



Yohei Miyanoiri, Mitsuhiro Takeda, and Masatsune Kainosho

Contents

Introduction	470
Methyl Labeling	471
The Stereo-Array Isotope Labeling (SAIL) Method	474
NMR Studies of Proteins Using Residue-Selective Labeling	478
Large-Amplitude Slow Breathing Motion of Proteins Probed by the Ring Flipping Motions of Phenylalanine and Tyrosine Residues Embedded in the Hydrophobic Core	479
Hydrogen Exchange Study of Polar Side-Chain Groups of Proteins	480
NMR Studies of Conformations and Isomerizations of Protein Disulfide Bonds Using Proteins Selectively Labeled with Site-Specifically ¹³ C-Labeled Cysteine	482
Concluding Remarks	484
References	484

Y. Miyanoiri

Structural Biology Research Center, Graduate School of Science, Nagoya University, Nagoya,
Japan

e-mail: miyanoiri.youhei@h.mbox.nagoya-u.ac.jp

M. Takeda

Structural Biology Research Center, Graduate School of Science, Nagoya University, Nagoya,
Japan

Department of Structural BioImaging, Faculty of Life Sciences, Kumamoto University, Kumamoto,
Japan

e-mail: takeda@structbiol.com

M. Kainosho (✉)

Structural Biology Research Center, Graduate School of Science, Nagoya University, Nagoya,
Japan

Graduate School of Science and Engineering, Tokyo Metropolitan University, Tokyo, Japan

e-mail: kainosho@tmu.ac.jp

Abstract

Boundless progress in isotope-aided NMR methods still continues to provide the driving force for developing novel NMR strategies for structural biology research of proteins. In the first edition of this book, we described an overview of the isotope labeling methods available at that time. In this second edition, we will mainly focus on newer isotope-aided NMR methods, such as the methyl-specific labeling and stereo-array isotope labeling (SAIL) methods, which have rapidly developed during the past decade. The methyl-specific labeling is currently used as the most practical technique applicable to large protein complexes and membrane proteins. The standard methyl labeling protocols employ isotope-labeled α -keto acid precursors, which enable selective observations of the methyl groups of Ile, Leu, and Val residues. More recently, the stereo-specific isotope labeling methods of *prochiral* methyl groups have become available, using either regio-selectively isotope-labeled precursors or stereo-specifically $^{13}\text{CH}_3$ -labeled amino acids. We also focus on the stereo-array isotope labeling (SAIL) method, which is a breakthrough isotope labeling technology using stereo- and regio-selectively [^2H , ^{13}C , ^{15}N]-labeled amino acids with isotope labeling patterns optimized for NMR studies. Various applications of SAIL and related methods to structural studies, including protein dynamics such as aromatic ring-flipping motions, hydrogen-deuterium exchange rates, conformational analysis, and dynamics about disulfide bonds, will be discussed.

Keywords

Isotope-aided NMR methods · Stereo-array isotope labeling (SAIL) method · Large proteins · Residue selective labeling · Aromatic ring NMR signal · Stereo-specific methyl labeling · Aromatic ring flipping motion · Large-amplitude slow breathing motion (LASBM) · Hydrogen exchange rates for side-chain polar groups · Deuterium induced isotope shifts · Disulfide bond isomerization

Introduction

Over the history of biomolecular NMR, stable isotope labeling methods have always played key roles in providing the foundation for protein structural investigations. The range of biomolecular NMR studies has expanded along with the development of protein isotope labeling methods [1–5]. Numerous isotope labeling schemes are now available for NMR studies. The conventional [^{13}C , ^{15}N] labeling approach, which is widely employed in NMR investigations of proteins smaller than 25 kDa, allows the application of heteronuclear multidimensional experiments to resolve highly overlapped proton NMR signals from each other, by correlating the heteronuclei chemical shifts of the attached ^{15}N and/or ^{13}C signals [6, 7]. As a result, it is now almost routine work to determine the three-dimensional structures of small proteins using [^{13}C , ^{15}N] samples, which can readily be prepared by growing *E. coli* cells harboring the targeted protein genes in minimal medium containing

[^{13}C]-glucose and/or [^{15}N]-ammonium salt as the sole carbon and nitrogen sources, respectively. However, the largest size amenable to NMR structure determination using uniformly [^{13}C , ^{15}N] proteins is around 25 kDa, due to the excessive signal overlapping and line broadening. To address the size problem, protein deuteration to improve the spectral quality has been explored. The substitution of ^1H to ^2H mitigates the unwanted dipolar and scalar couplings in proteins [8]. In concert with transverse relaxation spectroscopy (TROSY), the size limitation for NMR investigations has been somewhat relieved [9]. In this chapter, we will mainly focus on the further developments along this direction since the turn of the century, i.e., selective methyl labeling, stereo-array isotope labeling (SAIL), and other NMR methods using advanced site-specific isotope labeling technologies.

Methyl Labeling

The methyl-containing amino acids, i.e., alanine, isoleucine, leucine, valine, methionine, and threonine, are important constituents of hydrophobic core structures and are also involved in ligand interactions and conformational changes in proteins. In addition, the NMR signal intensity of a methyl group is very strong, due to the high proton density and fast rotation around the methyl carbon axis. Therefore, methyl groups are valuable probes in NMR studies of the structures and dynamics of large proteins. During the past two decades, methyl-specific isotope labeling techniques on perdeuterated proteins have been progressing, and their combination with methyl transverse relaxation optimized spectroscopy (methyl TROSY) [10] has allowed structural and dynamic information to be obtained for protein particles as large as 1 MDa [11].

For producing methyl-specific-labeled proteins in the conventional *E. coli* protein expression system, isotope-labeled amino acid precursors are commonly used. Among them, α -ketobutyrate and α -ketoisovalerate are well-established precursors for the specific isotope labeling of the Ile $\delta 1$, Leu $\delta 1/\delta 2$, and Val $\gamma 1/\gamma 2$ methyl groups in proteins, respectively [12, 13]. A variety of isotope-labeled α -keto acids and their methyl isotopomers (i.e., $^{13}\text{CH}_3$, $^{13}\text{CH}_2\text{D}$, and $^{13}\text{CHD}_2$) have been developed [14] and are commercially available at affordable prices. Therefore, these precursors are widely used in various NMR experiments. In the case of using α -ketoisovalerate, in which the $\gamma 1$ or $\gamma 2$ -methyl is alternatively labeled with $^{13}\text{CH}_3$ and $^{12}\text{CD}_3$ (Fig. 1a), all inter-residue methyl-methyl NOEs for Leu and Val residues can be observed, because of the alternative methyl labeling scheme. In protein dynamics studies, $^{13}\text{CH}_2\text{D}$ or $^{13}\text{CHD}_2$ isotopomers might be useful for obtaining ^2H or ^1H relaxation information for the methyl groups [15, 16]. The sequence-specific assignment for Ile $\delta 1$, Leu $\delta 1/\delta 2$, and Val $\gamma 1/\gamma 2$ methyl signals can also be achieved by using COSY-based HN-detected and methyl-detected “out-and-back” experiments [17, 18]. For larger proteins, such as the 468 kDa PhTET2 complex, a systematic site-directed mutagenesis method was used for the methyl signal assignment [19]. However, these approaches alone cannot establish the stereospecific assignments for Leu $\delta 1/\delta 2$ and Val $\gamma 1/\gamma 2$ methyl groups. For the stereospecific assignment of these *prochiral* methyl

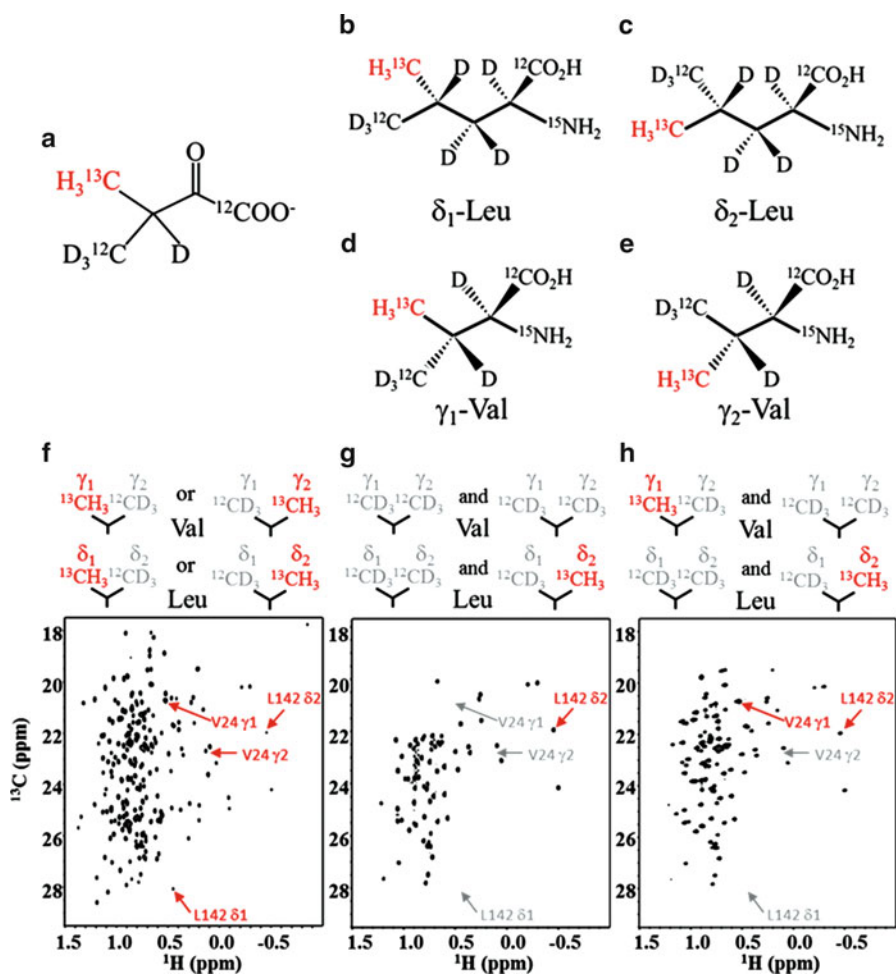


Fig. 1 Methyl-specific isotope labeling for Leu/Val residues. Structures of the amino acid precursor and the stereo-specifically ^{13}C -labeled leucines and valines. (a) $[3\text{-}^{13}\text{C}; 3,4,4,4\text{-}^2\text{H}_4]\text{-}\alpha\text{-ketoisovalerate}$, (b) $L\text{-}[\delta 1\text{-}^{13}\text{C}; \alpha, \beta, \gamma, \delta 2\text{-}^2\text{H}_7; \alpha\text{-}^{15}\text{N}]$ -leucine, $\delta 1\text{-Leu}$, (c) $L\text{-}[\delta 2\text{-}^{13}\text{C}; \alpha, \beta, \gamma, \delta 1\text{-}^2\text{H}_7; \alpha\text{-}^{15}\text{N}]$ -leucine, $\delta 2\text{-Leu}$, (d) $L\text{-}[\gamma 1\text{-}^{13}\text{C}; \alpha, \beta, \gamma 2\text{-}^2\text{H}_5; \alpha\text{-}^{15}\text{N}]$ -valine, $\gamma 1\text{-Val}$, and (e) $L\text{-}[\gamma 2\text{-}^{13}\text{C}; \alpha, \beta, \gamma 1\text{-}^2\text{H}_5; \alpha\text{-}^{15}\text{N}]$ -valine, $\gamma 2\text{-Val}$. 2D $^1\text{H}\text{-}^{13}\text{C}$ HMQC spectra of Leu/Val-labeled MSG, labeled with $[3\text{-}^{13}\text{C}; 3,4,4,4\text{-}^2\text{H}_4]\text{-}\alpha\text{-ketoisovalerate}$ (f), $\delta 2\text{-Leu}$ (g), and “ $\gamma 1\text{-Val} + \delta 2\text{-Leu}$ ” (h). The labeling patterns of the Leu/Val methyl groups are shown at the top of each spectrum. Signal assignments for Leu142 $\delta 1/\delta 2$ and V24 $\gamma 1/\gamma 2$ methyls are shown in each spectrum [18]

groups, additional experiments, such as biosynthetic fractional ^{13}C labeling or selective labeling with ^{13}C -“block”-labeled Leu and Val, are required [5, 20]. There are some other limitations of $\alpha\text{-keto}$ acid precursors for Ile, Leu, and Val methyl labeling. For example, $\alpha\text{-keto}$ butyrate cannot be applied for labeling the $\gamma 2$ methyl group of Ile. When using $\alpha\text{-ketoisovalerate}$, all of the Leu $\delta 1/\delta 2$ and Val $\gamma 1/$

γ 2 methyl groups are simultaneously labeled, causing serious signal overlapping, particularly in large proteins, and thus hampering precise signal assignments. Moreover, when the alternatively methyl ^{13}C -labeled α -ketoisovalerate is used, the labeling rates for each Leu δ 1/ δ 2 and Val γ 1/ γ 2 remain 50% or lower. As a consequence, the signal intensities of the inter-residue methyl-methyl NOEs, which are critical data for precise structural analyses of proteins, are only 25% of the maximum values. Therefore, the widely used α -ketoisovalerate is not ideal for the residue- and stereo-selective Leu/Val methyl labeling (Fig. 1a, f).

To overcome these problems, other methyl labeling methods have been developed. The precursor, γ - ^{13}C -aceto- α -hydroxybutyrate, can specifically label the Ile γ 2 methyl group in proteins and was successfully used for the sequence-specific signal assignments of the 360 kDa half proteasome complex [21]. For Leu selective labeling, methyl ^{13}C -labeled α -ketoisocaproate has been developed [22]. Among the several new precursors, the regio-specifically methyl ^{13}C -labeled acetolactate seems to be very useful for Leu and Val methyl labeling. Using γ - ^{13}C - α -acetolactate, in lieu of α -ketoisovalerate, the Val γ 2 and Leu δ 2 methyls are specifically labeled with high incorporation rates [23]. The specific labeling of the Val γ 1 and Leu δ 1 methyl groups in proteins can be achieved by using β - ^{13}C - α -acetolactate. Therefore, the overcrowded Val γ 1/ γ 2 and Leu δ 1/ δ 2 signals in 2D methyl TROSY are improved, and the methyl signal assignments and inter-residue methyl-methyl NOE analyses are achieved clearly and effectively, even in large proteins, such as 82 kDa MSG and 468 kDa PhTET2 [23].

Isotope-labeled amino acids for Val/Leu methyl signal analyses have also been developed. In early studies, stereo-selective deuterated leucine, (*2S,4R*) [5,5,5- $^2\text{H}_3$] leucine, was used for the stereospecific methyl signal assignment of Leu residues [24]. Recently, residue- and stereo-specific Val methyl isotope labeling has been achieved, by using ^{13}C - α -acetolactate in combination with uniformly deuterated leucine [25]. As a straightforward isotope labeling protocol for Leu and Val residues in a protein, stereo-specifically methyl-labeled leucine and valine were produced [26] (Fig. 1b–e). The *E. coli* cellular expression using these amino acids, in lieu of isotope-labeled amino acid precursors, allowed the preparation of proteins with labeled Leu δ 1/ δ 2 and/or Val γ 1/ γ 2 methyl groups in residue- and stereo-specific manners. Actually, when we prepared 82 kDa malate synthase G (MSG) samples using 2 mg of δ 2-Leu in 100 ml deuterated M9 medium (i.e., 20 mg/L δ 2-Leu), all 70 Leu residues in MSG gave highly sensitive and well-dispersed ^1H - ^{13}C HMQC signals for Leu δ 2, without serious signal overlapping (Fig. 1g). Therefore, the residue- and stereo-specific signal assignments can be achieved very efficiently and precisely. Moreover, this method allows the differential isotope labeling for Leu δ 1/ δ 2 and Val γ 1/ γ 2 methyl groups in any combination. For example, we could obtain Val γ 1 and Leu δ 2 methyl signals in the same HMQC spectrum, by using γ 1-Val and δ 2-Leu (Fig. 1h). This methyl labeling scheme cannot be achieved by using isotope-labeled amino acid precursors. The Leu-Val combinatorially labeled protein gives inter-residue methyl-methyl NOEs with all stereochemical pairs, which are extremely useful for obtaining precise structural information about large proteins. The incorporation efficiencies attained by adding relatively small amounts of labeled

Leu (20 mg/L) and Val (100 mg/L) are over 90% and 80%, respectively. Accordingly, highly sensitive measurements and unambiguous assignments of the inter-methyl NOEs for the Leu and Val residues of 82 kDa MSG were achieved [26].

Several practical protocols for preparing ^{13}C -methyl-labeled proteins have also been developed, using the cell-free protein synthesis system and the *E. coli* protein expression system. Recently, new auxotrophic *E. coli* strains have been created [27, 28]. An Ile, Leu, and Val auxotrophic strain of *E. coli*, derived from *E. coli* BL21(DE3), lacks the *ilvD* and *leuB* genes that encode dihydroxy acid dehydratase and β -isopropylmalate dehydrogenase, respectively. As a result, the biosynthetic pathways from pyruvate to L-isoleucine, L-leucine, and L-valine in this *E. coli* strain are completely aborted. Using this auxotrophic *E. coli* mutant strain for preparing ^{13}C -methyl-labeled proteins, the incorporation efficiencies of exogenous isotope-labeled isoleucine, leucine, and valine are nearly 100% without any scrambling, even with the supplementation of only 10 mg/L of each amino acid [28]. These methods are particularly useful when expensive isotope-labeled amino acids, such as stereo-array isotope-labeled (SAIL) amino acids, must be used for NMR studies, as described in the next section.

The Stereo-Array Isotope Labeling (SAIL) Method

The elucidation of the three-dimensional structure of a protein is an important focus of NMR studies. The determination of a protein structure is performed by collecting the NOE-derived distance restraints between the protons in the protein. If the size of the target protein is less than 25 kDa, then the structure determination can be performed using the uniformly $^{13}\text{C}/^{15}\text{N}$ -labeled (UL) protein. In this case, the backbone and side-chain resonances in the target protein are assigned by ^{13}C - or ^{15}N -edited multidimensional experiments and then NOESY experiments are applied. However, the application of this conventional method to proteins larger than 25 kDa may not be practical. With an increase in the molecular weight of a protein, the NMR analysis of the protein becomes more challenging due to the increased line broadening and serious overlapping of NMR signals. The stereo-array isotope labeling (SAIL) method was developed to solve these fundamental problems encountered in NMR studies of large proteins. The SAIL method employs proteins exclusively composed of stereo- and regio-specifically ^2H -, ^{13}C -, and ^{15}N -labeled amino acids (SAIL amino acids) for NMR experiments. The SAIL amino acids are chemically synthesized and then incorporated into a target protein by a cell-free protein expression system. The isotope labeling pattern profoundly improves the sensitivity and mitigates the signal overlap [29, 30]. In the methylene groups in SAIL proteins, for example, one of the two methylene protons is stereo-specifically substituted by ^2H . While the NOE peaks involving the substituted proton are undetectable, the peak intensity of the other methylene proton is exceptionally improved, and the signal overlapping between the two methylene protons is eliminated. As a result, the NOEs involving the methylene proton can be observed. Notably, since the methylene proton is stereo-assigned, the NOE-derived restraints contribute to defining the

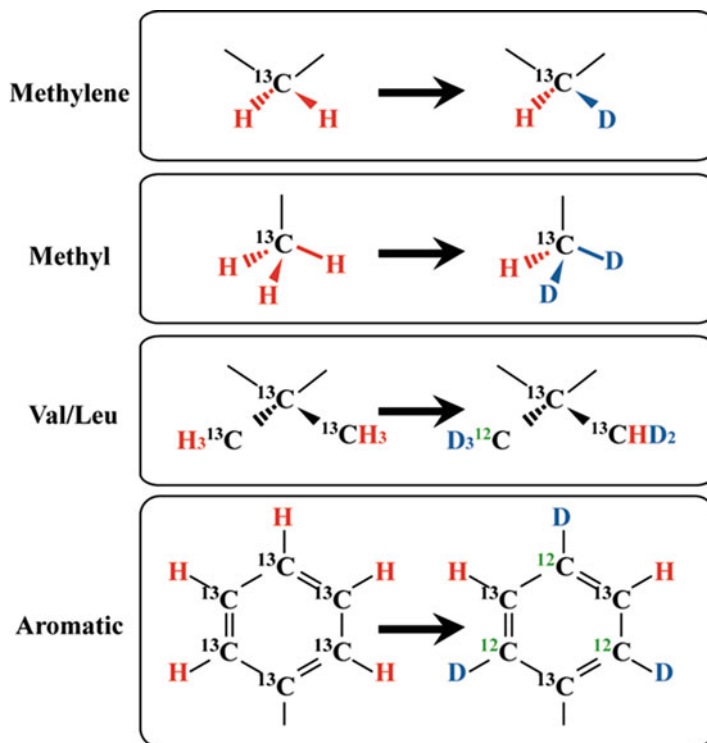


Fig. 2 SAIL amino acids. Design concepts in the SAIL amino acids

structure of the protein. Besides the methylene groups, such stereo- and regio-specific labeling are applied for the *prochiral* methyl groups of Leu and Val residues. Furthermore, the methyl groups are kept as $^{13}\text{CHD}_2$, such that the proton density of the proton is reduced, and thus more quantitative NOEs are obtainable. In the aromatic rings, specific C-H moieties are kept as $^{13}\text{C}-^1\text{H}$, and the adjacent C-H moieties are $^{12}\text{C}-^2\text{H}$ (Fig. 2).

The technical foundation for producing SAIL proteins was established during the NMR structure determinations of 18 kDa calmodulin and 42 kDa maltose-binding protein (MBP). We devised chemical synthesis methods for all of the 20 proteinaceous amino acids with stereo- and regio-specific ^2H -, ^{13}C - and ^{15}N -isotope labeling patterns [29, 31–33]. The SAIL amino acids were then incorporated into the target proteins, using an *E. coli* cell-free protein production system [34, 35]. As compared to the conventional cellular expression system, the cross-labeling of amino acids remains minimal in the cell-free production system. In addition, the incorporation rate of the amino acids is much higher, as compared to the *in vivo* system.

The use of SAIL proteins drastically improves the quality of the NMR spectra, especially in the ^1H - ^{13}C correlation spectra of side-chain aliphatic and aromatic

regions. The line width of each resonance in the SAIL protein is much narrower than that in a conventional uniformly $^{13}\text{C}/^{15}\text{N}$ -labeled (UL) protein, and the signal overlap problem is significantly mitigated. In the case of an aromatic moiety, a constant time scheme is not needed due to the absence of one-bond ^{13}C - ^{13}C coupling in the ring. Since fewer protons serve as the source of NOEs, as compared to UL proteins, the spin diffusion is highly suppressed, enabling the acquisition of more quantitative and sensitive NOEs [36].

The use of SAIL proteins offers an advantage in that the NMR signals of aromatic moieties, which are difficult to observe in a UL protein, can be detected with high quality, thus enabling the observation of NOEs involving the aromatic atoms. In the aromatic rings of SAIL amino acids, a CH moiety is labeled as ^{13}C - ^1H , and the other protons and carbons are kept as ^{12}C and ^2H . To examine whether the position of the ^{13}C - ^1H moieties in the phenyl ring influences the NMR analysis, we synthesized three types of SAIL Phe and Tyr residues to observe the aromatic signals at the δ , ϵ , and ζ positions. The δ -, ϵ -, and ζ -SAIL Phe residues were incorporated into an 18.2 kDa protein, *E. coli* peptidyl prolyl *cis-trans* isomerase b (EPP1b), and the aromatic CH signals were observed. As compared to UL proteins, each peak was clearly detected by virtue of the elimination of the scalar and dipolar couplings. Twelve signals were observed for the ζ -SAIL Phe-labeled EPP1b, in agreement with the number of Phe residues in the protein. Interestingly, however, only nine CH signals were detected for the δ - and ϵ -SAIL Phe-labeled proteins (Fig. 3). This phenomenon was ascribed to the exchange broadening derived from the ring flipping for the three missing CH signals, which will be described later in detail. The NOE constraints involving the δ , ϵ , and ζ protons of Phe/Tyr residues are complementary, and thus by combining the NOE restraints obtained using different SAIL proteins (i.e., δ -SAIL Phe/Tyr, ϵ -SAIL Phe/Tyr, and ζ -SAIL Tyr), the quality of the NMR structure can be further improved [37]. As with the Phe and Tyr residues, the spectral improvement in Trp residues is also enormous. In the SAIL Trp, the ^{13}C - ^1H moieties are located at the $\delta 1$, $\epsilon 3$, and $\eta 2$ positions. In terms of a protein structural study, an interesting point in the analysis of Trp residues is that the χ^2 conformation can be determined, based on the intra-residue NOEs involving $\text{H}_{\beta 3}$ and either the $\text{H}_{\delta 1}$ or $\text{H}_{\epsilon 3}$ proton [38].

Many NMR structures have been successfully determined using SAIL-labeled proteins [29, 39–41] (Fig. 4). The quality of the structures was improved by the collection of NOEs with aromatic amino acids, which contributed to the accuracy and convergence of the structures. Since the spectral quality is highly improved, by combining SAIL with the FLYA system, automatic assignment and structure determination will become more accessible [42, 43].

By using the SAIL method, the molecular weights of proteins amenable to NMR structure determination can be increased up to about 50 kDa. However, the molecular weights of functionally important proteins and complexes often exceed 50 kDa. Thus, further efforts are needed to increase the size range. As a prospect, it is important to further reduce the proton density of proteins via ^2H substitutions. The protons that offer key information are kept, and the other protons are extensively deuterated.

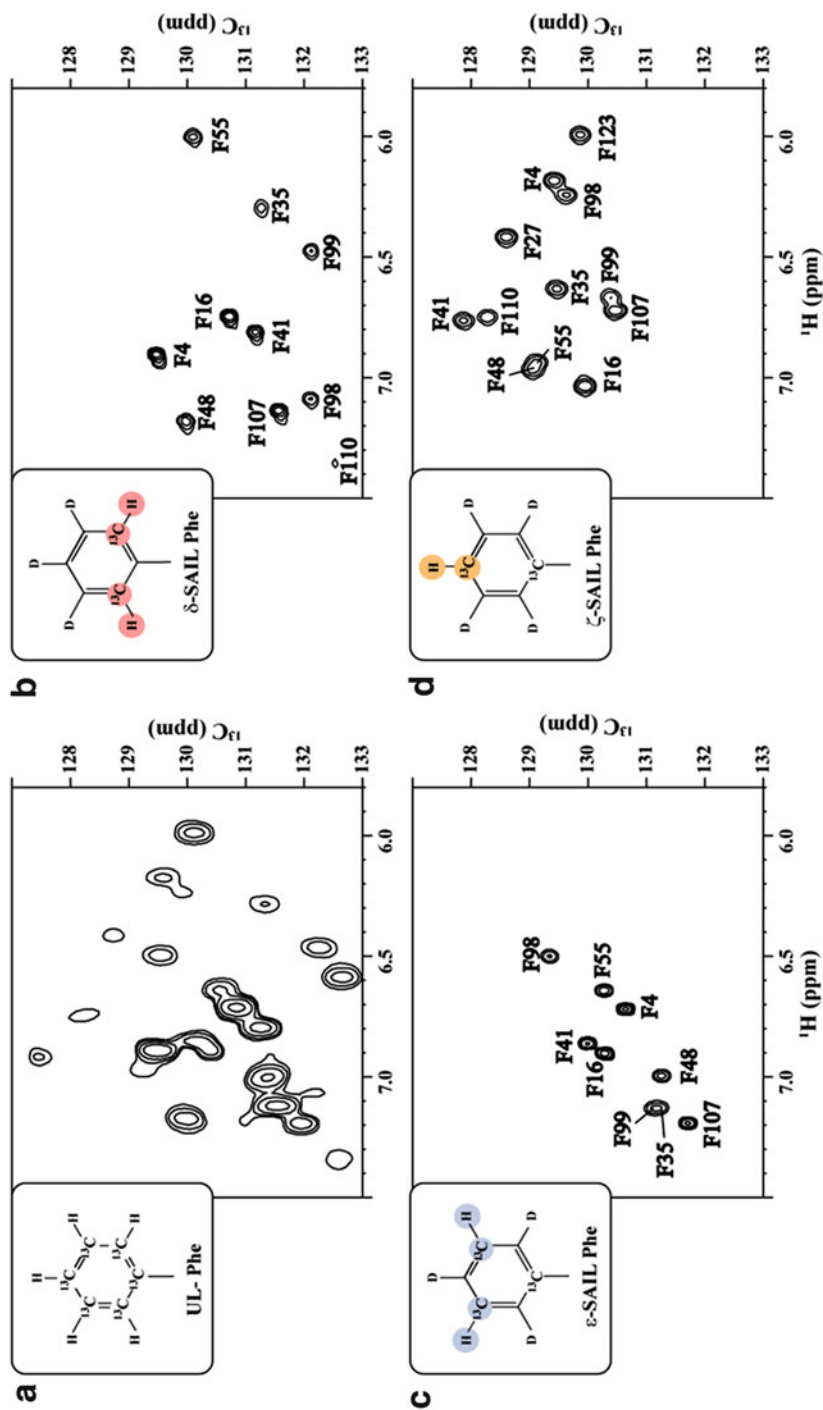


Fig. 3 NMR spectral comparison of the aromatic region of EPP1b specifically labeled with (a) uniformly ^{13}C -labeled Phe, (b) δ -SAIL Phe, (c) ϵ -SAIL Phe, and (d) ζ -SAIL Phe. The [^1H , ^{13}C]-HSQC spectra were measured at 30 $^\circ\text{C}$, using a DRX600 spectrometer (Bruker). Peaks are labeled with their assignments (Reproduced from Takeda et al. [37] with permission)

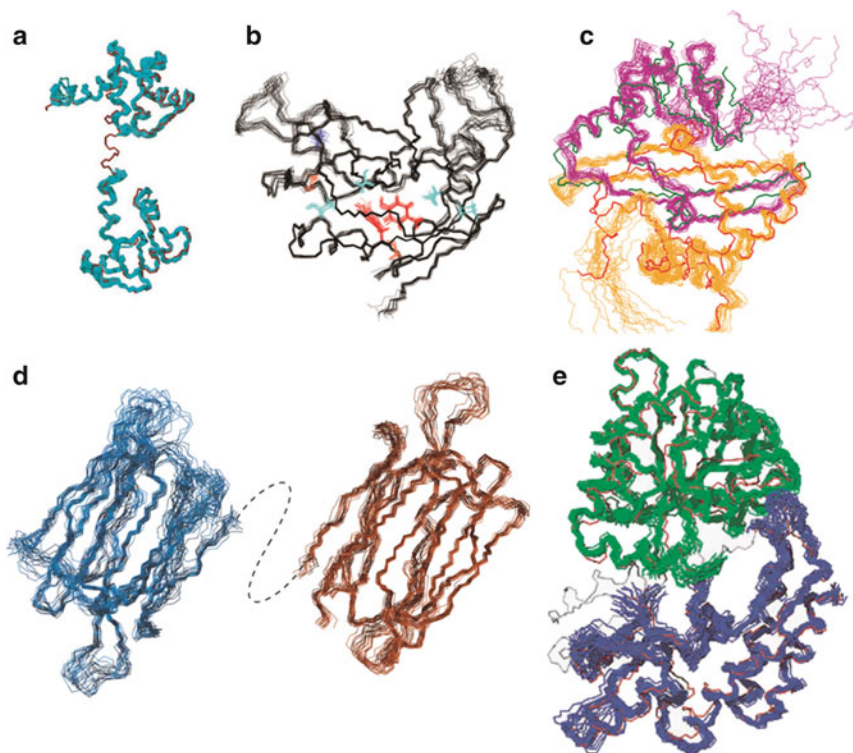


Fig. 4 NMR structures of SAIL proteins. (a) SAIL-CaM (backbone in cyan, Ca^{2+} in white), CaM X-ray structure (red), and three solution conformers of UL-CaM determined from residual dipolar coupling data (blue). (b) SAIL-EPP1b calculated by the combined use of all NOEs obtained for δ -, ϵ -, and ζ -SAIL Phe and δ - and ϵ -SAIL Tyr. (c) SAIL the C-terminal RNA-binding domain of SARS corona virus nucleocapsid protein. (d) SAIL-At3g16450.1 composed of the N-terminal domain (blue) and the C-terminal domain (red). (e) MBP solution and crystal structures. Backbone of the N-terminal domain (SAIL-MBP in green, X-ray structure in red) and the C-terminal domain (SAIL-MBP in blue, X-ray structure in red)

NMR Studies of Proteins Using Residue-Selective Labeling

Through the application of the SAIL method to large proteins, the detection of NMR signals has become possible by optimizing the isotope labeling pattern of the protein. Such site-specific isotope labeling is available for a variety of purposes besides structure determination, including investigations of the conformational dynamics and biomolecular interactions. If the amino acid to be labeled is unlikely to be scrambled in cells, then a robust *E. coli* cellular expression system can be used as the production method. In many cases, the expression level in cells is higher than that in cell-free synthesis, thus facilitating the application of the advanced isotope-labeling method to a wider range of proteins.

Large-Amplitude Slow Breathing Motion of Proteins Probed by the Ring Flipping Motions of Phenylalanine and Tyrosine Residues Embedded in the Hydrophobic Core

Folded proteins undergo a variety of conformational fluctuations on a wide range of time scales under physiological conditions, and the conformational fluctuations are assumed to be associated with the thermal stabilities and biological functions of the proteins. Among the conformational fluctuations, the infrequently occurring large-amplitude slow breathing motion (LASBM) represents a potentially important conformational fluctuation, which is accessible only by NMR. The LASBM was first detected through NMR observations of the rotations of the side-chain aromatic rings of phenylalanine (Phe) and tyrosine (Tyr) residues, embedded in the hydrophobic core of proteins, about their C β -C γ axis in solution. Since the embedded aromatic rings are tightly packed by the surrounding atoms of other residues in the crystal form, the explanation for the ring-flipping phenomenon is that proteins transiently undergo large-amplitude fluctuations involving the formation of a large space, which allows the ring-flipping motion to occur.

At the δ and ϵ positions of the aromatic rings of Phe and Tyr residues, a pair of CH moieties, equivalent with respect to the axis on the C β -C γ bond (e.g., $\delta 1$ vs. $\delta 2$), interconvert with each other due to the 180° ring-flipping motion. In this situation, if the ring-flipping motion is slower than the size of the chemical shift difference between the two equivalent positions, then the NMR signals of the two sites are resolved. However, if the flipping rate is faster than the chemical shift difference, then the signals are averaged. Based on the dependency of the line shape of the CH signals at the δ and ϵ positions on the ring-flipping rate, information about the frequency of the ring-flipping motion can be obtained.

The observation of NMR signals from aromatic atoms is generally difficult due to the complex spin system in the aromatic ring, with tight ^{13}C - ^{13}C coupling and ^1H - ^1H dipolar interactions between adjacent CH pairs. To overcome this problem, the site-specific isotope labeling pattern in the SAIL aromatic amino acids is quite useful. So far, we have designed and synthesized three types of phenylalanine and tyrosine residues, in which the ^1H - ^{13}C pairs at the δ , ϵ , or ζ positions have alternate isotope labeling patterns [31, 37]. Importantly, the CH bond at the ζ position is located on the rotational axis, and thus its line shape is insensitive to the ring-flipping rate. However, if another kind of conformational exchange is present, then the effect is manifested on the line shape of the CH signals at the ζ position. Therefore, through the combined analysis of the δ/ϵ and ζ positions, the exchange broadenings due to ring flipping and other conformational exchanges can be distinguished.

We have applied this method to study the protein-drug interactions of FKBP12-FK506 and FKBP12-rapamycin [44]. The interactions between FKBP12 and the immunosuppressant serve as a model for designing protein-ligand interactions. FK506 and rapamycin are both macrolides with the FKBP12-binding region on one side. The chemical structures of the FKBP12-binding sites

are identical in the two macrolides. When the two complexes are compared, the atomic coordinates of the molecular interfaces are indistinguishable between them. The binding interface includes several aromatic residues, and thus the ring-flipping motions of the interfacial aromatic rings were compared, based on the NMR line shapes and exchange experiments. In the free form, the δ - and ϵ -CH signals of Phe in FKBP12 are averaged. Upon binding to rapamycin, the signals of Tyr26 located in the binding interface were broadened, while these signals were resolved upon binding to FK506 (Fig. 5). The distinct line broadening indicates that the LASBMs occurring in the binding interfaces are different between the two complexes. The relationship between the different interface dynamics and the biological functions serves as an interesting future subject of study.

Hydrogen Exchange Study of Polar Side-Chain Groups of Proteins

Hydrogen atoms attached to polar groups, such as backbone amide groups and hydroxyl (OH), sulfhydryl (SH), and amino (NH₂) groups, exchange with the solvent water. The exchange frequency is closely associated with the environment at the position, such as the involvement in hydrogen bonding, the depth from the molecular surface, and the occurrence of conformational fluctuations allowing access to the solvent water. Therefore, the hydrogen exchange of polar

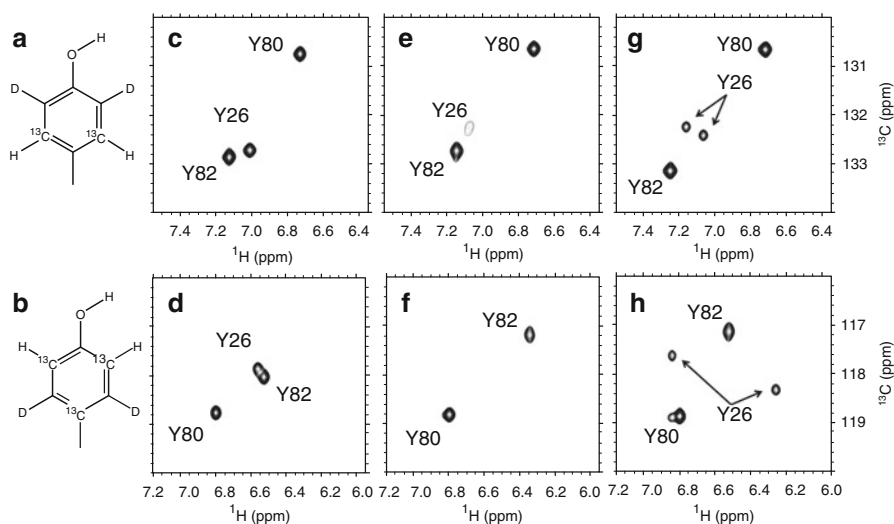


Fig. 5 The 2D ^1H - ^{13}C HSQC spectra of the aromatic region recorded on (a) δ - or (b) ϵ -SAIL Tyr-labeled FKBP12 in the apo (c, d), rapamycin-bound (e, f), and FK506-bound (g, h) states at 30 °C. Separated signals of Tyr26 were observed due to the slower ring-flipping motion upon FK506 binding (Reproduced from Yang et al. [44] with permission)

groups provides detailed information about the structures and dynamics of proteins, and numerous NMR hydrogen exchange studies have been reported. However, almost all of them focused on the non-exchangeable protons of amino acid residues, and there are very few NMR studies of polar side-chain groups. One reason for this is that the protons attached to polar groups exchange quickly with the water protons, and thus they are assumed to be rarely detectable by NMR. To expand the range of NMR hydrogen exchange studies to polar side-chain groups, we have developed NMR methods to investigate their hydrogen exchange rates in proteins by observing the line shapes of the ^{13}C -NMR signals of carbons attached to the polar groups, in a mixture of H_2O and D_2O [41, 45, 46, 47]. The ^{13}C chemical shifts of carbon atoms attached to protonated and deuterated polar groups are slightly, but significantly, different due to the H/D isotope shift effect. Given a hydrogen exchange rate slower than the size of the isotope shift, the ^{13}C signals attached to protonated and deuterated polar groups are resolved in the $\text{H}_2\text{O}/\text{D}_2\text{O}$ mixture; otherwise, the ^{13}C signals are observed as an averaged single peak.

The key requirement for this method is that the line widths of the ^{13}C signals from the carbon attached the polar groups are narrowed sufficiently to resolve the protonated and deuterated states. For this purpose, we synthesized amino acids in which the target carbon is site-specifically enriched by ^{13}C carbon and the protons, if present, are deuterated. For investigating Tyr OH groups, for example, we synthesized the ζ -SAIL Tyr. In this labeling pattern, the ζ -carbon attached to the OH group is enriched by ^{13}C , and the other aromatic carbons were kept as ^{12}C . To assign the ζ -carbon, the $\delta 1$ and $\delta 2$ protons were kept as ^1H , while the $\epsilon 1$ and $\epsilon 2$ were ^2H , such that the ζ -carbon is correlated with the δ -protons via three-bond J coupling, and then to the $\beta 3$ proton via the NOE. This ζ -SAIL Tyr was incorporated into the 18.2 kDa *E. coli* peptidyl-prolyl *cis-trans* isomerase b (EPPIb), and the Tyr ζ -carbon signals were observed in H_2O , D_2O and a 1:1 mixture. EPPIb contains three Tyr residues, Tyr30, Tyr36, and Tyr120. Interestingly, the signals of Tyr30 and Tyr36 were observed as a doublet, comprising the protonated and deuterated isotopomers, while that of Tyr120 was an averaged peak, indicating that the hydrogen exchange rates of the OH groups in Tyr30 and Ty36 were slower than the isotope shift ($\sim 10 \text{ s}^{-1}$) and that in Tyr120 was faster than the size of the line width (Fig. 6). Notably, the slowly exchanging OH protons identified through the ^{13}C observations in H_2O and D_2O are detectable by ^1H NMR. Such slowly exchanging and thus detectable ^1H s serve as sources of NOEs for determining the conformations of polar side-chain groups.

One important finding obtained through applications to protein NMR studies is that the proportion of slowly exchanging polar side-chain groups on the isotope shift time scales ($\sim 10 \text{ s}^{-1}$) may be larger than the generally assumed proportion. Interestingly, many polar side-chain groups were identified as slowly exchanging on the isotope-shift time scales throughout the analysis. (Ser: 1 of 6; Thr, 4 of 12; Cys, 2 of 2; Tyr, 2 of 3), and the numbers are larger than the general assumptions [41]. This method is also applicable to studies of biomolecular interactions [44].

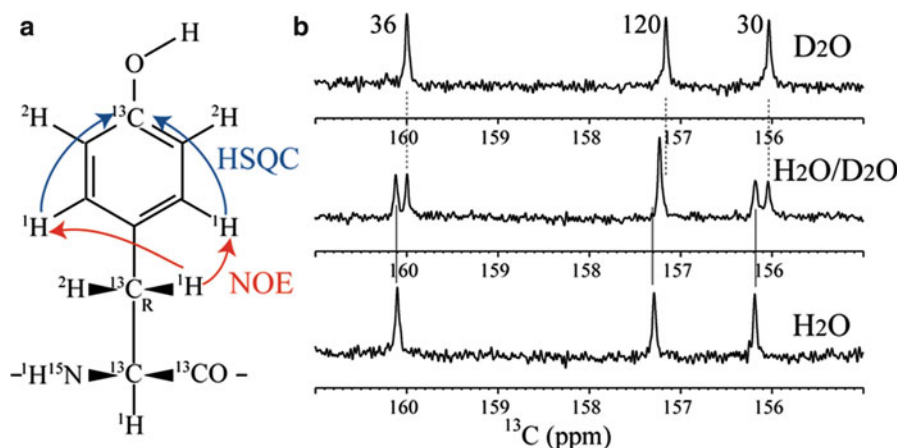


Fig. 6 (a) Structure of ζ -SAIL Tyr and the magnetization pathway related to the assignment of the $^{13}\text{C}^\zeta$ atom. (b) Carbon spectra of EPP1b selectively labeled by ζ -SAIL Tyr in 100% H_2O , 50% $\text{H}_2\text{O}/50\%$ D_2O , and 100% D_2O solutions. These spectra were acquired using a Bruker DRX600 spectrometer equipped with a TCI cryogenic probe at 40 °C (Reproduced from Takeda et al. [45] with permission)

NMR Studies of Conformations and Isomerizations of Protein Disulfide Bonds Using Proteins Selectively Labeled with Site-Specifically ^{13}C -Labeled Cysteine

Disulfide bonds in a protein adopt a specific conformation and in some cases exist in equilibrium between different conformations. Such conformational and dynamic properties of protein disulfide bonds represent an important subject for NMR studies. We initially developed a method for determining the connectivity and conformation of a protein disulfide bond by observing the quantitative NOEs between the β -protons in the two Cys residues across the disulfide bond [48]. The problem encountered in observing the quantitative NOEs is that fast spin diffusion between the *geminal* β -protons occurs in conventional $[\text{U-}^{13}\text{C}, ^{15}\text{N}]$ proteins. To overcome this issue, we prepared proteins selectively labeled with an equimolar mixture of $(2R,3S)$ - $[\beta\text{-}^{13}\text{C};\alpha,\beta\text{-}^2\text{H}_2]$ Cys and $(2R,3R)$ - $[\beta\text{-}^{13}\text{C};\alpha,\beta\text{-}^2\text{H}_2]$ Cys, but otherwise fully deuterated, to observe the quantitative NOEs. In the labeling pattern, since either one of the *prochiral* methylene protons, namely, β_2 (*proS*) or β_3 (*proR*), is always replaced with a deuteron and no other protons remain in proteins prepared by this labeling scheme, all four of the expected NOEs for the β -protons across disulfide bonds could be measured without any spin diffusion interference, even with long mixing times (Fig. 7). With the NOE information, accurate distances were evaluated for the four β -proton pairs, and thus the conformation could be determined. This method was applied to the three disulfide bonds in bovine pancreatic trypsin inhibitor (BPTI), demonstrating the feasibility of this method [48].

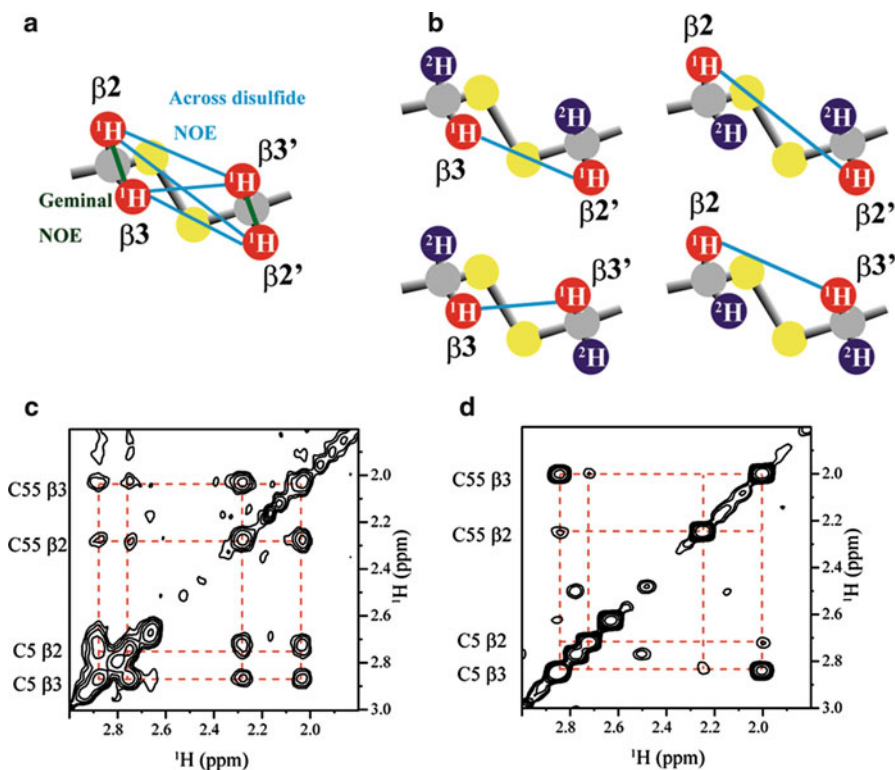


Fig. 7 The observation of across-disulfide NOEs. (a) In a disulfide bond, dipolar interactions between the geminal protons cause problematic spin diffusion and hamper the quantitative analysis of the four across-disulfide NOEs. (b) By incorporating equimolar amounts of (2*R*,3*S*)- and (2*R*,3*R*)-[β-¹³C;α,β-²H₂] Cys into a target protein, quantitative across-disulfide NOEs are observed in the complete absence of the geminal dipolar interactions. (c) In conventional UL proteins, the geminal NOEs are observed, and spin diffusion hampers the quantitative analysis of the across-disulfide NOEs. (d) In the ((2*R*, 3*RS*)-[3-¹³C;2,3-²H₂] Cys, [U-²H]-labeled protein, the geminal NOEs are eliminated, and quantitative NOEs are observed (Reproduced from Takeda et al. [48] with permission)

However, the conformation of a disulfide bond is not necessarily inert, and in some cases it exists in dynamic equilibrium between different conformations. We probed the conformational isomerization of the Cys14-Cys38 disulfide bond in bovine pancreatic trypsin inhibitor (BPTI), by an NMR line-shape analysis of its Cys carbon peaks [49]. In this case, the targeted carbons are site-specifically enriched by ¹³C, and the proton attached to the carbon is deuterated. The 1D ¹³C spectra were then recorded at small temperature intervals for the BPTI samples, and the recorded peaks were displayed in the order of the temperature. As the result, the exchange broadening that became altered with temperature was manifested for the carbon peaks of Cys14 and Cys38 over the profile of the line shape. Interestingly, biphasic exchange broadening was observed for the Cys14 α-carbon peak, which is

consistent with the report that the Cys14-Cys38 disulfide bond exists in equilibrium between a high-populated state (M) and two low-populated states (m_{c14} and m_{c38}). This line-shape analysis is useful for detecting and characterizing the conformational isomerization of protein disulfide bonds.

Concluding Remarks

The labeling of proteins with stable isotopes enhances NMR methods for analyses of structures, dynamics, and interactions. In our efforts to develop SAIL and related methods, we have demonstrated the feasibility and utility of stereo- and regio- specific isotope labeling for various purposes. With new technological advancements, we believe that the isotope-aided NMR studies of proteins will start from the design of the isotope labeling pattern of the target protein for the intended purposes, such that the spin relaxation occurring in the protein is optimally controlled.

References

1. Ohki S, Kainosho M. Recent developments in stable-isotope-aided methods for protein NMR spectroscopy. In: Modern magnetic resonance. The Netherlands: Springer; 2006. p. 211–8.
2. Ohki S, Kainosho M. Stable isotope labeling methods for protein NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc.* 2008;53:208–26.
3. GCK R. NMR of macromolecules: a practical approach. Oxford: Oxford University Press; 1993.
4. DM LM. Isotope labeling in solution protein assignment and structural analysis. *Prog Nucl Magn Reson Spectrosc.* 1994;26:371–419.
5. Kainosho M. Isotope labelling of macromolecules for structure determinations. *Nat Struct Biol.* 1997;4:854–7.
6. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry.* 1990;29:4659–67.
7. Clore GM, Gronenborn AM. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* 1994;239:349–63.
8. Markley JL, Putter I, Jardetzky O. High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science.* 1968;161:1249–51.
9. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A.* 1997;94:12366–71.
10. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc.* 2003;125:10420–8.
11. Ruschak AM, Kay LE. Proteasome allostery as a population shift between interchanging conformers. *Proc Natl Acad Sci U S A.* 2012;109(50):E3454–62.
12. Gardner KH, Kay LE. Production and incorporation of ^{15}N , ^{13}C , ^2H (^1H -61 methyl) isoleucine into proteins for multidimensional NMR studies. *J Am Chem Soc.* 1997;119:7599–600.
13. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR.* 1999;13(4):369–74.

14. Tugarinov V, Kay LE. An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR*. 2004;28:165–72.
15. Kay LE, Muhandiram DR, Farrow NA, Aubin J, Forman-Kay JD. Correlation between dynamics and high affinity binding in an SH2 domain interaction. *Biochemistry*. 1996;35:361–8.
16. Ishima R, Louis JM, Torchia DA. Optimized labeling of ^{13}C HD₂ methyl isotopomers in perdeuterated proteins: Potential advantages for ^{13}C relaxation studies of methyl dynamics of larger proteins. *J Biomol NMR*. 2001;21:167–71.
17. Gardner KH, Konrat R, Rosen MK, Kay LE. An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated ^{15}N , ^{13}C -labeled proteins. *J Biomol NMR*. 1996;8:351–6.
18. Tugarinov V, Kay LE. Ile, Leu, and Val methyl assignments of the 723-Residue Malate Synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc*. 2003;125:13868–78.
19. Amero C, Asunción DM, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, Vernet T, Franzetti B, Boisbouvier J. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J Biomol NMR*. 2011;50:229–36.
20. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional carbon-13 labeling. *Biochemistry*. 1989;28:7510–6.
21. Ruschak AM, Velyvis A, Kay LE. A simple strategy for ^{13}C , ^1H labeling at the Ile- γ 2 methyl position in highly deuterated proteins. *J Biomol NMR*. 2010;48:165–72.
22. Lichtenecker RJ, Weinhäupl K, Reuther L, Schörghuber J, Schmid W, Konrat R. Independent valine and leucine isotope labeling in *Escherichia coli* protein overexpression systems. *J Biomol NMR*. 2013;57:205–9.
23. Gans P, Hamelin O, Sounier R, Ayala I, Durá MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Eng*. 2010;49(11):1958–62.
24. Ostler G, Soteriou A, Moody CM, Khan JA, Birdsall B, Carr MD, Young DW, Feeney J. Stereospecific assignments of the leucine methyl resonances in the ^1H NMR spectrum of *Lactobacillus casei* dihydrofolate reductase. *FEBS Lett*. 1993;318:177–80.
25. Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J. Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J Biomol NMR*. 2013;57(3):251–62.
26. Miyanoiri Y, Takeda M, Okuma K, Ono AM, Terauchi T, Kainosho M. Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereospecifically methyl labeled amino acids. *J Biomol NMR*. 2013;57:237–49.
27. Monneau YR, Ishida Y, Rossi P, Saio T, Tzeng SR, Inouye M, Kalodimos CG. Exploiting *E. coli* auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *J Biomol NMR*. 2016;65(2):99–108.
28. Miyanoiri Y, Ishida Y, Takeda M, Terauchi T, Inouye M, Kainosho M. Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic *E. coli* strain. *J Biomol NMR*. 2016;65:109–19.
29. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P. Optimal isotope labelling for NMR protein structure determinations. *Nature*. 2006;440:52–7.
30. Kainosho M, Güntert P. SAIL—stereo-array isotope labeling. *Q Rev Biophys*. 2009;42:247–300.
31. Torizawa T, Ono AM, Terauchi T, Kainosho M. NMR assignment methods for the aromatic ring resonances of Phenylalanine and Tyrosine residues in proteins. *J Am Chem Soc*. 2005;127:12620–6.
32. Terauchi T, Kobayashi K, Okuma K, Oba M, Nishiyama K, Kainosho M. Stereoselective synthesis of triply isotope-labeled Ser, Cys, and Ala: amino acids for stereoarray isotope labeling technology. *Org Lett*. 2008;10:2785–7.

33. Okuma K, Ono AM, Tsuchiya S, Oba M, Nishiyama K, Kainosho M, Terauchi T. Asymmetric synthesis of (2*S*,3*R*)- and (2*S*,3*S*)-[2-¹³C;3-²H] glutamic acid. *Tetrahedron Lett.* 2009;50:1482–4.
34. Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T, Yokoyama S. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* 1999;442:15–9.
35. Torizawa T, Shimizu M, Taoka M, Miyano H, Kainosho M. Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. *J Biomol NMR.* 2004;30:311–25.
36. Takeda M, Ikeya T, Güntert P, Kainosho M. Automated structure determination of proteins with the SAIL-FLYA NMR method. *Nat Protoc.* 2007;2:2896–902.
37. Takeda M, Ono AM, Terauchi T, Kainosho M. Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. *J Biomol NMR.* 2010;46:45–9.
38. Miyanoiri Y, Takeda M, Jee J, Ono AM, Okuma K, Terauchi T, et al. Alternative SAIL-Trp for robust aromatic signal assignment and determination of the $\chi(2)$ conformation by intra-residue NOEs. *J Biomol NMR.* 2011;51:425–35.
39. Takeda M, Chang CK, Ikeya T, Güntert P, Chang YH, Hsu YL, et al. Solution structure of the C-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol.* 2008;380:608–22.
40. Takeda M, Sugimori N, Torizawa T, Terauchi T, Ono AM, Yagi H, et al. Structure of the putative 32 kDa myrosinase binding protein from Arabidopsis (At3g16450.1) determined by SAIL-NMR. *FEBS J.* 2008;275:5873–84.
41. Takeda M, Jee J, Ono AM, Terauchi T, Kainosho M. Hydrogen exchange study on the hydroxyl groups of Serine and Threonine residues in proteins and structure refinement using NOE restraints with polar side-chain groups. *J Am Chem Soc.* 2011;133:17420–7.
42. Ikeya T, Takeda M, Yoshida H, Terauchi T, Jee JG, Kainosho M, Güntert P. Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system. *J Biomol NMR.* 2009;44:261–72.
43. Schmidt E, Ikeya T, Takeda M, Löhr F, Buchner L, Ito Y, Kainosho M, Güntert P. Automated resonance assignment of the 21 kDa stereo-array isotope labeled thiol-disulfide oxidoreductase DsbA. *J Magn Reson.* 2014;249:88–93.
44. Yang CJ, Takeda M, Terauchi T, Jee J, Kainosho M. Differential large-amplitude breathing motions in the interface of FKBP12-Drug complexes. *Biochemistry.* 2015;54:6983–95.
45. Takeda M, Jee J, Ono AM, Terauchi T, Kainosho M. Hydrogen exchange rate of tyrosine hydroxyl groups in proteins as studied by the deuterium isotope effect on C ζ chemical shifts. *J Am Chem Soc.* 2009;131:18556–62.
46. Takeda M, Jee J, Terauchi T, Kainosho M. Detection of the sulfhydryl groups in proteins with slow hydrogen exchange rates and determination of their proton/deuteron fractionation factors using the deuterium-induced effects on the ¹³C β NMR signals. *J Am Chem Soc.* 2010;132:6254–60.
47. Takeda M, Miyanoiri Y, Terauchi T, Yang CJ, Kainosho M. Use of H/D isotope effects to gather information about hydrogen bonding and hydrogen exchange rates. *J Magn Reson.* 2014;241:148–54.
48. Takeda M, Terauchi T, Kainosho M. Conformational analysis by quantitative NOE measurements of the β -proton pairs across individual disulfide bonds in proteins. *J Biomol NMR.* 2012;52:127–39.
49. Takeda M, Miyanoiri Y, Terauchi T, Kainosho M. ¹³C-NMR studies on disulfide bond isomerization in bovine pancreatic trypsin inhibitor (BPTI). *J Biomol NMR.* 2016;66:37–53.