochloride or urea. However, since protein is made to denature once, the probability of obtaining a protein having a native structure and function again by refolding becomes low. Therefore, it is important to obtain soluble protein without the refolding. In this research, we resolved the problem of the expression system by using a cold shock vector with a GST tag (pCold-GST vector) in order to allow the production of soluble protein.

The pCold-GST vector used in this study has the advantage of promoting the production of protein expression and protein solubility (Hayashi K and Kojima C. 2008). It is described in more detail in the first section for more information, pCold-GST vector is composed of the Cold which is cold shock vector expresses the protein at low temperature and the GST tag which is soluble tag (fig. 4). By inducing the expression of protein at low temperature, the translation of protein is promoted in the pCold vector. As a result, the expression of the protein increase. On the other hand, GST are generally known as a tag to facilitate solubilization. GST is a soluble protein with a molecular weight of 28 kDa. The pCold-GST vector is a vector for

co-expression of the target protein and GST. By co-expression, result of GST having relatively large molecular weight solubilizes, protein is induced for solubilized. In this study, we allowed the expression of soluble loop-rich protein and the large-scale production using the pCold-GST vector having these characteristics of above-mentioned. In the structure determination of the protein, it is very important that a soluble protein in a state having the conventional bond such a disulfide bond are obtained.

The second problem is that the assignment rate of nuclear overhauser effect (NOE) peaks is low. NOE containing the distance information for the nuclear spin peaks influences the cross-relaxation time of spins (Muhandiram D. R *et al.* 1993, Zhang O et al. 1994). In the chemical shifts of the two nuclei, NOE results in cross peaks when there is an interaction between the peaks. Cross-peak intensity by NOE is inversely proportional to the sixth power of the internuclear distance. NOE become large if the distance between nuclear spins is near, and NOE become small if the distance between nuclear spins is far. Moreover, the distance between the nuclear spins which can be observed with the solution NMR is 5 Å, and it cannot observe a far distance beyond it. The Loop portion in a solvent is flexible and takes several conformation. In

the case of loop-rich protein, it may not have a single structure. As a result, it becomes difficult that we obtain internuclear distance information by NOE assignment using the NMR analysis. In this study, we were not possible to obtain NOE spectra of high quality for the reason that the assignment rate of the side chains are low. Therefore, we established the method for the determine of structure using the distance information derived from paramagnetic relaxation enhancement (PRE) in addition to NOE. PRE are the distance information caused by a nuclear spin and an unpaired electron which the metal ion or paramagnetism compound have. Protein has no unpaired electrons. Therefore, in this study, we added MTSL to a protein by a disulfide bond. MTSL is a paramagnetic compound containing an unpaired electron. Peak intensity by PRE is inversely proportional to the sixth power of the internuclear distance. The PRE-derived distance which can be observed with the solution NMR is 15 Å~24 Å. PRE is used as a complement to NOE sometimes. However, so far, the influence of the PRE-derived distance restraints have not been researched for the structure determination of the soluble proteins. By we established these method, we hopes that contributed significantly to the determination of the structure of soluble loop-rich protein.

In this research, we used the samples that were class B GPCRs and Sin1, which contains a lot of long loop domains.

Biological context of sample

G protein-coupled receptor belonging to class B

G protein-coupled receptors (GPCR) containing seven trans-membrane helices play important roles in the biological function, such as signal transduction and homeostasis. Therefore, structure elucidation of GPCRs has been attempted to obtain important clues of new drug development. Now, as a result of the analysis of the genomic sequence, presence of about 800 kinds human GPCRs has been elucidated (Lundstrom K. 2005, Gloroam DE *et al.* 2007) and GPCR is classified into six types (class A ~ class F) based on amino acid sequence information.

Major receptor belonging to class A is rhodopsin. Adrenergic β2 receptor. The crystal structure of the β2-adrenergic receptor has been reported in 2007. Class B has a hormone binding domain at the N terminus, and major receptor is secretin receptor. Class C has a ligand binding domain at N terminus and a long C-terminal loop. Major receptor is metabotropic glutamate receptor. Receptors belonging to the class D only exist in yeast. Major receptor major receptor is STE2/STE3, and ligand is pheromone of

fungi. Receptor of cAMP belongs to the class E and Smoothened (serpentine receptor) and frizzled receptor belong to the class F. These classifications are based on GPCRDB (http://www.GPCR.org/7tm/) (Horn F *et al.* 2003).

Structural analysis of GPCRs is very difficult, because the sample preparation of recombinant protein of GPCRs using *E.Coli* is difficult. GPCRs undergo posttranslational modification such as glycosylation, samples tend to be uneven by using detergent, and structure of GPCRs changes between activation and inactivation. Therefore, structure analysis of GPCRs is seldom progressing in spite of play the important function.

A total of 15 Human GPCRs belong to class B: CALCR, CALRL, CRFR1, CRFR2, GHRHR, GIPR, GLP1R, GLP2R, GLR, PAC1, PTH1R, PTH2R, SCTR, VPAC1 and VPAC2. Class B plays an important role in inhibition or activation of intracellular signaling, immune system, metabolic pathway, and nervous system in vivo. Class B GPCRs have an N-terminal extracellular domain containing about 100-150 amino acids. This domain has a major ligand-binding site; a peptide hormone specifically binds to the N-terminal extracellular domain. Therefore, structural analysis of this domain is required in order to understand the mechanism of interaction with the ligand (Fig.1).

The N-terminal extracellular domain of class B contains conserved six cysteine residues and has three disulfide bonds (Grace *et al.* 2007, Parthier C *et al.* 2007, Tenno *et al.* 2008). According to a recent study, it was revealed

that α-helix of the N-terminus and two pairs antiparallel β-sheets in the region of the core are conserved. Further, tryptophan residues, cysteine residue and disulfide bond are conserved in each receptors, and these has been present in approximately the same position in conformation. Therefore, basic fold of this domain has been maintained by conserved some amino acid residues, most of the N-terminal extracellular domain of GPCR belonging to the class B is expected to have a similar structure (Grace *et al.* 2007, Sun *et al.* 2007, Parthier C *et al.* 2007, Runge *et al.* 2008, Pioszak A.A *et al.* 2008). In addition, this domain has many loop regions. The loop does not have a unique structure. The amino acid sequence of the loop region of each receptor has not been conserved in particular, the length of the loop is also different. However the loop has a high motility, which interacts with ligand. The major ligand interaction site of the N-terminal extracellular domain is constituted of the second loop, the linker portion ranging from first β1 strand to β2 strand (Fig.2). These portions perform hydrophobic interactions with the ligand.

Some structures of this domain have been reported. Recognition of the N-terminal domain site is different in each receptor. The N-terminal domain interacts in the different topologies, and conformation of the loop region is also different (Grace *et al*, 2007; Sun *et al*, 2007; Parthier C *et al*, 2007; Runge *et al*, 2008; Pioszak A.A *et al*,2008). Therefore, structure analysis of all domain is necessary.

However, according to previous studies (Grace *et al*, 2007; Sun *et al*, 2007; Parthier C *et al*, 2007; Pioszak A.A *et al*,2008), the expression rate of this domain is low and insoluble protein, we need the refolding process in the step of sample preparation. In general, the efficiency and success rate of the refolding process is expected to be low for the extracellular domains. Here, we attempted to obtain the soluble protein using pCold-GST.

Furthermore, a long loop like contained in this domain tends to cause protein proteolysis in order to receive the attack of proteolytic enzyme. If protein purification takes time, we can obtain hardly to the monomeric protein suitable for structure determination because loop-rich protein often tend to be formed a dimer. From the above thing, sample preparation of a loop-rich protein is very difficult. Therefore, we tried also solve these problems by using the pCold-GST system.

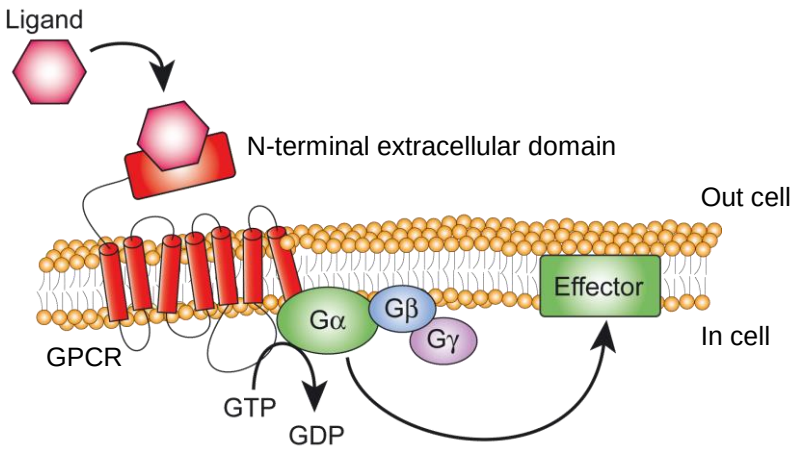


Fig.1. Schematic diagram of intracellular signaling through GPCR belonging class B

GPCR is protein containing seven trans-membrane helices. In the GPCR belonging to class B, the signaling begins by the binding of a ligand to the extracellular domain of the GPCR. GPCR has an active state and an inactive state. When the other molecules do not bond, GPCR is an inactive. In the extracellular, ligand binds with the N-terminal extracellular domain of the GPCR. In addition, in the intracellular, heterotrimeric of G protein (G_α, G_β, G_γ; GDP-bound) binds with GPCR. As the result, GPCR becomes an active state. Active-GPCR induces a conformational change of the G_α subunit and GDP dissociate. After, GPCR binds GTP again. GTP-bound G_α dissociates from the GPCR, and dissociate from G_β and G_γ also. The G_α subunit binds to the effector alone resulting the signal is transmitted.

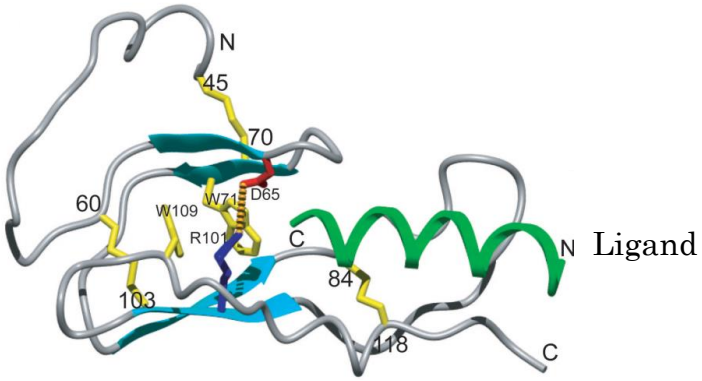


Fig.2. Basic structure of class B GPCR and Complex structure of the peptide hormone (complex structure of CRF2β and astressin). (Grace *et al*, 2007)

Green α-helix is ligand. Ligand-recognition site is different for each receptor. The GPCR belonging to class B of N-terminal extracellular domain is constituted of the second loop, the linker portion ranging from first β1 strand to β2 strand.

Sp Sin1$_{CRIM}$

TOR is a serine/threonine-specific protein kinase and the subcellular target of immunosuppressive drug rapamycin. TOR constitutes TOR complex 1 (TORC1) and TOR complex 2 (TORC2), TORC1 and TORC2 responds differently to the drug rapamycin and have different cellular functions. Case of mammalian, TORC1 is composed of mTOR, Raptor and mLST8, controls cell growth. Whereas TORC2 is composed of mTOR, Rictor, Sin1 and mLST8, regulates cell proliferation by functioning as the regulatory kinase of oncogene protein Akt and other members of the AGC kinase family (Fig.3). Therefore, TORC2 has attracted attention as a new target for anticancer drug development. However, TORC1 has been characterized in great detail, but the regulation of mTORC2 has not been characterized yet. In this study, we focused on TORC2-specific subunit Sin1, which plays an important role for forming the TOR2.

Sin1 stands for SAPK-interaction protein 1. Sin1 is required for the assembly of mTOR2 and the substrate binding of mTORC2 (Yang *et al*, 2006; Jacinto *et al*). Gene is conserved in a wide range between species. We studied for *Schizosaccharomyces pombe* Sin1 (*sp* Sin1). In the model system experiment, a system using the yeast has been preceded, and a model system of TORC2 has been established in fission yeast *Schizosaccharomyces pombe* to elucidate the function (Ikeda et al, 2008). Sin1 homologs contain three

common domains; the conserved region in the middle (CRIM) domain, the Ras-binding domain and the pleckstrin homology (PH) domain (Schroder *et al.* 2004, Schroder *et al.* 2007). Recently, the CRIM domain of *sp* Sin1 is shown to be sufficient to bind a substrate of TORC2 (Tatebe *et al.* in preparation). The CRIM domain of mammalian Sin1 is also shown to be required for substrate binding of mTORC2 (Cameron *et al.* 2011). Although the crystal structures of the PH domains of *S. cerevisiae* and *human* Sin1 were recently determined (Pan and Matsuura. 2012), structural information on the CRIM domain of Sin1 has not yet been reported. Structure of spSin1 is unknown. Therefore, first, we attempted construction of an expression system using pCold-GST system.

Here, we report the result of expression system using pCold-GST for sp Sin1 (amino acid 247-400) which contains the CRIM domain. The fragment of *sp* Sin1 is referred to as $Sin1_{CRIM}$ herein.

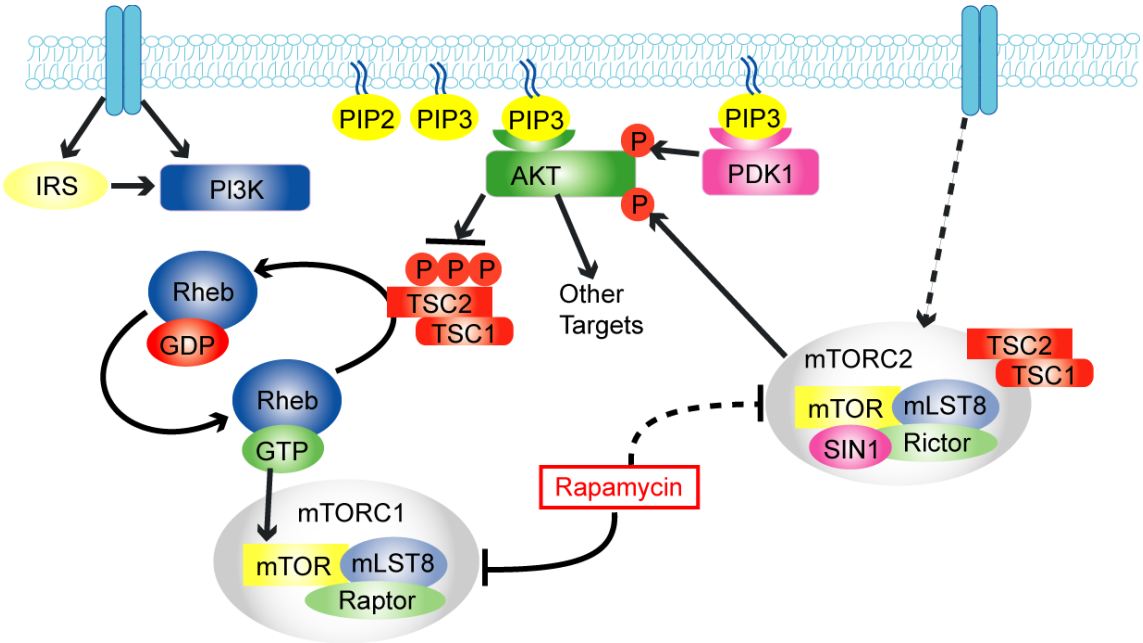


Fig. 3. Schematic diagram of TOR complex.

Sin1 constitute the complex with mTOR, mLST8 and Rictor. mTORC2 phosphorylates Akt.

2. Development of sample preparation methods for protein NMR study

2-1. Refinement of pCold-GST expression system

2-1-1. introduction

As general strategies in an effort to get sufficient amount of recombinant proteins for structural biological study, at present, there are several in vivo and in vitro protein over-production methods (Brondyk WH. 2009). The "Cell-free" protein synthesis, target polypeptides are translated in a test tube by mixing cDNA or mRNA coding target protein and cell extracts, is established as in vitro protein expression system (Murray CJ *et al.* 2013). In the case of in vivo protein expression system, on the other hand, cDNA coding target protein is genetically introduced into living prokaryotic or eukaryotic cells, and the transformed cells over-produce desired protein using their own ribosomal machinery. In many cases, in vivo protein expression systems are superior to in vitro systems in terms of powerfulness of recombinant protein biosynthesis and stability/reproducibility of the bioactivities. Furthermore, in vivo protein expression systems utilizing

prokaryotic cells are major strategies due to its easy handling character. As prokaryotic cells, gram-negative bacterial cells such as *Escherichia coli* (*E. coli*), gram-positive bacterial cells such as *Brevibacillus* and *Corynebacterium* are used. Especially, *E. coli* expression systems are the most widely used systems for the following reasons: easy genetic manipulation and cultivation procedures, rapid cell growth and large amount of biomass, toughness and adaptability to a wide range of culture conditions, and cost effectiveness for isotope labeling. However, in many cases, some eukaryotic heterologous proteins, which demand to form complex disulfide bonds network, membrane proteins such as G-protein coupled receptor (GPCR), play a significant role in biological function, are difficult to over-express with retaining their own native structure and biological activities using *E. coli* and other prokaryotic cell expression systems. In an effort to promote structural biological analyses of biologically significant molecules, therefore, it is valuable to develop and/or refine experimental techniques to overcome the bottlenecks in protein expression systems.

In an effort to facilitate expression level and/or solubility of target proteins using *E. coli* expression system, fusion of solubility enhancement tags (SET)

such as maltose-binding protein (MBP), thioredoxin (Trx), GB1, glutathione-S transferase (GST) to the target protein is one of the standing strategies.

As another strategy, decreasing cultivation temperature during induction of protein over-expression has been proposed to assist proper folding of biosynthesized target proteins and to prevent forming insoluble inclusion bodies in *E. coli*. Inouye et al. developed pCold expression vectors. When *E. coli* cultivation temperature is changed from 37°C to lower (< 15°C), expression of cold-shock proteins are specifically enhanced by activation of cold-shock promoter, cspA promoter, despite biosynthesis of other most of background proteins become significantly slow. The pCold vectors utilize the cspA promoter to selectively over-express target protein in lower temperature. As those features, purity of target protein expressed in *E. coli* is higher compared to other over-expression systems such as utilizing T7 promoter, and it facilitates protein purification. Furthermore, since the cspA promoter exists natively in almost all of *E. coli*, the pCold vector can be applied to all the *E. coli* strains. In several cases, however, target protein expression level cannot be improved due to severe slowdown of cell growth rate and/or gene translation speed although the low-temperature cultivation and pCold vector system has many advantages as described above.

In an effort to overcome the bottleneck of low-temperature cultivation, here, we have established pCold-GST expression vector by inserting cDNA coding GST into the upstream of multi-cloning site of pCold, and N-terminal GST-fused target protein is generated. So far, ten proteins which could not be obtained in soluble fraction using normal pCold or common T7 expression systems were applied to the pCold-GST, nine proteins of the ten could be drastically expressed in soluble fraction. However, especially in the case of membrane protein such as G-protein coupled receptor (GPCR) and loop-rich proteins which have many unstructured regions, protein expression did not succeed even if the pCold-GST vector system was applied to. Therefore, I further developed the pCold-GST vector system in an effort to expand its general versatility and effectiveness. I chosen seven of the N-terminal extracellular domain of Class B GPCRs to validate my study to refine the pCold-GST expression system since they are loop-rich protein and were difficult to express in soluble fraction using normal pCold-GST expression system. In addition, I conducted protein over-expression, purification, and tertiary structure determination by solution NMR techniques of $Sin1_{CRIM}$ prepared by pCold-GST system.

Outline of pCold-GST vector

Basic format of pCold-GST vector was previously designed and developed by Hayashi et al. (Hayashi K and Kojima C. 2008). Main area of the pCold-GST vector consists of 5'UTR, translation enhancing element (TEE) regions, hexa-histidine and following GST sequences, and multi-cloning site (MCS) downstream of the *cspA* promoter (Fig.4). Those components are basically derived from pCold vector. The GST also allows for one-step purification of target proteins by using glutathione-immobilized resin, not only as SET.

Uniquely, human rhinovirus 3C (HRV3C) protease recognition site (amino acid sequence of Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro) has been inserted between the GST sequence and MCS in the pCold-GST vector (Fig.5). The HRV3C protease cleaves C-terminus of the Gln residue in the recognition site (Cordingley MG *et al.* 1990), and the GST can be eliminated from target proteins after purification. In addition, Factor Xa cleavage sequence has been also inserted between the hexa-histidine and the GST sequences (Fig. 5). Users optionally digest and eliminate the N-terminus hexa-histidine tag from GST-fused target protein.

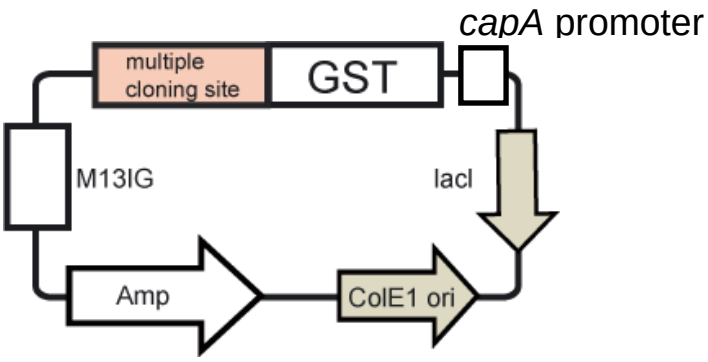


Fig. 4. Vector map of pCold-GST.

Fig.5 HRV3C site in pCold-GST.

At the N-terminus of the GST, in addition, we incorporated hexa-histidine tag. It may be utilized not only for purification of GST-fused target proteins but also for elimination of GST fragments after HRV3c protease digestion of the GST-fused target proteins.

2-1-2. Materials and Methods

Construction of expression plasmids

Eight cDNA, seven N-terminal extracellular domain of class-GPCRs (CALCR, CALRL, GHRHR, GLP2R, GLR, PTH2R, and SCTR) and $Sin1_{CRIM}$ (Fig. 6 and 7), were amplified by PCR and genetically inserted into multiple cloning site of pCold-GST vector.

Estimation of protein expression level

The constructed plasmids were transformed into *E. coli* Rosetta™ (DE3) (Novagen). Initially, the transformants were cultured in 15mL LB media at 37 °C. At the point of the cell suspension reaches OD_{660} of 0.5, the cultivation temperature was promptly downshifted to 15°C. Protein expression was induced by addition of 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) (Wako) in the medium and continuing the culture for overnight. In the case of using *E. coli* ArcticExpress (DE3) (Stratagene) as host cells, recombinant protein over-expression was carried out in the same manner. Following the induction periods, the cells were harvested by centrifugation (2,400 G for 10 min at 4°C). The harvested cells were resuspended in 50 mM Tris-HCl buffer (pH 8.0) containing 300 mM KCl, 0.1 mM EDTA and 1 mM DTT

(buffer A). The cell suspensions were physically disrupted by ultrasonication and crude membranes were eliminated by centrifugation (15,800 G for 20 min at 4°C). Expression level and yield in soluble fraction of target proteins were estimated by performing SDS-PAGE.

Protein purification

The supernatants which were prepared as described in the previous section were loaded onto Glutathione Sepharose 4B resin (GE Healthcare), and the resin was washed extensively with buffer A. Next, the GST-tagged target proteins were eluted from the resin by using buffer A containing 50 mM reduced glutathione. Residues between area of the N-terminal GST and target proteins were digested by using HRV3C protease, and then, the GST fragments were completely eliminated from protein solution by performing gel filtration chromatography using a HiLoad Superdex 75 (26/60) column (GE Healthcare), which was equilibrated with 50 mM potassium phosphate buffer (pH 6.8) containing 50 mM KCl and 1 mM DTT as its running buffer.

DGTITIEEQI VLVLKAKVQC ELNITAQLQE GEGNCFPEWD GLICWPRGTV GKISAVPCPP
YIYDFNHKGV AFRHCNPNGT WDFMHSLNKT WANYSDCLRF LQPDISIGKQ EFFERLYVM

PTH2R (Number of amino acid residues: 119 / MW: 13587.5 Da)

GALPRLSDVL QVLWEEQDQC LQELSREQTG DLGTEQPVPG CEGMWDNISC WPSSVPGRMV
EVECPRFLRM LTSRNGSLFR NCTQDGWSET FPRPNLACGV NVNDSSNEKR HSYLLKLKVM

SCTR (Number of amino acid residues: 120 / MW: 13511 Da)

ASQPQVPSAQ VMDFLFEKWK LYGDQCHHNL SLLPPPTELV CNRTFDKYSC WPDTPANTTA
NISCPWYLPW HHKVQHRFVF KRCGPDGQWV RGPRGQPWRD ASQCQMDGEE IEVQKEVAKM
YSSFQVM

GLR (Number of amino acid residues: 127 / MW: 14737.3 Da)

KQVTGSLLEE TTRKWAQYKQ ACLRDLLKEP SGIFCNGTFD QYVCWPHSSP GNVSVPCPSY
LPWWSEESSG RAYRHCLAQG TWQTIENATD IWQDDSECSE NHSFKQNVDR YALLSTLQLM

GLP2R (Number of amino acid residues: 120 / MW: 13745.3 Da)

MHPESDFITQ LREDESACLQ AAEEMPNTTL GCPATWDGLL CWPTAGSGEW VTLPCPDFFS
HFSSESGAVK RDCTITGWSE PFPPYPVACP VPLELLAEEE SYFSTVKII

GHRHR (Number of amino acid residues: 109 / MW: 11965.4 Da)

AFSNQTYPTI EPKPFLYVVG RKKMMDAQYK CYDRMQQLPA YQGEGPYCNR TWDGWLCWDD
TPAGVLSYQF CPDYFPDFDP SEKVTKYCDE KGVWFKHPEN NRTWSNYTMC NAFTPEKLKN
AYVLYYLAIV G

CALCR (Number of amino acid residues: 131 / MW: 15430.6 Da)

AELEESPEDS IQLGVTRNKI MTAQYECYQK IMQDPIQQAE GVYCNRTWDG WLCWNDVAAG TESMQLCPDY FQDFDPSEKV TKICDQDGNW FRHPASNRTW TNYTQCNVNT HEKVKTALNL FYLTIIG

CALRL (Number of amino acid residues: 127 / MW: 14659.3 Da)

Fig.6. Amino acid sequences of the N-terminal extracellular domain of classB GPCRs

SVSNAKAPTS ALRALLEHKE NSSQNGPLAE NFATFSGHAE SNALRLNIYF PSSESPSKPL FVELRKNVLV SEAIGYILLQ YVNQQLVPPI EDEAQNPNYW NLRIVEDDGE LDEDFPALDR VGPLSKFGFD AFALVKATPA QIKENQAAYP FKSK

$Sin1_{CRIM}$ (Number of amino acid residues: 154 / MW: 16948 Da)

Fig.7. Amino acid sequence of $Sin1_{CRIM}$

2-1-3. Results

Development of pCold-GST expression system

We examined protein expression level in soluble fraction about seven N-terminal extracellular domain of classB GPCRs and $Sin1_{CRIM}$ by applying those cDNA to the pCold-GST system. As *E. Coli* host cell, we used Rosetta™ (DE3) since it is one of the standard strain for over-expression of heterologous proteins. The amount and soluble/insoluble ratio of the expressed target proteins were evaluated by performing SDS-PAGE.

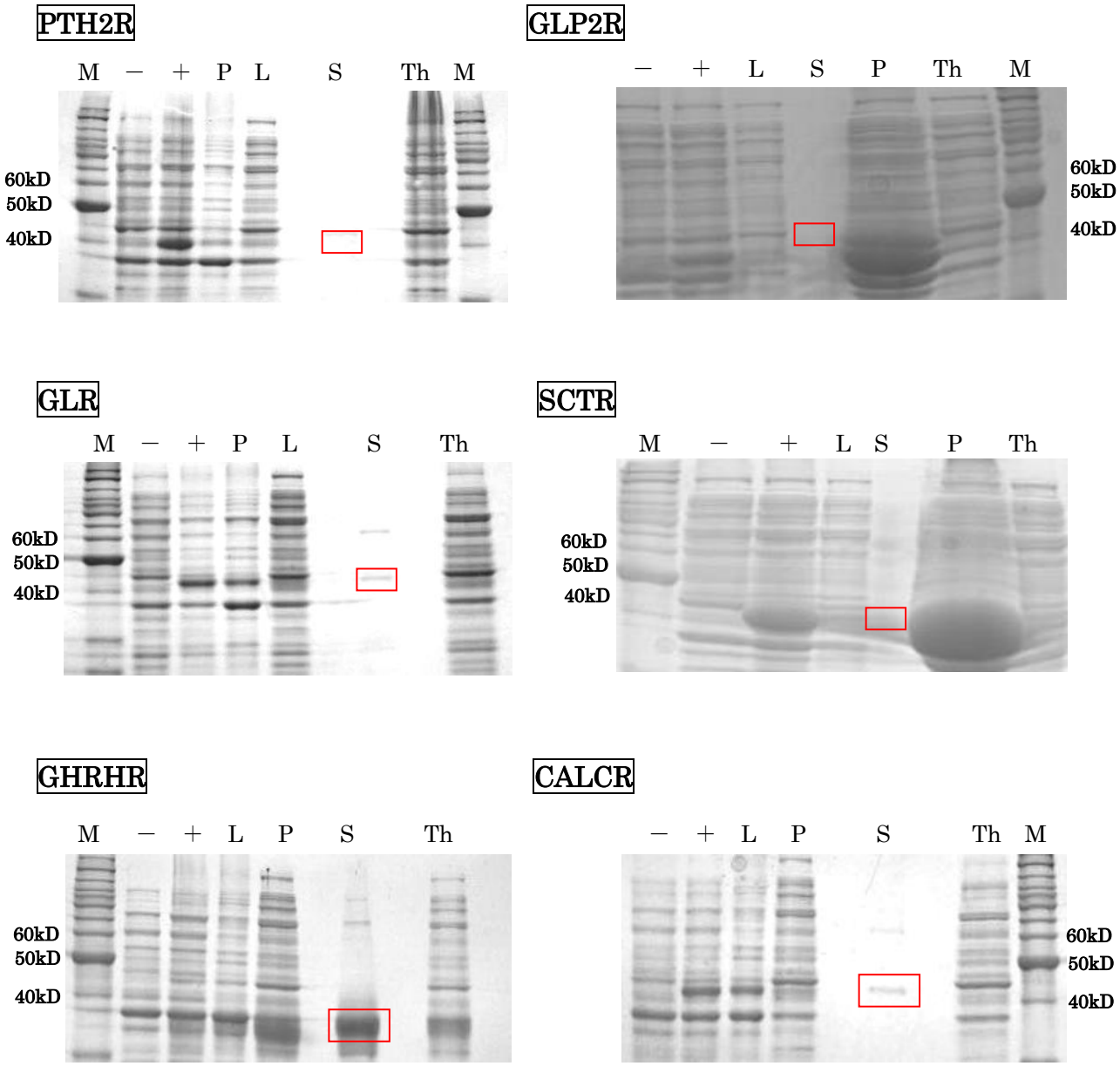
As a result, all of the examined target proteins were successfully expressed. In the case of GPCRs, four (GLR, GHRHR, CALCR, and CALRL) and residual three (PTH2R, GLP2R, and SCTR) were mainly existed in soluble and insoluble fractions of lysate of E. coli host cells, respectively (Fig. 8). In the cases of the $Sin1_{CRIM}$ and GHRHR, sufficient amount of proteins was successfully expressed in soluble fractions (Fig. 10). On the other hand, in the almost cases of GPCRs, total protein expression level and yield in soluble fraction were considerably low even if applying pCold-GST system. Therefore, we examined other *E. coli* host strain from Rosetta™ (DE3) to ArcticExpress (DE3) for all of the eight proteins.

As a result, in all of the GPCRs except for GHRHR, which were insufficient expression level in the case of using Rosetta strain, total protein expression level and its yield in soluble fraction were drastically improved (Fig. 9, Table

1).

Protein purifications of GHRHR and $Sin1_{CRIM}$

As next step, in an attempt to elucidate solution structure of the N-terminal extracellular domain of GHRHR and $Sin1_{CRIM}$, we established large scale protein preparation protocols based on the protein expression level, which was revealed in the previous paragraph. From 1.5 L culture using nutritionally minimal media, we obtained approximately 3.0 g and 3.2 g of cells in wet weight which over-expresses GHRHR and $Sin1_{CRIM}$, respectively. As a result of protein purification work as described in Materials and Methods section, in the case of GHRHR, there were many residual contaminants apart from the target protein and it could not be eliminated sufficiently even if any chromatographic techniques were applied (Fig. 11). On the other hand, sufficient amount of the $Sin1_{CRIM}$ protein of high purity could be prepared (Fig. 12). Finally, we could prepare a 0.5 mL of 0.5 mM $Sin1_{CRIM}$ sample for NMR measurement. However, when concentration of the $Sin1_{CRIM}$ exceed 0.5 mM, insoluble aggregates were formed as time passes.



(Fig.8. continues to the next page)

CALRL

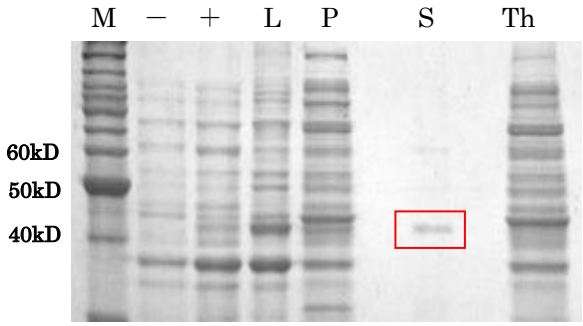
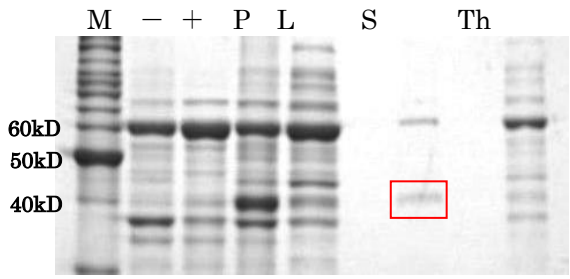


Fig.8 Protein expression of N-terminal extracellular domain of classB GPCR using pCold-GST system
(the result of SDS-PAGE)

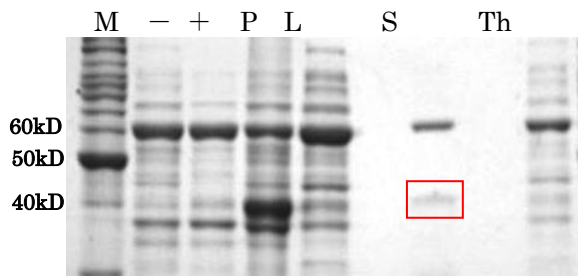
M; Bench Mark protein radder, −and +; before and after of IPTG injections, L; lysate, P and S; insoluble and soluble fractions separated from the lysate, Th; flow-through fraction from glutathione sepharose resin.

The red-lined square indicates band of target protein.

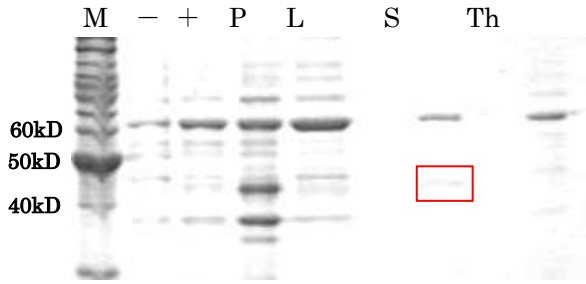
PTH2R



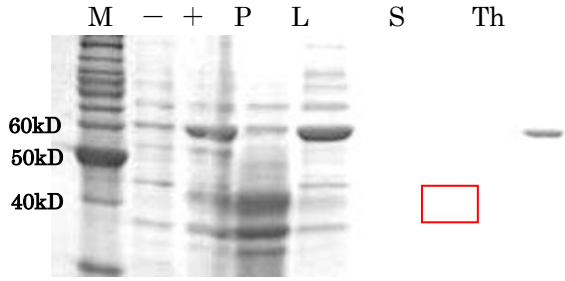
GLP2R



GLR

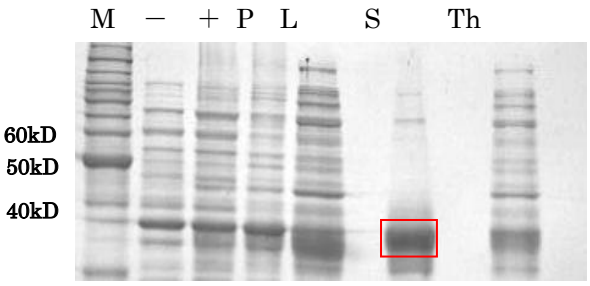


SCTR

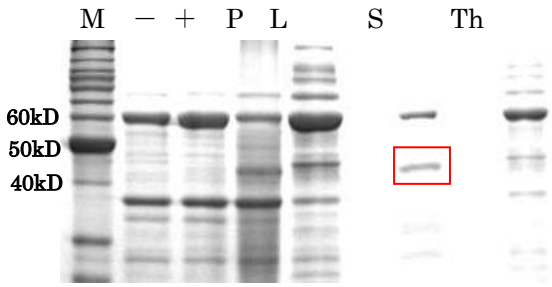


(Fig.9. continues to the next page)

GHRHR



CALRL



CALCR

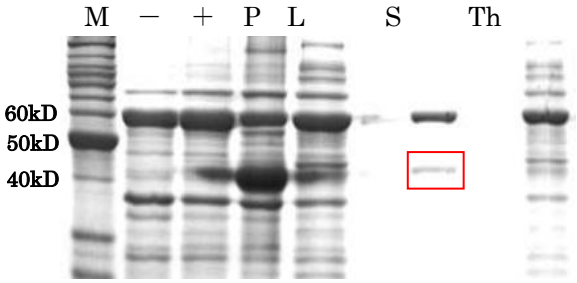


Fig.9 Protein expression of N-terminal extracellular domain of classB GPCR using pCold-GST system. (the result of SDS-PAGE)

M; Bench Mark protein radder, −and +; before and after of IPTG injections, L; lysate, P and S; insoluble and soluble fractions separated from the lysate, Th; flow-through fraction from glutathione sepharose resin.

The red-lined square indicates band of target protein.

Table.1 Result of protein expression of N-terminal extracellular domain of classB GPCR using pCold-GST system.

	Competent cell	
Receptor name	BL21(DE3)Rosetta	ArcticExpress (DE3)
PTH2R	×	○
GLP2R	×	○
GLR	○	○
SCTR	×	×
GHRHR	○	○
CALCR	○	○
CALRL	○	○

×; insoluble protein, ○; soluble protein.

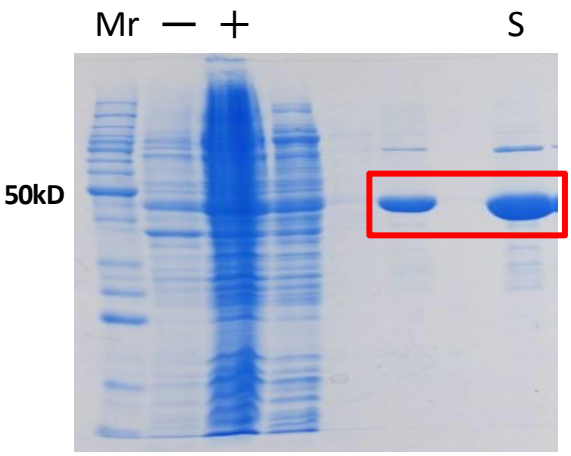


Fig.10 Protein expression of using pCold-GST system Sin1$_{CRIM}$ (the result of SDS-PAGE)

M; Bench Mark protein radder, −and +; before and after of IPTG injections, L; lysate, P and S; insoluble and soluble fractions separated from the lysate, The red-lined square indicates band of target protein.

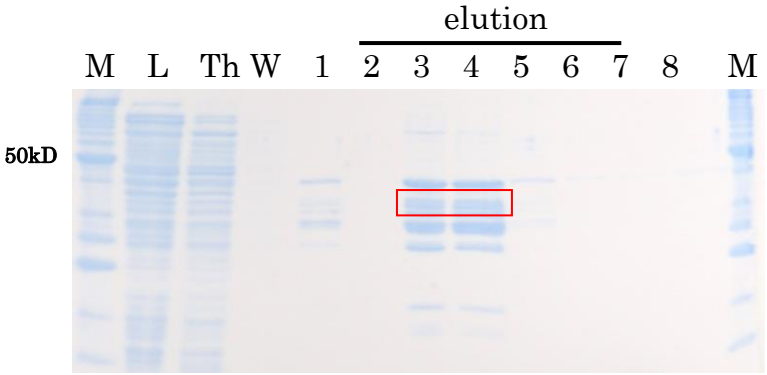


Fig.11. Purification of pCold-GST/GHRHR using affinity column (the result of SDS-PAGE)

M; Bench Mark protein radder, －and +; before and after of IPTG injections, L; lysate, P and S; insoluble and soluble fractions separated from the lysate, Th; flow-through fraction from glutathione sepharose resin. W; wash, 1; before elution, 2-7; eluate, 8; after elution.
The red-lined square indicates band of target protein.

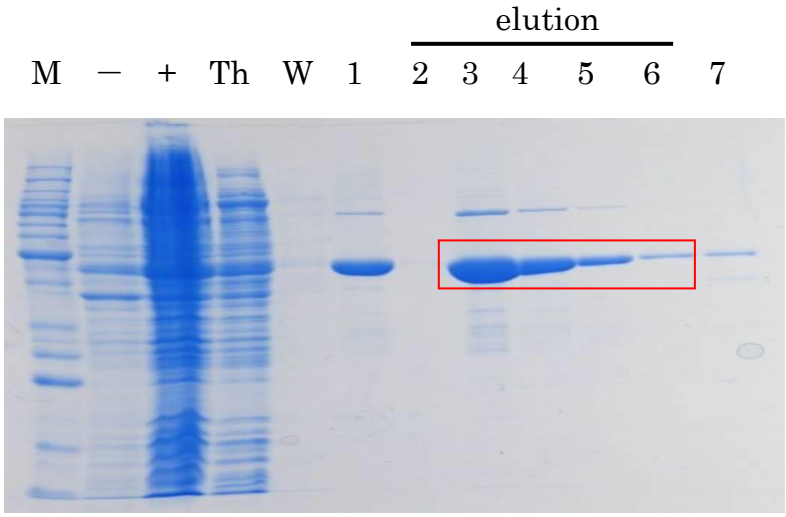


Fig.12. Purification of pCold-GST/spSin1$_{CRIM}$ using affinity column (the result of SDS-PAGE)

M; Bench Mark protein radder, −and +; before and after of IPTG injections, L; lysate, P and S; insoluble and soluble fractions separated from the lysate, Th; flow-through fraction from glutathione sepharose resin. W; wash, 1; before elution, 2-6; eluate, 7; after elution.
The red-lined square indicates band of target protein.

2-1-4. Discussion

Appling pCold-GST system, all of the examined 8 proteins, N-terminal extracellular domain of class B GPCRs and $Sin1_{CRIM}$, could be expressed as soluble form in Rosetta™ (DE3) *E. coli* host cells. Furthermore, by changing the host *E. coli* strain to ArcticExpress (DE3), protein expression level in soluble fraction was drastically improved in all of the cases. From the results, we could further confirm that the pCold-GST system has great potential to produce "high-difficulty" proteins in soluble form, and combining *E. coli* strain ArcticExpress (DE3) and pCold-GST system is one of the effective strategies to produce loop-rich proteins in soluble fraction. Since the GPCRs contain three disulfide bonds, I assumed that correct formation of these disulfide bonding is essential for proper protein folding. The ArcticExpress (DE3) includes two molecular chaperonin (Cpn) 60 and 10, whose chaperonin-activity is optimal in low-temperature (4-12°C) circumstance. It seems that the Cpn60 and Cpn10 facilitated protein folding of GPCRs under the cold-shock condition, and it prevented exposure hydrophobic core of the proteins and it led to improvement of protein solubility.

The protein of $Sin1_{CRIM}$ could be sufficiently expressed using the pCold-GST system in soluble fraction of Rosetta™ (DE3) strain, and the protein could be purified successfully in sufficient quality for NMR measurement.

Therefore, we conducted a structural analysis of $Sin1_{CRIM}$ using solution NMR as described in the next chapter.

2-2. Sample preparation for PRE measurements

2-2-1. Introduction

In general, a loop region, which is connecting secondary structure elements in proteins, does not form apparent secondary structures. However, its flexibility varies depending on individual proteins, one is highly disordered and the others are relatively rigid, It seems to be unique tertiary structure. In many cases, properly flexible or disordered loop regions play a significant role in its biological function such as their own enzymatic activity and molecular-molecular interaction. Therefore, especially for proteins which have rigid and "structured" loops, it is valuable to elucidate its tertiary structure and character of molecular dynamics at atomic resolution. However, in general, structural biological analyses of loop-rich protein are difficult because it is hard to crystalize the proteins and to obtain sufficient resolution of X-ray diffraction. Even if using NMR spectroscopy, line width of NMR signals of the loop regions may be severely broadened when the time scale of molecular fluctuation of the loop region is comparable to the chemical shift offset. In addition, since the regions which form no specific secondary structure are poor in hydrogen bonds and number of NOE signals, accurate and converged tertiary structure calculation of the fluctuated

region by NMR techniques is difficult. Measuring multi-dimensional NMR spectra and analyzing chemical shift of backbone ^{1}H, ^{13}C, and ^{15}N of $Sin1_{CRIM}$, we revealed that the approximately 50% region of the whole $Sin1_{CRIM}$ consists of loop region. It did not form specific secondary structure. Furthermore, it was also suggested by the chemical shift analyses that the almost overall of loop of the $Sin1_{CRIM}$ were not fully disordered but relatively rigid. These results indicate that the loop region of the $Sin1_{CRIM}$ forms distinct higher-order structure. In addition, especially at the loop regions, large moieties of NMR signals of side-chains were severely broadened and/or disappeared, and the number of NOE signals which required as distance constraints for structural calculation were few.

In this study, we tried to establish new methodology for structure determination of the structured-loop-rich proteins by using solution NMR in the process of tertiary structure determination of $Sin1_{CRIM}$.

We focused on paramagnetic relaxation enhancement (PRE) effect for structure determination of $Sin1_{CRIM}$ since PRE provide long-range semi-quantitative distance information in the range of 15 to 24 Å, and it may complement shortage of NOE-derived distance constraints.

PRE is caused by the spin labeled chemical compounds which containing a

stable lone-pair electron. Therefore, the chemical compounds which have a lone-pair electron must be attached to the target protein. In this study, we used (1-oxy1-2,2,5,5-tetramethyl-Δ3-pyrrriline-3-methyl) methanethiosulfonate (MTSL) (Fig. 13) as spin labeled chemical compound. Since the MTSL has a free thiol group, it can be chemically conjugated to the thiol moiety of cysteine residues on the surface of the target proteins by spontaneous disulfide bond formation. In order to determine high resolution tertiary structure of proteins, it is required for gather variation of PRE-derived distance constraints as many as possible. Therefore, amino acid residues which exposed to the surface of the target protein should be substitute to cysteine, and MTSL-labeled various type of target protein samples must be prepared as many as possible. In this section, I will describe sample preparation of various $Sin1_{CRIM}$ mutants and procedures of MTSL labeling. Details of the structure determination of $Sin1_{CRIM}$ by applying the PRE constraints are described in Chapter 3.

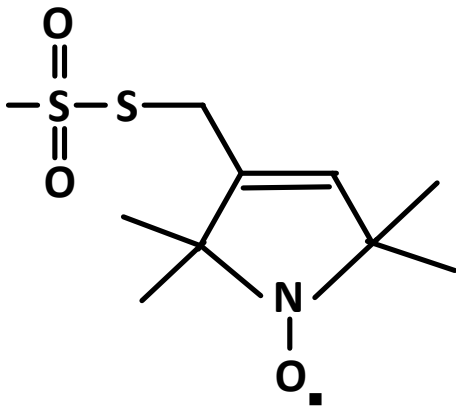


Fig. 13. The structural formula of MTSL ($C_{10}H_{18}NO_3S_2$).

MTSL has an unpaired electron, it is used as a paramagnetic spin label. MTSL has cysteine residue. MTSL is introduced into the protein specifically by cysteine residue of MTSL bounds to cysteine residue of object with disulfide bond. As a result, protein is spin labeled.

2-2-2. Materials and Methods

Cysteine mutagenesis and spin-labeling of $Sin1_{CRIM}$

The Table 1 is a list of single cysteine mutants which were designed for site directed spin labeling of $Sin1_{CRIM}$. Single cysteine mutations were introduced into the pCold-GST/$Sin1_{CRIM}$ plasmid by QuikChange site-directed

mutagenesis method (Stratagene). Uniformly ^{15}N-labeled $Sin1_{CRIM}$ mutant proteins were overexpressed and purified by the same procedures as described in the previous section, except for the composition of running buffer of size-exclusion column chromatography, which did not contain reducing agents such as DTT. After the mutated $Sin1_{CRIM}$ proteins were purified, MTSL was covalently attached to the thiol moiety of newly introduced cysteine residues as follows. The MTSL was added to the mutants of $Sin1_{CRIM}$ samples (0.1 mM) at a 10:1 (MTSL:proetein) molar ratio, and it was incubated at 20 °C for 4h. Unreacted MTSL was removed by performing size-exclusion column chromatography and sufficiently purified MTSL-labeled $Sin1_{CRIM}$ proteins were applied to NMR measurements. Following the collection of paramagnetic NMR data, diamagnetic NMR data were collected by using the same NMR samples, whose unpair electron moiety of the MTSL attached to the $Sin1_{CRIM}$ was completely reduced by adding 3-fold molar excess of ascorbic acid and incubating at 25°C for more than 1 h. Conjugation of MTSL on the $Sin1_{CRIM}$ proteins was confirmed by performing mass spectroscopy after the collection of both paramagnetic and diamagnetic NMR data.

2-2-3. Results

In order to obtain PRE-derived distance restraints, a total of 35 single cysteine mutants of $Sin1_{CRIM}$ were designed for site-directed spin labeling (Table 2). Among them, plasmids of 20 mutants, S269C, T280C, S282C, R291C, S301C, K312C, S317C, G321C, Q331C, L322C, Q341C, G355C, F361C, R366C, S371C, T384C, A386C, A393C, A394C, and S399C, were successfully constructed. With regard to the other mutants listed above, S248C, S256C, D260C, S287C, S298C, K304C, S319C, V333C, R349C, E359C, D360C, A363C, K382C, Q392C and Y395C, accurate gene modification to single cysteine substitution could not succeed. Within the 20 mutants, the 19 mutants (except for A393C) of proteins could be successfully over-expressed and collected in soluble fraction of *E. coli* lysate. Those 19 mutant proteins could be purified. Within the 19 mutants, however, protein expression level of K312C and Q331C were extremely low. In addition, since Q331C is mutant protein prone to be insoluble aggregate, we just obtained total 0.02 mM of that protein as soluble form.

2-2-4. Discussion

We have designed the insertion site of the spin label in the domain in order to obtain the PRE information without information being concentrated. Finally, we succeeded in the sample preparation of a total of 19 mutants. We chose the parts of loop which do not take the secondary structure in the insertion site.

Table. 2. Mutants of $Sin1_{CRIM}$ designed for the site directed spin labeling.

Mutant	construction[a]	expression	solubility	purification	NMR measurement
S248C	×	-	-	-	-
S256C	×	-	-	-	-
D260C	×	-	-	-	-
S269C	○	○	○	○	○
T280C	○	○	○	○	○
S282C	○	○	○	○	○
S287C	×	-	-	-	-
R291C	○	○	○	○	○
S298C	×	-	-	-	
S301C	○	○	○	○	○
K304C	×	-	-	-	-
K312C	○	Δ	○	○	○
S317C	○	○	○	○	○
S319C	×	-	-	-	-
G321C	○	○	○	○	○
Q331C	○	Δ	Δ	○	○

L332C	○	○	○	○	○
V333C	×	-	-	-	-
Q341C	○	○	○	○	○
R349C	×	-	-	-	-
G355C	○	○	○	○	○
E359C	×	-	-	-	-
D360C	×	-	-	-	-
F361C	○	○	○	○	○
A363C	×	-	-	-	-
R366C	○	○	○	○	○
S371C	○	○	○	○	○
K382C	×	-	-	-	-
T384C	○	○	○	○	○
A386C	○	○	○	○	○
Q392C	×	-	-	-	-
A393C	○	-[b]	-	-	-
A394C	○	○	○	○	○
Y395C	×	-	-	-	-
S399C	○	○	○	○	○

[a]The plasmids containing appropriate mutation was obtained (○) or not (×).
[b]Not tried.

3. Development of structural analysis method using PRE information for loop rich protein

3-1. Introduction

In this chapter, we performed the structural analysis of the *sp* $Sin1_{CRIM}$ whose which sample preparation was successful as described in the previous chapter, and where we performed the structure determination of the *sp* $Sin1_{CRIM}$ using Paramagnetic Rrelaxation Eenhancement (PRE) effect.

Structure determinations of proteins by solution NMR relyies on the assignments of backbone and side chain and the information from distance measurements of nuclear Overhauser effect (NOE), this information is necessary to determinate of structure. In general procedure of structure determination of proteins by NMR, quality of result of the calculation mainly depends on the number and quality of NOE-derived distance restraints. In order to achieve a high-resolution structure determination, sufficient number of NOE peaks and nearly complete and accurate assignments of the NOE peaks are required. For NOE signal assignments, nearly complete assignments of ^{1}H chemical shifts are essential step. Therefore, obtaining a series of high quality NMR spectra is necessary for successful structure determination. This cannot be always achieved depending on the physical

properties of the target protein. Widely used computational programs such as CYANA (Güntert *et al.* 1997; Herrmann *et al.* 2002), ARIA (Nilges *et al.* 1997) and AutoStructure (Huang *et al.* 2000), generate structures of the protein by molecular dynamics calculation in combination with automated NOE signal assignments program. In the case of CYANA, it has been shown that NOE peaks which measured more than 8.4 per a residue and not less 90% of completeness of chemical shift assignments is necessary for successful structure determination by performing automated NOE assignments (Jee and& Güntert. 2003).

PRE has been recognized as an approach for obtaining long-range conformational restraints for following reasons: PRE effect caused by the spin labeled chemical compounds, which contains a stable unpair electron, provides the semi-quantitative distance information in the range of 15 to 24 Å, that is, PRE restraints have the advantage to cover longer distances as compared with NOEs. The distance measurements by NOEs are limited to distances of up to 5 Å, and NOE signals are sometimes becomes undetectably weak. Therefore, the information of PRE can complement the information of NOEs. PRE restraints have the advantage to cover longer distances as compared with NOEs. For the above reasons, PRE has been recognized as an

approach for obtaining long-range conformational restraints.

PRE is caused by the spin labeled chemical compounds which containing a stable lone-pair electron. In this study we tried analysis of the backbone and side-chain, and acquisition of the distance information by NOEs. In addition, we tried acquisition of the distance information for loop-rich protein using PREs in order to supplement the information of NOE; we insert the spin label in the loop part and tried getting PRE information.

Recently, PRE is utilized to determine the inter-domain orientation or dimer interfaces (Madl *et al.* 2010, Yang *et al.* 2010), and tertiary structure of the membrane proteins in detergent micelles (Roosild *et al.* 2005, Zhou *et al.* 2008, Van Horn *et al.* 2009, Reckel *et al.* 2011). Detailed analyses of the influence of PRE-derived distance restraints on the structure determinations of α-helical membrane proteins with limited long-range NOE information have been conducted using simulated NMR data (Gottstein *et al.* 2012). It showed that how the number and the location of spin labels affect qualities of the result of structure calculation. Regardless of the extent of large or membrane proteins, in some cases, soluble proteins suffer from obtaining high quality spectra and the high-resolution structures cannot be determined. To our knowledge, influences of employing PRE-derived

distance restraints to determine high-resolution structures of such proteins have not been investigated.

In this study, we tried analyses of the backbone and side-chain, and acquisition of the distance information by NOEs for a loop-rich protein. In addition, we tried acquisition of the distance information for the loop-rich protein using PREs in order to supplement the information from NOE; we introduced spin labels in the loop parts and tried getting PRE information.

3-2. Materials and Methods

Purification for solution NMR anyalysis

For the preparation of ^{15}N- and ^{13}C,^{15}N-labeled proteins, the transformant cells were grown in M9 minimum medium containing 4 g/L unlabeled glucose and 0.5 g/L $^{15}NH_4Cl$, or 1 g/L ^{13}C-glucose and 0.5 g/L $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. Initially, the cells were cultured at 37°C. At the point of the cell suspension reaches OD_{660} of 0.5, the cultivation temperature was downshifted to 15°C. Protein expression was induced by addition of 1 mM isopropyl-ß-d-thiogalactopyranoside (Wako) in the medium and continuing the culture for overnight. Unlabeled protein was overexpressed by using LB medium in a same manner as described above.

Expressing cells expressing the proteins were harvested by centrifugation (2,400 G for 10 min at 4°C). The harvested cells were resuspended in buffer A. The cell suspensions were physically disrupted by ultrasonication and crude membranes were eliminated by ultracentrifugation (138,000 G for 30 min at 4°C). The supernatants were then loaded onto a Glutathione Sepharose 4B column, and the resin was washed extensively with buffer A. GST-tagged protein was eluted using buffer A containing 50 mM reduced glutathione. The N-terminal GST-tag of the eluted protein was removed by digestion using human rhinovirus 3C protease. After the protease treatment, the sample was further purified by Gel filtration chroimnatography using a

HiLoad Superdex 75 (26/60) column (GE Healthcare), which equilibrated with 50 mM potassium phosphate buffer (pH 6.8) containing 50 mM KCl and 1 mM DTT.

NMR measurements and data analyses

NMR samples were prepared at a protein concentration of 0.5 mM in 90/10% H_2O/D_2O containing 50 mM potassium phosphate (pH 6.8), 50 mM KCl and 1 mM DTT.

NMR experiments were performed using AVANCE I 800 and AVANCE III 950 spectrometers at the ^{1}H resonance of 950 MHz (Bruker). For backbone resonance assignments, Non-Uniformly Sampled (NUS)-HNCO, NUS-HNCA, NUS-HN(CO)CA, NUS-HN(CA)CB and NUS-HN(COCA)CB spectra (Rovnyak, D. *et al.* 2004) were recorded. For side-chain resonance assignments, HBHA(CCO)NH (Grzesiek and, S. & Bax, A. 1993) , C(CO)NH (Montelione, G. T. *et al.* 1992), NUS-H(CCO)NH (Lyons, B. A. & and Montelione, G. T. 1993), HCCH-TOCSY (L. E. Kay *et al.* 1993) and ^{1}H-^{1}H NOESY (S. Macura *et al.* 1981) spectra were recorded. To obtain distance restraints, ^{13}C-edited NOESY (Muhandiram D. R. *et al.* 1993) and ^{15}N-edited NOESY spectra (Zhang, O. *et al.* 1994) were recorded. To obtain χ_1 angles, three bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings (Hu J. S. and Bax A. 1997) were measured. ^{1}H-^{15}N heteroNOE was measured with and without 3 sec of proton saturation in an interleaved fashion (Farrow N. A. *et al.* 1994). PRE

effects were measured from ^{1}H-^{15}N HSQC spectra of MTSL-conjugated proteins recorded in the oxidized (paramagnetic) and reduced (diamagnetic) states. MTSL quenching was achieved by addition of 3-fold molar excess ascorbic acid. N-H residual dipolar couplings (RDCs) were measured using inphase (IP) and antiphase (AP) HSQC sequences under both isotropic and anisotropic conditions (Ottiger M. *et al.* 1998).

Uniformly sampled NMR spectra were processed using NMRPipe (Delaglio F. *et al.* 1995), while NUS NMR spectra were processed using Rowland NMR toolkit (http://rnmrtk.uchc.edu/rnmrtk/RNMRTK.html). NUS NMR spectra is able to increase the information content of the multi-dimensional NMR spectrum of the particularly difficult biopolymer in normal measurement (Sven G. H. 2013). Spectra were analyzed using Magro NMRView and Sparky 3.115 (Goddard, T. D. and Kneller, D. G. SPARKY 3, University of California, San Francisco).

Structure calculation

In order to obtain distance restraints form PRE data, first, the contribution of oxidized spin label to relaxation rates was calculated from intensity ratios of HSQC spectra in the oxidized and reduced states, according equations (1). (Battiste J. L. and Wagner G. 2000)

$$I_{ox}/I_{red} = R_2\exp(-R_2^{sp}t)/(R_2+R_2^{sp}) \qquad .....(1)$$

where I_{ox} and I_{red} are the peak intensities in the oxidized and reduced states, respectively, t is the total INEPT evolution time of the HSQC (10 ms). R_2 and R_2^{sp} are the transverse relaxation rate for amid spin in the reduced states

and the contribution of electron spin in the oxidized states to the relaxation rate. R_2^{sp} was then converted into distances using the following equation (2),

$$r = [K/ R_2^{sp}(4\tau_c + 3\tau_c/ 1+\omega_h^2\tau_c^2)]^{1/6} \quad .....(2)$$

where r is the distance between the electron (unpaired electrons of MTSL) and nuclear spins, τ_c is the correlation time for the electron-nuclear interaction, ω_h is the Larmor frequency of the proton nuclear spin, and K is 1.23×10^{-32} cm^6s^{-2} composed of physical constants (Battiste J. L. and Wagner G. 2000). For calculating distances, the approximation was made that τ_c was equal to the global correlation time of the protein estimated from the molecular weight of the protein, and R_2 was estimated from the line width at half-height ($\Delta\nu_{1/2}$) in proton dimension by using the equation, $R_2 = \pi\Delta\nu_{1/2}$, in reduced spectra. Line width and peak intensities were measured using Sparky.

PRE distance restraints were classified into three; (1) Peaks with an intensity ratio < 0.8 and detectable in the oxidized spectra, (2) severely broadened peaks not detectable in the oxidized spectra, and (3) peaks with an intensity ratio > 0.8. Peaks in the class (1) were restrained as the calculated distance. Peaks in class (2) were restrained with no lower distance limit and upper distance limits of distances estimated from the noise level. Peaks in class (3) were restrained with no upper distance limit and lower distance limits of distances calculated from intensity ratio of 0.8.

φ and ψ backbone torsion angle restraints were derived using the TALOS program (Delaglio F. *et al.* 1995). Twelve χ_1 angle restraints were obtained by measuring three covalent bonds $J_{C'C_\gamma}$ and J_{NC_γ} coupling

constants. RDC values were measured using nmrDraw program (Delaglio F. *et al.* 1995).

Structure calculations were performed by CYANA 3.95 (Güntert *et al.* 1997, Herrmann *et al.* 2002). In the case of the structure calculation without PRE, the input data consisted of ^{1}H, ^{15}N and ^{13}C chemical shifts of $Sin1_{CRIM}$, NOE peak positions in the ^{13}C- and ^{15}N-edited NOESY-HSQC, and φ, ψ and χ_1 dihedral angles. In the case of the structure calculation with PRE, distance restraints derived from PRE measurements for 9 spin-labeled mutants (T280C, S282C, R291C, S301C, K312C, L332C, S371C, T384C and A394C) were included in addition to the input data described above. PRE distance restraints were introduced for distances between C^{β} atoms of residues mutated to cysteine for the paramagnetic labeling and amide protons with an error of 7Å. The NOE peaks in the ^{13}C- and ^{15}N-edited NOESY-HSQC were automatically assigned in seven cycles of the structure calculations, and the NOE assignment tables were utilized to final structure calculation. Fifty structures were calculated and 10 structures with the lowest target function were selected in each cycle and in the final structure calculation. The 10 structures selected in the final structure calculations were selected as representative. The NOE cross-peaks in the ^{13}C-edited NOESY, ^{15}N-edited NOESY were automatically assigned through seven cycles of structure calculations of CYANA 2.1 (Güntert P. *et al.* 1997, Herrmann T. *et al.* 2002). Throughout the calculations, PRE distance and dihedral angle restraints were employed. In CYANA structure calculations, PRE distance restraints were introduced for distances between C^{β} atoms of residues mutated to

cysteine for the paramagnetic labeling and amide protons with an error of 7 Å. Subsequently a total of 795 NOE upper distance limits were obtained. With NOE distance, PRE distance and dihedral angle restraints, 100 structures were calculated by CYANA 2.1. The 10 structures with the lowest target function were further refined using Xplor-NIH 2.31 (Schwieters C. D. *et al.* 2003). The initial structures for Xplor-NIH structure calculations were generated with a single MTSL nitroxide label at each mutated position. In Xplor-NIH structure calculations, PRE distance restraints were introduced for distances between NS1 atoms of MTSL labels and amide protons with an error of 4 Å. Structure calculations by Xplor-NIH were initially performed with NOE distance, PRE distance and dihedral angle restraints. Ten structures were calculated starting from each structure. Then, the 10 lowest energy structures were verified by examining fits of the RDC data to the structures using PALES (Zweckstetter M. and Bax A. 2000). The Q-factors (a goodness-of fit measure for RDCs; low Q indicates better agreement) range from 0.002 to 0.004 (average, 0.003). Therefore, the NOE- and PRE-based structures were considered to be reasonable.

Next, structure refinements by Xplor-NIH were performed with RDC restraints in addition to NOE distance, PRE distance and dihedral angle restraints. Only RDC values of residues that were classified helix or extend region by TALOS program were employed. Again, 10 structures were calculated starting from each structure. The 10 lowest energy structures were not so different from those obtained from the calculations without RDC restraints: the backbone RMSD between the mean structures was 0.375.

However, structural convergence was improved by the addition of RDC restraints during structure calculations. Thus, the 10 lowest energy structures obtained from structure calculations with RDC restraints were selected as representatives. These 10 structures were analyzed using PROCHECK-NMR (Laskowski R. A. *et al*, 1996) and PALES (Zweckstetter M. and Bax A. 2000). The details of restraints used for the structure calculations by Xplor-NIH and structural statistics for the 10 lowest energy structures calculated with RDC restraints were summarized in Table 3.

3-3. Results

^{1}H, ^{15}N and ^{13}C resonance assignment of *sp* $Sin1_{CRIM}$

We have succeeded in preparing a Max the sample with concentrations of up to 0.5 mM sample. Therefore we performed ^{1}H-^{15}N HSQC measurement using this sample. As a result, although the majority of the ^{1}H-^{15}N HSQC peaks of $Sin1_{CRIM}$ were well resolved with high signal intensities. However, it was quite difficult to obtain sufficient quality of 3D NMR spectra of $Sin1_{CRIM}$ in a conventional manner. Many signals in the 3D NMR spectra were weak or missing. Additionally, the NMR sample of $Sin1_{CRIM}$ precipitated within 3 days of the NMR measurements at 30 °C. To overcome this difficulty, the NUS method was employed in measurements of the 3D spectra of $Sin1_{CRIM}$ in order to acquire as many scans as possible within the limited time period.

Following a standard sequential assignment procedure, 94% of the backbone $^{1}H^{N}$ and ^{15}N resonances of the non-proline residues were assigned (Fig. 13). S247, N251, A252, S257, F279, S288 and A320 were not assigned. Additionally, resonances for 92% of $^{13}C'$, 95% of $^{13}C\alpha$, 91% of $^{1}H\alpha$, 72% of side chain ^{1}H and 69% of side chain ^{13}C were assigned. In Fig.1, side chain signals are connected by lines for each asparagine and glutamine residue. Chemical shifts have been deposited in the BioMagResBank (BMRB) under accession number 11546.

Fig. 13 shows the random coil index-predicted order parameters (RCI-S^2) and the secondary structure elements estimated from the $^{1}H^{N}$, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$

$^{13}C'$ and $^{13}C\beta$ chemical shifts using the TALOS+ program (Shen *et al.* 2009; Berjanskii MV and Wishart DS. 2005)(Fig.15). The RCI is combination of the chemical shift data from six different nuclei ($^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C_O$, ^{15}N, ^{1}HN and $^{1}H\alpha$—or any combinations thereof), and in closely correlates with amplitudes of backbone protein motions such as order parameters (S2) and root mean square fluctuations (RMSFs) of structural ensembles. Namely, RCI can provide quantitatively estimate for backbone RMSFs of structural ensembles and order parameters using only chemical shifts (Mark V. Berjanskii and David S. Wishart. 2007). RCI values and S2 are calculated using the according to equations (3).

$$RCI=(\langle A|\Delta\delta C_\alpha|+B|\Delta\delta CO|+C|\Delta\delta C_\beta|+D|\Delta\delta N|+E|\Delta\delta NH|+F|\Delta\delta H\alpha|\rangle)^{-1} \quad .....(3)$$

$|\Delta\delta C_\alpha|$, $|\Delta\delta C_o|$, $|\Delta\delta C_\beta|$, $|\Delta\delta N|$, $|\Delta\delta NH|$, $|\Delta\delta H\alpha|$ are the absolute values of the secondary chemical shifts (in p.p.m.) of C_α, C_O, C_β, N, NH and Hα, respectively. A, B, C, D, E and F are weighting coefficients. Left angle and right angle brackets (< >) indicate that the average is being calculated (Wishart DS. 1994).

The N-terminal 24 consecutive residues except for S249 were predicted to have RCI-S^2 lower than 0.5, while the remaining residues were estimated to have RCI-S^2 higher than 0.5. This indicated that $Sin1_{CRIM}$ comprises an N-terminal flexible region and a C-terminal structured domain. The structured domain of $Sin1_{CRIM}$ possessed a low content of secondary

structure elements. In the structured domain, P273-T280, V316-N329, and P385-A393 were predicted to form α-helices, while R291-Y295, L306-L310, W346-V351 and A377-K382 were predicted to form β-strands.

In addition to the secondary structure estimation by the program TAROS+, and performing {^{1}H-^{15}N} heteronuclear NOE measurements revealed that the more than half of the regions except for secondary structured moieties of Sin1$_{CRIM}$ forms flexible loop (Fig. 16). From these results, I have found that Sin1 is a loop-rich protein. Usually, the structural analysis of loop-rich protein is difficult because loop is flexible.

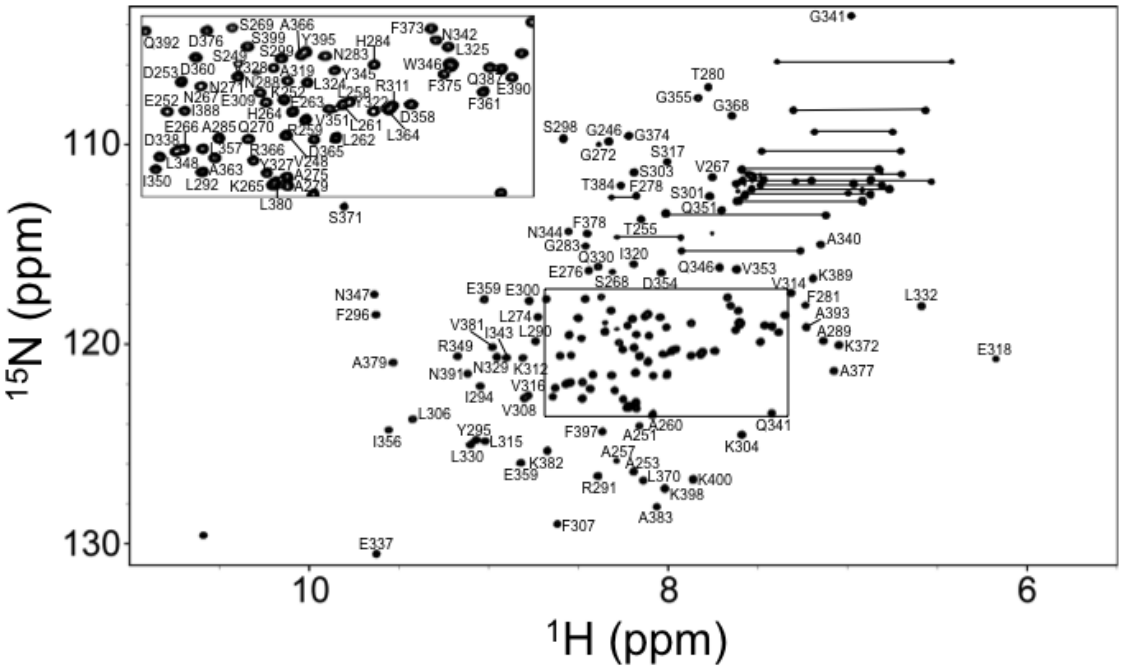


Fig. 14. ^{1}H-^{15}N HSQC spectrum of $^{15}N/^{13}C$-labeled $Sin1_{CRIM}$.
Side chain signals are connected by lines for each asparagine and glutamine residue.

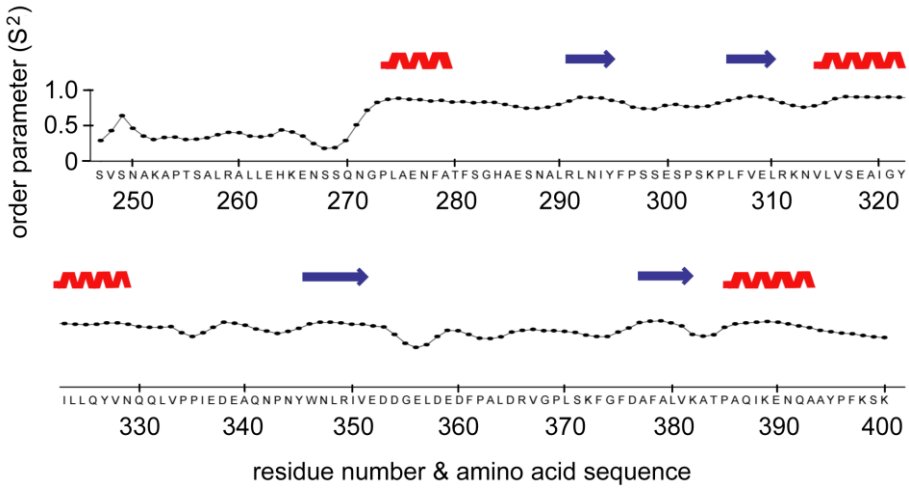


Fig. 15. Random coil index-predicted order parameter and secondary structure of $Sin1_{CRIM}$ estimated using the TALOS+ program based on the $^{1}H^{N}$, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$ $^{13}C'$ and $^{13}C\beta$ chemical shifts.
Beta-strands are blue and α-helices are red.

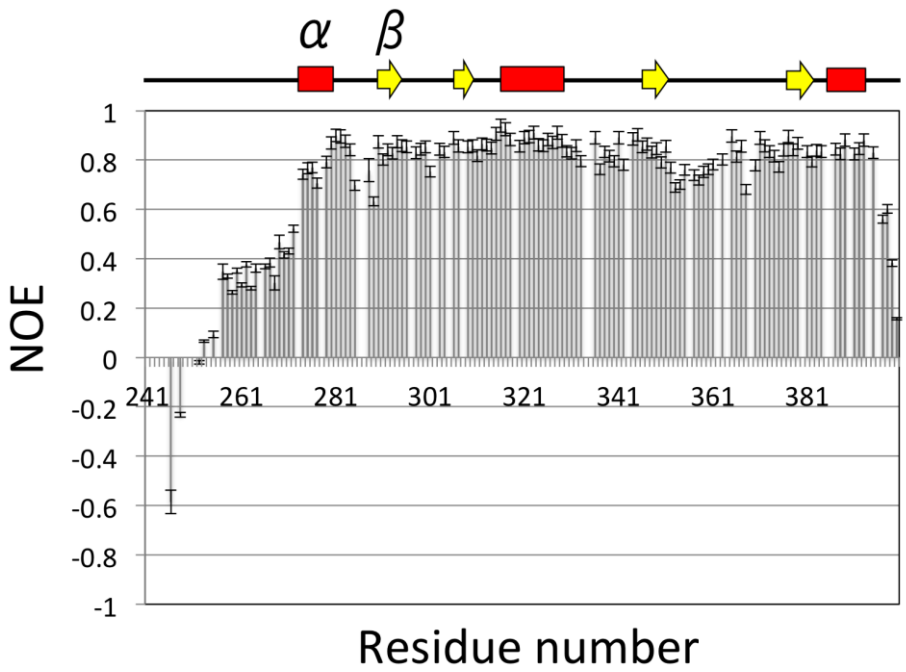


Fig. 16. $\{^{1}H\text{-}^{15}N\}$ heteronuclear NOE values of $Sin1_{CRIM}$.

Structure calculation withouput PRE-derived distance restraints

Structure calculations of $Sin1_{CRIM}$ were performed by the program CYANA with automated NOE assignments in a general structure calculation procedure. However, generated structures were not converged sufficiently (Fig. 17). The backbone root mean square deviation (RMSD) values of the structured region (amino acids 275-395) and the correlation coefficient between experimental RDC values which as input restraints and back-calculated one from the generated tertiary structures were 3.06 ± 0.89 Å and 0.56 ± 0.12, respectively (Table. 3).

The unambiguous assignments of the chemical shift were difficult and the NOESY spectra could not be measured with high signal-to-noise (S/N) ratio.

Therefore, we did not obtain sufficient convergence by normal structural calculation.

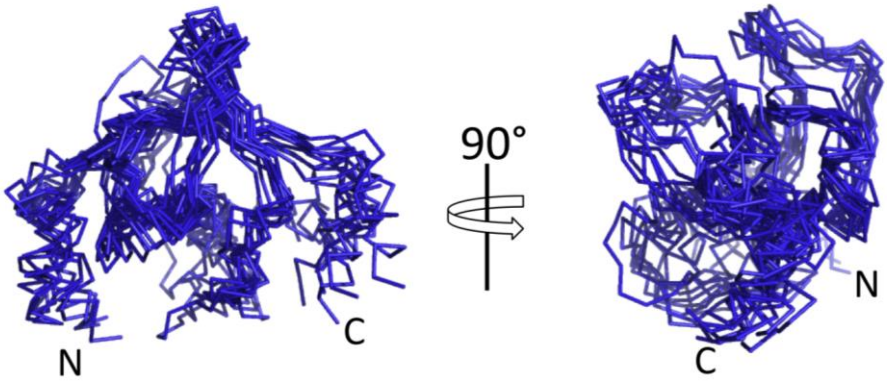


Fig. 17. Superimposition of the final 10 structures calculated by the program CYANA without PRE distance restraints.

Table.3. The RMSD values of backbone atomsStructural statistics for the NMR structures of $Sin1_{CRIM}$ calculated without PRE distance restraints.

NOESY peaks	
^{13}C	3911
^{15}N	1250
Total	5161
Completeness of chemical shift assignments	
	84.80%
Dihedral angle restraints	
φ	110
ϕ	90
χ1	12
NOE upper distance limit	
short $\|i-j\| \leq 1$	645
middle $1 < \|i-j\| \leq 5$	136
long $5 \leq \|i-j\|$	143
Total	924
RMSD (amino acids 275-395) Å	
backbone	3.06 ± 0.89
all heavy	3.57 ± 0.84
RDC correlation	0.56 ± 0.12

Structure calculation with PRE-derived distance restraints

The high-resolution tertiary structure of *sp* $Sin1_{CRIM}$ could not be determined by general NOE-based solution structure determination procedure as shown in the previously (Fig. 17, Table. 3). Wherein, PRE-derived distance restrains were additionally employed on the structure calculation together with performing automated NOE assignments by CYANA. In order to obtain PRE-derived distance restraints, a total of 34 single cysteine mutants were designed for site-directed spin labeling (Table. 1).

Twenty mutants were successfully introduced into the *Sp*Sin1 gene by genetic engineering, and their recombinant proteins overexpressed and purified. First, we combined MTSL and thiol moiety of cysteine residues. Next, we measured each mutant by ^{1}H-^{15}N HSQC for each mutant in the oxidized and reduced solution conditions.

As a result of the ^{1}H-^{15}N HSQC measurements, among prepared mutants, S269C, G321C, Q331C, Q341C, G355C, R366C and S399C mutants were not used to collect PRE-derived distance restraints. The substituted cysteine sites of S269C and S399C mutants were not used to collect PRE-derived distance restraints since the substituted cysteine sites are located in flexible regions. G321C, Q341C and G355C mutationants caused drastic chemical shift changes. In the case of Q331C mutant, ^{1}H-^{15}N HSQC spectra could not be measured with sufficient signal-to-noise ratios. The ^{1}H-^{15}N HSQC

spectrum of R366C mutant globally altered from the oxidized to reduced states (Fig. 18). For the residual twelve mutants, distances between the lone spin of MTSL and amide protons were estimated based on signal intensity ratios of the ^{1}H-^{15}N HSQC spectra in the oxidized and reduced states (Fig. 19).

We decided that PRE-derived distance restraints for amino acid residues located at flexible regions were not applied to the structure calculations.

Initially, PRE-derived distance restraints which were obtained with each mutant were employed to the structure calculations in a one-by-one manner. None of the twelve sets of structures showed significant improvement of RMSD values and high RDC correlation coefficients (Table. 4). This result suggests that PRE-derived distance restraints obtained from one spin label alone were not sufficient to improve the quality of protein structure calculation. Then, all of the PRE-derived distance restraints obtained from nine mutants (T280C, S282C, R291C, S301C, K312C, L332C, S371C, T384C and A394C) were employed in the structure calculations at once. These mutants provided a total of 163 upper and 704 lower distance limits (Table. 5; as shown in "PRE distance restraints" and "RMSD" value). In this case, accuracy and convergence of the generated structures were drastically improved (Fig. 20). In the structure calculations described above, a total number of NOE upper distance limits were increased after the final step of the automated NOE assignment cycle from 924 to 967 by applying PRE-derived distance restraints.

Through NOE signals were observed throughout the structured region of the protein, distance restraints which were converted from the NOESY spectra were sparse especially in the N-terminal region (Figure Fig. 21-a and 20-b). In addition, the NOE-derived long-range distance restraints were relatively focused on the specific amino acid residues (Figure Fig. 21-b). The PRE-derived upper distance limits used for structure calculations were also mapped on the determined structure (Figure Fig. 21-c). The PRE-derived upper distance limits were distributed throughout the structured region as if it complements shortage of distance restraints on the N-terminal region (Figure Fig. 21-d).

In an effort to evaluate the contribution of the PRE-derived distance restraints to the structure determination, structure calculations were performed in the presence or absence of PRE-derived distance restraints using the list of NOE-based upper distance limits used to determine the structure shown in Fig. 20. By employing the PRE-derived distance restraints, the RMSD values were improved from 1.51 ± 0.59 Å to 0.98 ± 0.20 Å, and the correlation coefficient between experimental RDC values and back-calculated one from the generated structures were improved from 0.59 ± 0.04 to 0.87 ± 0.05, respectively (Fig. 22). This result indicates that the PRE-derived distance restraints could serve as highly respectable distance constraints and could complement the insufficient NOE-based distance restraints, though its restraint force was not so strong due to the large error range of ± 7 Å.

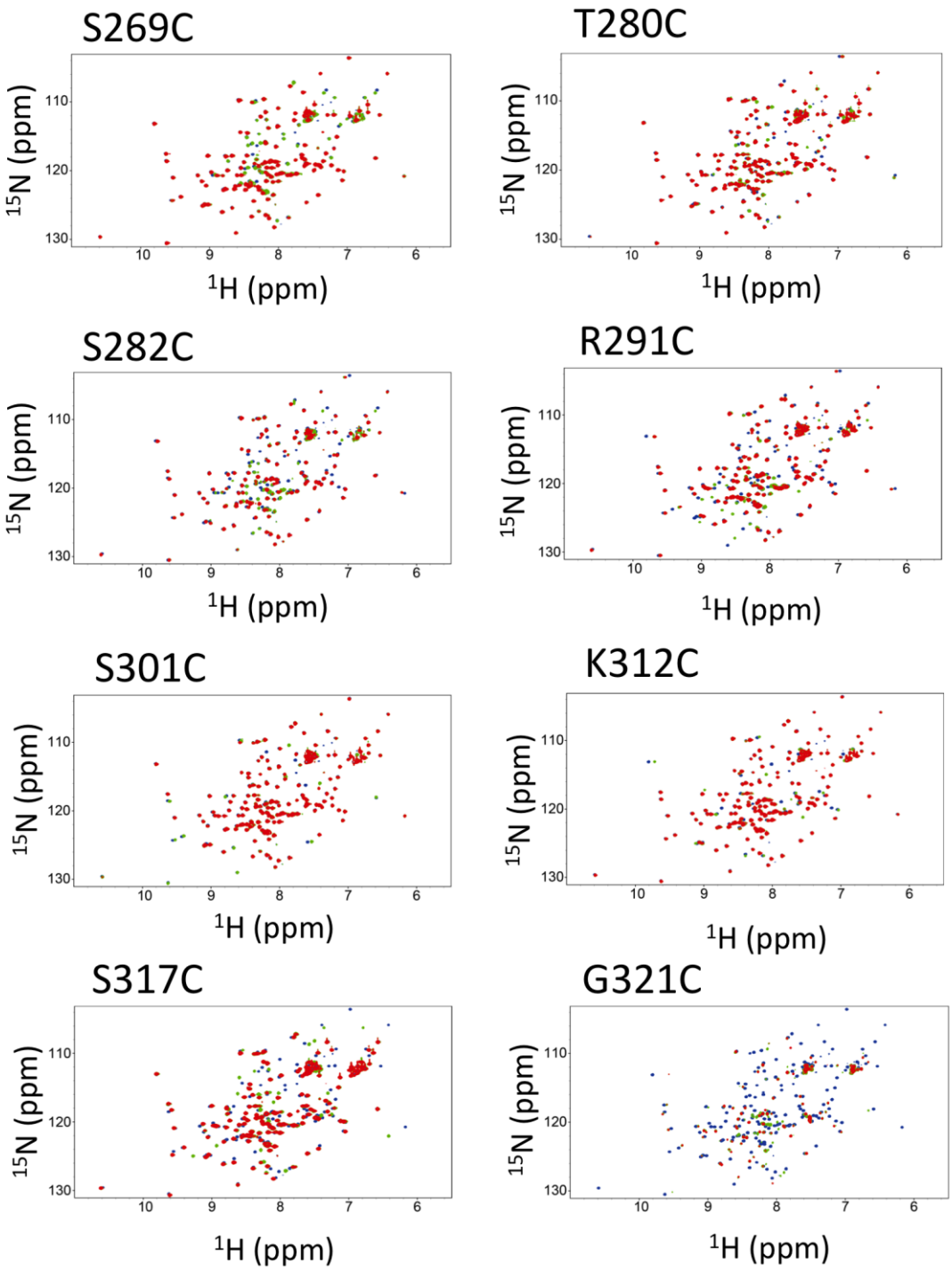
Evaluation of effective minimum number of PRE restraints

In order to examine the effective number of PRE-derived distance restraints, the PRE-derived distance restraints (163 and 704 of upper and lower distance limits, respectively) were decreased in stage as 87.5, 75.0, 62.5, 50.0, 37.5, 25.0 and 12.5 %, and structure calculations were performed in each condition. As a result, convergence of structure calculations and correlation between experimental RDC values and back-calculated ones from generated structures, were exponentially improved by applying ~50% number of the PRE-derived distance restraints, and its improvements were moreover promoted by applying 50-100% number of the PRE-derived distance restraints (Fig. 23). In an effort to achieve further improvement of the structure, additional 5 single cysteine mutants (S317C, G321C, G355C, F361C and A386C) were designed. These mutants were designed based on the determined structure so that PRE-derived distance restraints for the residues that were poor in PRE-derived distance restraints could be obtained. Among these mutants, G321C and G355C mutants caused drastic chemical shift changes. Therefore these mutants were not used to collect PRE-derived distance restraints.

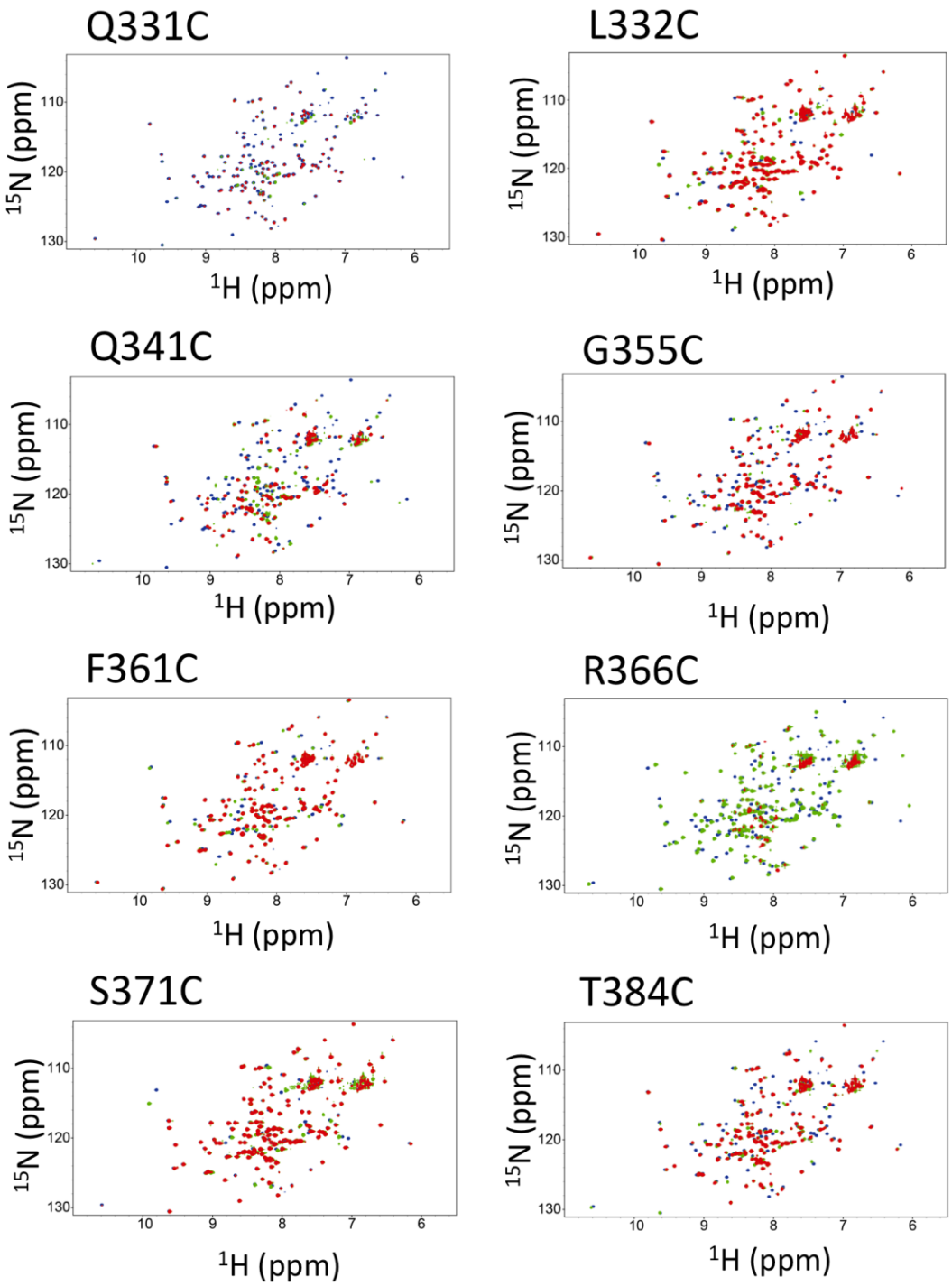
Table. 4. Backbone RMSDs and RDC correlation coefficients of structures determined in the presence of PRE-derived distance restraints derived from any one spin-labeled sample or 9 spin-labeled samples.

labeled residue	backbone RMSD	RDC correlation coefficient
280	2.32 ± 0.29	0.64 ± 0.10
282	2.35 ± 0.73	0.60 ± 0.10
291	2.80 ± 0.58	0.73 ± 0.11
301	3.70 ± 0.77	0.49 ± 0.15
312	2.40 ± 0.51	0.62 ± 0.06
317	2.92 ± 0.74	0.62 ± 0.14
332	1.96 ± 0.85	0.38 ± 0.20
361	3.81 ± 0.97	0.57 ± 0.11
371	3.60 ± 0.98	0.64 ± 0.07
384	3.19 ± 0.84	0.49 ± 0.11
386	3.05 ± 0.53	0.72 ± 0.09
394	2.67 ± 0.77	0.58 ± 0.05
9 residues[a]	3.06 ± 0.89	0.91 ± 0.17

[a]Residues 280, 282, 291, 301, 312, 332, 371, 384 and 394.



(**Fig. 18**, continues to the next page)



(**Fig. 18**, continues to the next page)

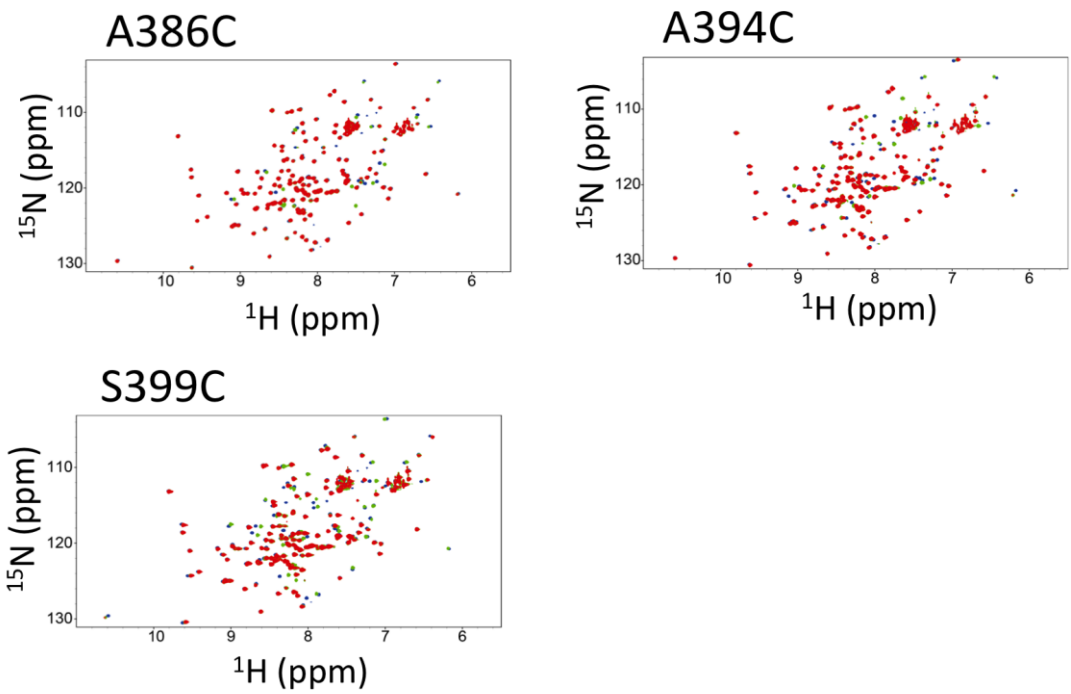
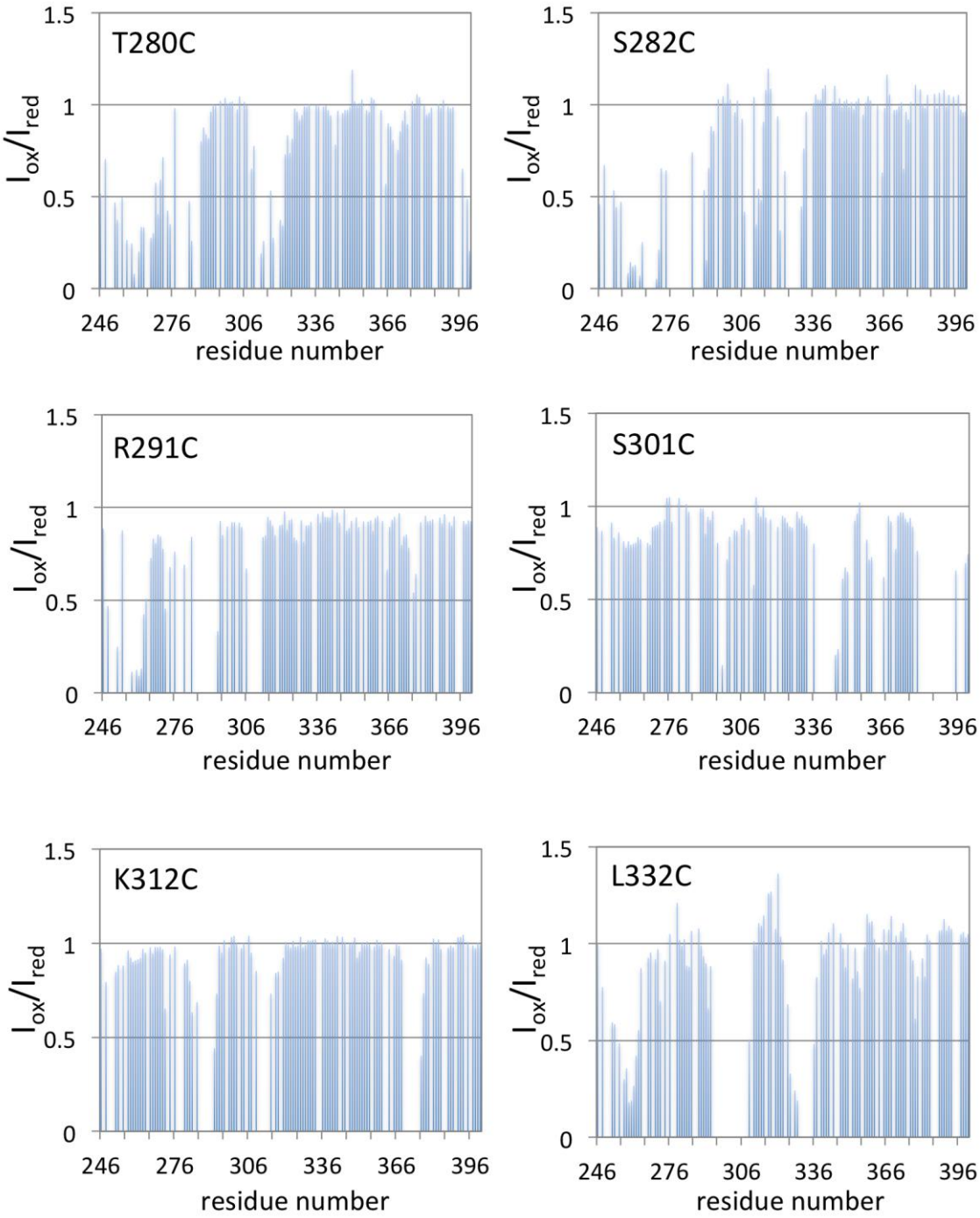


Fig. 18. Overlay of ^{1}H-^{15}N HSQC spectra of $Sin1_{CRIM}$.

WT (blue), MTSL attached mutant in the reduced state (green) and oxidized state (red).



(**Fig. 19**, continues to the next page)

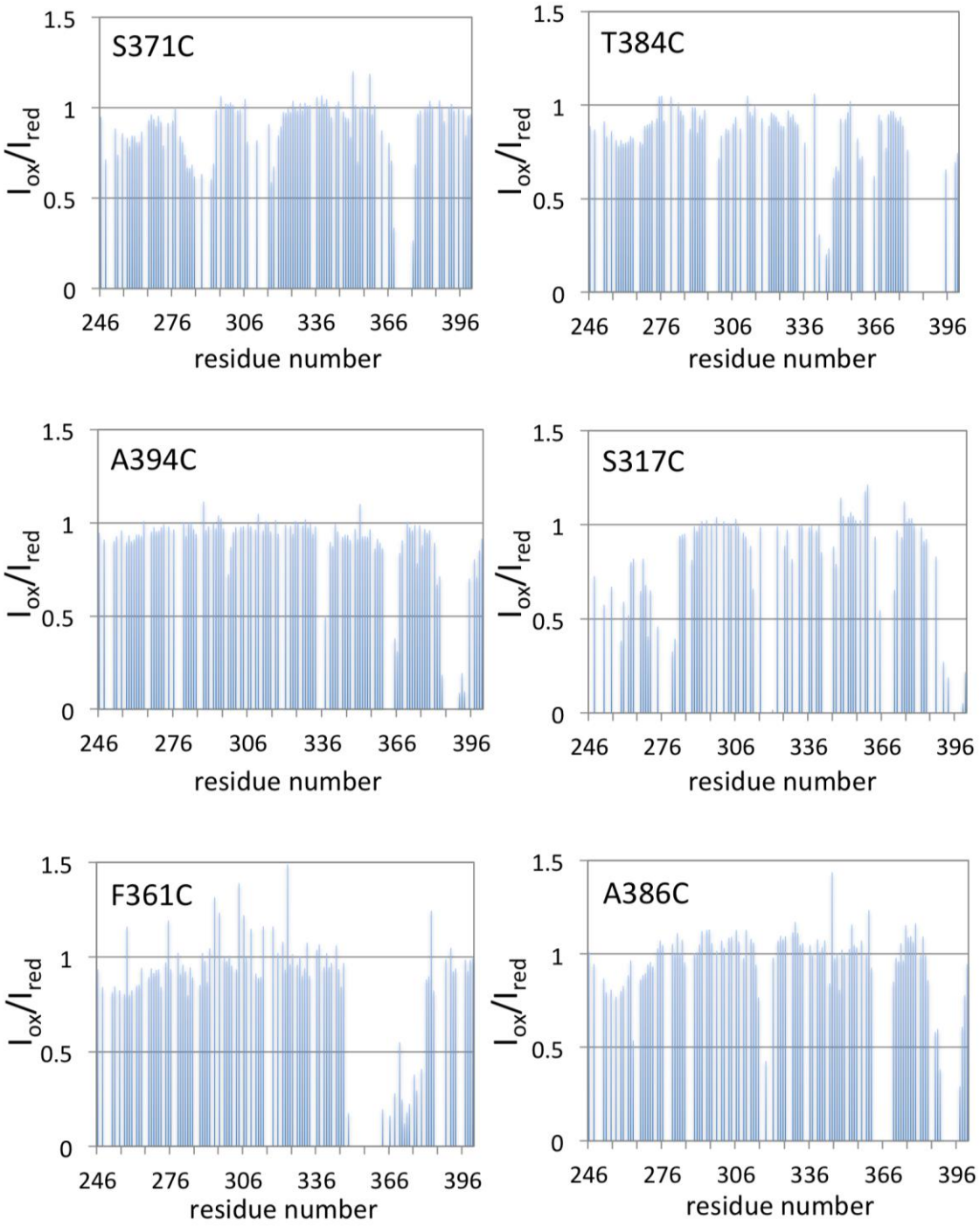


Fig. 19. Intensity ratios of ^{1}H-^{15}N HSQC peaks in the oxidized states against to those in the reduced states.

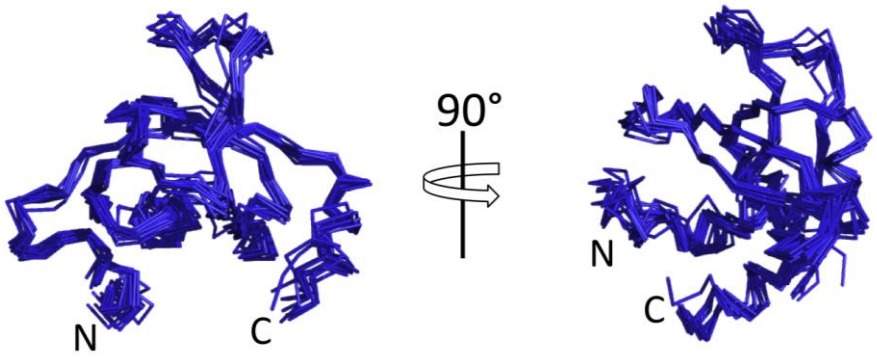


Fig. 20. The superposition of the final 10 structures calculated by the program CYANA with PRE distance restraints.

Table .5. Structural statistics for the NMR structures of $Sin1_{CRIM}$ calculated with PRE distance restraints.The RMSD values of backbone atoms addition to PRE data

NOESY peaks	
^{13}C	3911
^{15}N	1250
Total	5161
Completeness of chemical shift assignments	
	84.80%
Dihedral angle restraints	
φ	110
ϕ	90
χ1	12
PRE distance restraints	
Upper	163
Lower	704
Total	867
NOE upper distance limit	
short $\|i-j\| \leq 1$	636
middle $1 < \|i-j\| \leq 5$	132
long $5 \leq \|i-j\|$	199
Total	967
RMSD (amino acids 275-395) Å	
backbone	0.91 ± 0.17
all heavy	1.37 ± 0.19
RDC correlation	0.86 ± 0.05

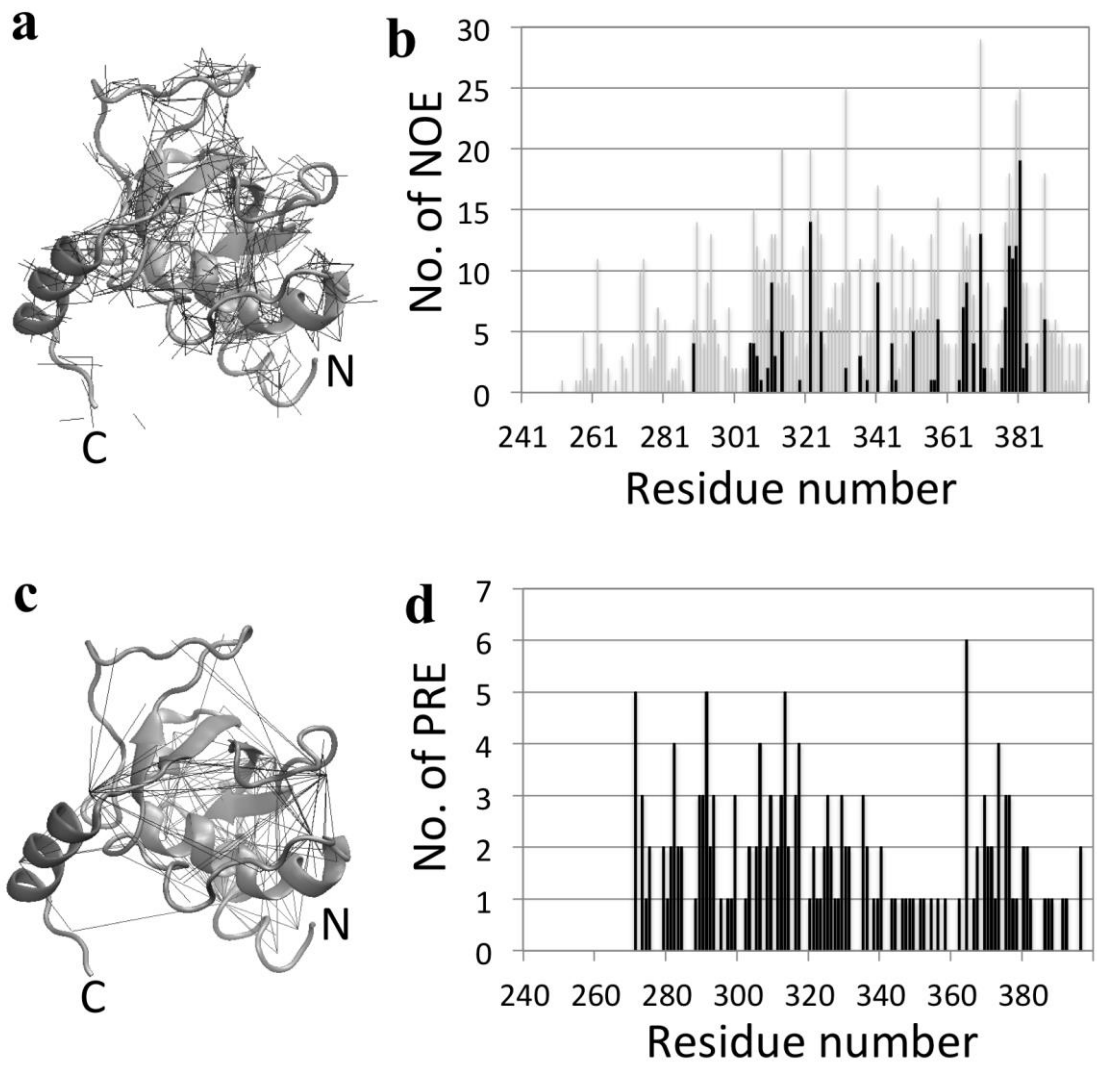


Fig. 21. Distance restraints which generated from NOE and PRE data.

NOEs (a) and PREs (c) used for the structure determination of $Sin1_{CRIM}$. NOEs and PREs are shown by lines on the lowest target function structure of $Sin1_{CRIM}$. (b, d) The number of NOEs (b) and PREs (d) used for the structure determination of $Sin1_{CRIM}$. Long-range NOEs are shown in black in (b).

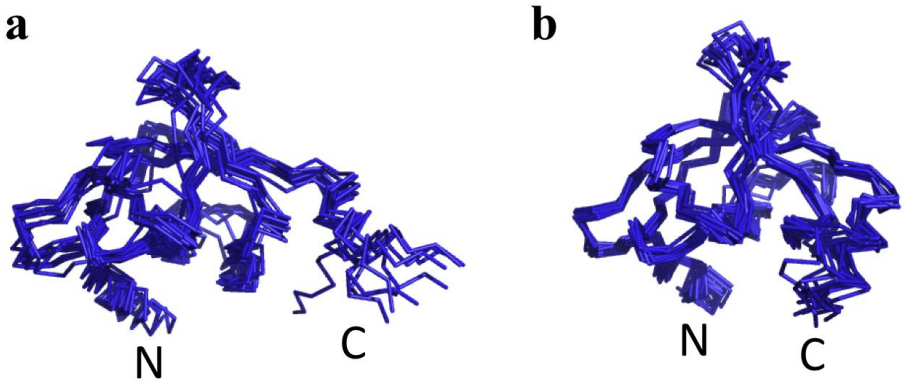


Fig 22. Superimposition of the final 10 structures which calculated in the absence (a) or presence (b) of PRE-derived distance restraints.

NOE upper distance limits, which were created through the automated NOE assignments in the presence of PRE-derived distance restraints, were applied to the both calculations.

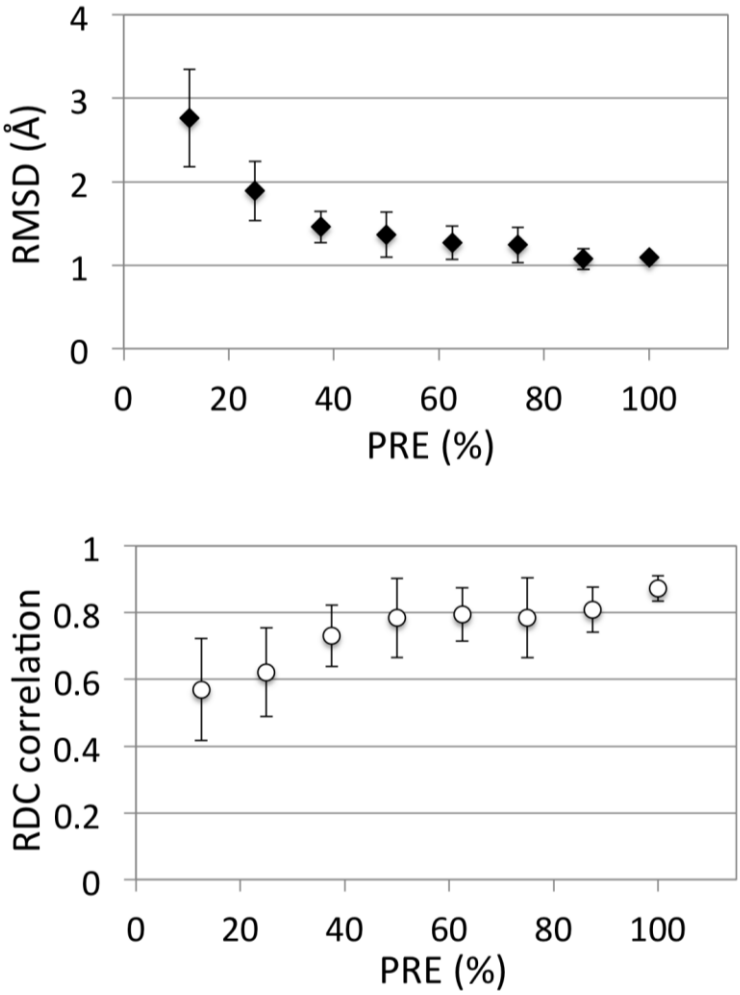


Fig. 23. Evaluation of effective minimum number of PRE.

Structure calculations were performed in the presence of 12.5, 25.0, 37.5, 50.0, 62.5, 75.0, 87.5 and 100 % PRE-derived distance restraints. The RMSD values of backbone atoms and the correlation coefficients between experimental RDC values and back-calculated ones from the generated structures.

3-4. Discussion

First, we have successfully assigned 94%, 94%, 82% and 80% of resonances of $Sin1_{CRIM}$ for backbone ^{1}H and ^{15}N, side chain ^{1}H and ^{13}C, respectively.

Next, in the structural calculations of $Sin1_{CRIM}$ using a conventional method, the RMSD values of the structured region (amino acids 275-395) were was 3.06 ± 0.89 Å. We could not determine the structure of $Sin1_{CRIM}$ with sufficient convergence and accuracy, because of the NOE data was insufficient for structure determination. In addition, the superposition of the final 10 structures calculated by the program CYANA showed that *sp* $Ssin1_{CRIM}$ is a loop-rich protein.

Thenrefore, we introduced nitroxide spin label MTSL into several different positions in the protein, and obtained distance restraints from paramagnetic relaxation enhancement (PRE) caused by the unpaired electron. By using PRE distance restraints derived from spin labels at nine different positions combined with NOE distance restraints, we successfully determined the structure of $Sin1_{CRIM}$ with sufficient convergence and accuracy.

Finally, the RMSD values of the structured region (amino acids 275-395) were was 0.91 ± 0.17 Å. We have succeeded in the structure determination of *sp* $Sin1_{CRIM}$ which had a new folding.

4. General conclusion

In the structural analysis of proteins Loop-rich using a solution NMR, first, the construction of expression system was important. In this study, it was particularly important that we obtain a large amount of soluble protein. In this study, we succeeded in producing the soluble proteins efficiently by using the pCold-GST vector. The key here is that we have succeeded in producing loop-rich protein with soluble protein, and this point is due to the use of the pCold-GST vector. This vector is very effective for loop-rich protein which expression level is small and insoluble protein. Thus, by using pCold-GST vector, we now allowed for perform a structure determination of loop-rich proteins under conditions similar to that in vivo. Therefore, the vector selection according to the protein is important.

Further, MBP tag (42.5 kDa) or His tag (1 kDa) are used commonly as a tag to promote the solubility of the protein, besides GST tag (28 kDa). When protein expression level is low by a pCold-GST vector, and it may be good to try these availability tags. But, it takes care when to use MBP tag having a large molecular weight. Under the influence of MBP with a large molecular

weight, the purpose protein may turn into soluble protein temporarily. However, it does not have solubility in practice. It is shown that protein was still insolubility. In this study, we used also a vector containing the MBP tag or the HIS tag with the cold shocking vector. However, the target protein has gone to aggregation when we separated MBP tag from the target protein. Therefore, it is necessary to pay attention to the state of protein during the sample preparation.

Next, the assignment rate is important for refinement the structure for loop-rich protein. In particular, the number of NOE assignments has a great influence on the refinement of the structure. In this study, except for the loop portion of the N-terminal, assignment rate was high. However, although the reason was not found, the assignment rate of the side chain was not enough. For the reason, assignment of NOE was insufficient, but we supplemented for shortage of distance information by using PRE.

In addition, usually, loop region is often be cut when sample preparation from the difficulty of assignment peaks. Therefore we believe that this established method have a major impact.

In this method, in order to obtain a PRE-derived distance constraints, there is also a demerit of spend the time and effort to introduces unpaired electron with MTSL. However, it is a big merit that it can be used for protein with insufficient NOE assignment. In fact, we are using for the other loop-rich proteins which could not obtain the distance information derived NOE, using the established method in this study. The presented method improved the protein structure determination. The RMSD of 4.5 Å for the structure determined with NOE distances was reduced to 1.3 Å by adding the PRE derived distances. As a result, we have succeeded in gradually to the refinement of the structure. We believe that this studying is useful in the elucidation of the motion of protein in vivo or the interaction mechanism of the ligand.

References

Amrein KE., Takacs B., Stieger M., Molnos J., Flint NA., and Burn P.(1995). Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an Escherichia coli expression system overproducing the bacterial chaperones GroES and GroEL. *Proc Natl Acad Sci U S A*. 92(4), 1048-1052.

Bokoch, M.P., Zou, Y.Z., Rasmussen, S.G.F., Liu, C.W., Nygaard, R., Rosenbaum, D.M., Fung, J.J., Choi, H.J., Thian, F.S., Kobilka, T.S., et al. (2010). Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463, 108-U121.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G.F., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., et al. (2007). High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor. *Science* 318, 1258-1265.

Cordingley MG., Callahan PL., Sardana VV., Garsky VM., and Colonno RJ.(1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J Biol Chem*. 265(16):9062-9065.

Day, PW., Rasmussen, SG., Parnot, C., Fung, JJ., Masood, A., Kobilka, TS., Yao, XJ., Choi, HJ., Weis, WI., Rohrer, DK., and Kobilka, BK. (2007). A monoclonal antibody for G protein-coupled receptor crystallography. *Nat Methods*. 4, 927-929.

Gloriam, DE., Fredriksson, R., and Schiöth, HB. (2007). The G protein-coupled receptor subset of the rat genome. *BMC Genomics*. 8, 338-405.

Grace, CR., Perrin, MH., DiGruccio, MR., Miller, CL., Rivier, JE., Vale, WW., and Riek, R. (2004). NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proc Natl Acad Sci U S A*. 101(35), 12836-12841.

Grace, CR., Perrin, MH., Gulyas, J., Digruccio, MR., Cantle, JP., Rivier, JE., Vale, WW., and Riek, R. (2007). Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand. *Proc Natl Acad Sci U S A*.. 104, 4858-4863.

Hayashi, K., and Kojima, C. (2008). pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. *Protein Expr Purif*. 62, 120-127.

Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, FE., and Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res*. 31(1), 294-297.

Lundstrom, K. (2005). Structural genomics of GPCRs. *Trends Biotechnol.* 23, 103-108.

Mishima, M., Wakabayashi, S., and Kojima, C. (2007). Solution structure of the cytoplasmic region of Na+/H+ exchanger 1 complexed with essential cofactor calcineurin B homologous protein 1. *J Biol Chem.* 282, 2741-2751.

Mujacic, M., Cooper, KW., and Baneyx, F. (1999). Cold-inducible cloning vectors for low-temperature protein expression in Escherichia coli: application to the production of a toxic and proteolytically sensitive fusion protein. *Gene*. 238, 325-332.

Murakami, M., and Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature*. 453(7193), 363-367.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, CA., Motoshima, H., Fox, BA., Trong, I.L., Teller, DC., Okada, T., Stenkamp, RE., Yamamoto, M., and Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*. 289(5480), 739-745.

Park, JH., Scheerer, P., Hofmann, KP., Choe, HW., and Ernst, OP. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*. 454(7201), 183-187.

Parthier, C., Kleinschmidt, M., Neumann, P., Rudolph, R., Manhart, S., Schlenzig, D., Fanghänel, J., Rahfeld, JU., Demuth, HU., and Stubbs, MT. (2007). Crystal structure of the incretin-bound extracellular domain of a Gprotein-coupled receptor. *Proc Natl Acad Sci U S A.* 104.

Pioszak AA., and Xu HE. Molecular recognition of parathyroid hormone by its G protein-coupled receptor.(2008). *Proc Natl Acad Sci U S A.* 105(13):5034-5039.

Pioszak AA., Parker NR., Suino-Powell K., and Xu HE.(2008). Molecular recognition of corticotropin-releasing factor by its G-protein-coupled receptor CRFR1. *J Biol Chem.* 283(47), 32900-32912.

Qing, G., Ma, LC., Khorchid, A., Swapna, GV., Mal, TK., Takayama, MM., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, GT., Ikura, M., and Inouye, M. (2004). Cold-shock induced high-yield protein production in Escherichia coli. *Nat Biotechnol.* 22, 877-882.

Rasmussen, S.G.F., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R.P., Sanishvili, R., Fischetti, R.F., et al. (2007). Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor. *Nature* 450, 383-U384.

Rasmussen, S.G.F., Choi, H.-J., Fung, J.J., Pardon, E., Casarosa, P., Chae, P.S., DeVree, B.T., Rosenbaum, D.M., Thian, F.S., Kobilka, T.S., et al. (2011). Structure of a nanobody-stabilized active state of the [bgr]2 adrenoceptor. *Nature* 469, 175-180.

Rosenbaum, DM., Cherezov, V., Hanson, MA., Rasmussen, SG., Thian, FS., Kobilka, TS., Choi, HJ., Yao, XJ., Weis, WI., Stevens, RC., and Kobilka, BK. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. 318, 1253-1254.

Runge, S., Thøgersen, H., Madsen, K., Lau, J., and Rudolph, R. (2008). Crystal structure of the ligand-bound glucagon-like peptide-1 receptor extracellular domain. *J Biol Chem*. 283(17), 11340-11347.

Shimamura, T., Hiraki, K., Takahashi, N., Hori, T., Ago, H., Masuda, K., Takio, K., Ishiguro, M., and Miyano, M. (2008). Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J Biol Chem*. 283(26), 17753-17756.

Sun, C., Song, D., Davis-Taber, RA., Barrett, LW., Scott, VE., Richardson, PL., Pereda-Lopez, A., Uchic, ME., Solomon, LR., Lake, MR., Walter, KA., Hajduk, PJ., and Olejniczak, ET. (2007). Solution structure and mutational analysis of pituitary adenylate cyclase-activating polypeptide binding to the extracellular domain of PAC1-RS. *Proc Natl Acad Sci U S A*. 104, 7875-7880.

Wilkins MR., Gasteiger E., Bairoch A., Sanchez JC., Williams KL., Appel RD., and Hochstrasser DF.(1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. **112**, 531-552.

Battiste, J. L. & Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39: 5355-65.

Cybulski, N., Hall, M. N. (2009) TOR complex 2: a signaling pathway of its own. Trends Biochem Sci 34: 620-627.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6: 277-293.

Facchinetti, V., Ouyang, W., Wei, H., Soto, N., Lazorchak, A., Gould, C., Lowry, C., Newton, A. C., Mao, Y., Miao, R. Q., Sessa, W. C., Qin, J., Zhang, P., Su, B., Jacinto, E. (2008) The mammalian target of rapamycin complex 2 controls folding and stability of Akt and protein kinase C. EMBO J 27: 1932-19243.

Farrow, N., Muhandiram, R., Singer, A., Pascal, S., Kay, C., Gish, G., Shoelson, S., Pawson, T., Formankay, J. & Kay, L. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src Homology-2 domain studied by N-15 nmr relaxation. Biochemistry 33: 5984-6003.

Frias, M. A., Thoreen, C. C., Jaffe, J. D., Schroder, W., Sculley, T., Carr, S. A., Sabatini, D. M. (2006). mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. Curr Biol 16: 1865-1870.

García-Martínez, J. M., Alessi, D. R. (2008) mTOR complex 2 (mTORC2) controls hydrophobic motif phosphorylation and activation of serum- and glucocorticoid-induced protein kinase 1 (SGK1). Biochem J 416: 375-385.

Gottstein, D., Reckel, S., Dötsch, V., Güntert, P. (2012). Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. Structure 20: 1019-1027.

Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. J Biomol NMR 3: 185-204.

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

Hayashi, K. & Kojima, C. (2008). pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. Protein Expr Purif 62: 120-127.

Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319: 209-227.

Hresko, R. C., Mueckler, M. (2005) mTOR.RICTOR is the Ser473 kinase for Akt/protein kinase B in 3T3-L1 adipocytes. J Biol Chem 280: 40406-40416.

Hu, J. S. & Bax, A. (1997). Chi 1 angle information from a simple two-dimensional NMR experiment that identifies trans $^{3}J_{NC}$ gamma couplings in isotopically enriched proteins. J Biomol NMR 9: 323-328.

Huang, Y. J., Swapna, G. V., Rajan, P. K., Ke, H., Xia, B., Shukla, K., Inouye, M. & Montelione, G. T. (2003). Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from Escherichia coli. J Mol Biol 327: 521-536.

Jacinto, E., Facchinetti, V., Liu, D., Soto, N., Wei, S., Jung, S. Y., Huang, Q., Qin, J., Su, B. (2006) SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity. Cell 127: 125-137.

Jee, J. & Güntert, P. (2003). Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. J Struct Funct Genom 4: 179-189.

Kay, L. E., Torchia, D. A., Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28: 8972-8979.

Kay, L., Xu, G., Singer, A., Muhandiram, D. & Formankay, J. (1993). A gradient-enhanced HCCH TOCSY experiment for recording side-chain H-1 and C-13 correlations in H_2O samples of proteins. Journal of Magnetic Resonance Series B 101: 333-337.

Lyons, B. & Montelione, G. (1993). An HCCNH triple-resonance experiment using C-13 isotropic mixing for correlating backbone amide and side-chain aliphatic resonances in isotopically enriched proteins. Journal of Magnetic Resonance Series B 101: 206-209.

Macura, S., Huang, Y., Suter, D. & Ernst, R. (1981). Two-dimensional chemical-exchange and cross-relaxation spectroscopy of coupled nuclear spins. Journal of Magnetic Resonance 43: 259-281.

Madl, T., Felli, I. C., Bertini, I. & Sattler, M. (2010). Structural analysis of protein interfaces from ^{13}C direct-detected paramagnetic relaxation enhancements. J Am Chem Soc 132: 7285-7287.

Mark V. Berjanskii and David S. Wishart. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. Nucleic Acids Res 35;W531-W537

Matsuo, T., Kubo, Y., Watanabe, Y., Yamamoto, M. (2003) Schizosaccharomyces pombe AGC family kinase Gad8p forms a conserved signaling module with TOR and PDK1-like kinases. EMBO J 22: 3073-3083.

Montelione, G., Lyons, B., Emerson, S. & Tashiro, M. (1992). An efficient triple resonance experiment using C-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. Journal of the American Chemical Society 114: 10974-10975.

Muhandiram, D., Farrow, N., Xu, G., Smallcombe, S. & Kay, L. (1993). A gradient C-13 NOESY-HSQC experiment for recording noesy spectra of C-13-labeled proteins dissolved in H_2O. Journal of Magnetic Resonance Series B 102: 317-321.

Nilges, M. (1997). Ambiguous distance data in the calculation of NMR structures. Fold Des 2: S53-S57.

Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. Journal of Magnetic Resonance 131: 373-378.

Reckel, S., Gottstein, D., Stehle, J., Loehr, F., Verhoefen, M.-K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., Bernhard, F., Schwalbe, H., Guentert, P. & Doetsch, V. (2011). Solution NMR Structure of Proteorhodopsin. Angewandte Chemie-International Edition 50: 11942-11946.

Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R. & Choe, S. (2005). NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science 307: 1317-1321.

Rovnyak, D., Frueh, D. P., Sastry, M., Sun, Z. Y., Stern, A. S., Hoch, J. C. & Wagner, G. (2004). Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. J Magn Reson 170: 15-21.

Sarbassov, D. D., Guertin, D. A., Ali, S. M., Sabatini, D. M. (2005) Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. Science 307: 1098-1101.

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

Sven G. H., Scotto A. R., Gerhard Wagner. (2013) Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J of Biomol NMR 55, Issue2: 167-178

Van Horn, W. D., Kim, H.-J., Ellis, C. D., Hadziselimovic, A., Sulistijo, E. S., Karra, M. D., Tian, C., Soennichsen, F. D. & Sanders, C. R. (2009). Solution Nuclear Magnetic Resonance Structure of Membrane-Integral Diacylglycerol Kinase. Science 324: 1726-1729.

Wishart DS., Sykes BD. (1994) Chemical shifts as a tool for structure determination. Meth. Enzymol 239: 363–392.

Wullschleger, S., Loewith, R., Hall, M. N. (2006) TOR signaling in growth and metabolism. Cell 124: 471-484.

Yang, Q., Inoki, K., Ikenoue, T., Guan, K. L. (2006) Identification of Sin1 as an essential TORC2 component required for complex formation and kinase activity. Genes Dev 20: 2820-2832.

Yang, Y., Ramelot, T. A., McCarrick, R. M., Ni, S., Feldmann, E. A., Cort, J. R., Wang, H., Ciccosanti, C., Jiang, M., Janjua, H., Acton, T. B., Xiao, R., Everett, J. K., Montelione, G. T. & Kennedy, M. A. (2010). Combining NMR and EPR methods for homodimer protein structure determination. J Am Chem Soc 132: 11910-11913.

Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994). Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR 4: 845-858.

Zhou, Y., Cierpicki, T., Jimenez, R. H., Lukasik, S. M., Ellena, J. F., Cafiso, D. S., Kadokura, H., Beckwith, J. & Bushweller, J. H. (2008). NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. Mol Cell 31: 896-908.

Zweckstetter, M. & Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. Journal of the American Chemical Society 122: 3791-3792.