J Biomol NMR (2010) 46:45–49
DOI 10.1007/s10858-009-9360-9

ARTICLE

Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination

Mitsuhiro Takeda · Akira M. Ono · Tsutomu Terauchi · Masatsune Kainosho

Received: 15 June 2009 / Accepted: 7 July 2009 / Published online: 7 August 2009

Abstract The extensive collection of NOE constraint data involving the aromatic ring signals is essential for accurate protein structure determination, although it is often hampered in practice by the pervasive signal overlapping and tight spin couplings for aromatic rings. We have prepared various types of stereo-array isotope labeled phenylalanines (ε- and ζ-SAIL Phe) and tyrosine (ε-SAIL Tyr) to overcome these problems (Torizawa et al. 2005), and proven that these SAIL amino acids provide dramatic spectral simplification and sensitivity enhancement for the aromatic ring NMR signals. In addition to these SAIL aromatic amino acids, we recently synthesized δ-SAIL Phe and δ-SAIL Tyr, which allow us to observe and assign δ-^{13}C/^{1}H signals very efficiently. Each of the various types of SAIL Phe and SAIL Tyr yields well-resolved resonances for the δ-, ε- or ζ-^{13}C/^{1}H signals, respectively, which can readily be assigned by simple and robust pulse sequences. Since the δ-, ε-, and ζ-proton signals of Phe/Tyr residues give rise to complementary NOE constraints, the concomitant use of various types of SAIL-Phe and SAIL-Tyr would generate more accurate protein structures, as compared to those obtained by using conventional uniformly ^{13}C, ^{15}N-double labeled proteins. We illustrated this with the case of an 18.2 kDa protein, *Escherichia coli* peptidyl-prolyl *cis-trans* isomerase b (EPPIb), and concluded that the combined use of ζ-SAIL Phe and ε-SAIL Tyr would be practically the best choice for protein structural determinations.

Keywords SAIL-NMR method · SAIL tyrosine · SAIL phenylalanine · EPPIb structure

M. Takeda · M. Kainosho
Structural Biology Research Center, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

A. M. Ono · T. Terauchi
SAIL Technologies Co., Inc., 1-40, 1 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

A. M. Ono · T. Terauchi · M. Kainosho (✉)
Center of Priority Areas, Graduate School of Science and Technology, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji 192-0397, Japan
e-mail: kainosho@nmr.chem.metro-u.ac.jp

Introduction

We have recently described a new isotope-assisted NMR approach, designated as the stereo-array isotope labeling (SAIL) method, which utilizes isotope-labeled protein samples optimally designed for NMR spectroscopy (Kainosho et al. 2006). In doing so, various stereo- and regio-specifically ^{2}H, ^{13}C, ^{15}N-triple labeled amino acids, SAIL amino acids, have been synthesized by chemical and enzymatic methods (for example, see Terauchi et al. 2008; Okuma et al. 2009). SAIL amino acids are incorporated most efficiently into the target proteins, without losing or scrambling the isotopes of labeled amino acids, by the *Escherichia coli* cell-free protein expression system (Torizawa et al. 2004; Takeda et al. 2007). SAIL proteins, which are exclusively composed of SAIL amino acids, generated enormously simplified NMR spectra with much higher sensitivity, as compared to those obtained for conventional, uniformly ^{13}C, ^{15}N-double labeled proteins. The SAIL proteins thus allowed us to determine the structures of proteins as large as 40 kDa very accurately (Kainosho et al. 2006; Takeda et al. 2008a, b). One of the remarkable advantages of the SAIL proteins can be seen in the dramatic improvement of the spectral quality for aromatic ring resonances, which are often very difficult to analyze for

uniformly double-labeled proteins, due to extensive signal overlapping and tight spin-couplings in this spectral region (Torizawa et al. 2005; Takeda et al. 2008a). Since the aromatic rings tend to exist in the hydrophobic cores of globular proteins, the NOEs related to aromatic rings are extremely important distance constraints to determine high-resolution protein structures. In this paper, we will present the comparative use of various SAIL Phe and Tyr, including new members, δ-SAIL Phe/Tyr, to evaluate the δ-, ε-, and ζ-proton derived NOE distance constraints on the qualities of the determined protein structures.

SAIL Phe and Tyr with different isotope-labeling patterns

It is often difficult to analyze the ring spectral regions of aromatic amino acids, especially for Phe and Tyr residues, in uniformly ^{13}C labeled proteins. One reason is that the chemical shifts of the ring carbons of Phe residues, namely the δ-, ε-, and ζ-carbons, and the δ-carbon of Tyr residues, have limited dispersions, and thus their resonances are prone to overlap each other in NMR spectra. Such limited chemical shift dispersions and large ^{13}C–^{13}C spin couplings result in tight spin couplings, which hamper the spectral analyses, especially for Phe residues. Slow ring flipping motions, which have occasionally been reported in the literature (Wagner et al. 1976; Campbell et al. 1976), make the NMR signals for the δ and ε atoms of some Phe and Tyr residues even more difficult, or sometimes impossible, to observe (Takeda et al. 2008a).

The isotope-labeling patterns of the aromatic rings in SAIL Phe and Tyr residues are designed to overcome all of these problems simultaneously. Figure 1 illustrates the repertories of the currently available SAIL Phe and Tyr. These include newly designed δ-SAIL Phe and δ-SAIL Tyr, in addition to the previously prepared ε-SAIL Phe, ε-SAIL Tyr, and ζ-SAIL Phe (Torizawa et al., 2005). The special characteristics of these SAIL amino acids reside in their alternative ^{13}C, ^{2}H-labeling patterns. The ring carbons of all of the SAIL Phe and Tyr residues lack directly bonded ^{13}C–^{13}C pairs to avoid strong one-bond ^{13}C–^{13}C scalar couplings, and the alternate ^{2}H-labeling eliminates large scalar and dipolar interactions among the ring protons. Therefore, the unambiguous assignments can readily be achieved by the straightforward NOE connectivity between H_δ and H_β, as shown in Fig. 1, and by the through ^{13}C–^{13}C coupling connectivities (Torizawa et al. 2005).

Assignment of the ring proton and carbon resonances of twelve Phe residues of EPPIb by SAIL Phe with different isotope-labeling patterns

Since the use of various SAIL Phe and Tyr residues dramatically simplifies the ^{1}H–^{13}C HSQC signals of the aromatic ring signals in proteins, the analysis of aromatic ring signals becomes quite straightforward, even for proteins containing many aromatic amino acids. Figure 2 illustrates the aromatic ring NMR regions of an 18.2 kDa protein, *E. coli* peptidyl-prolyl *cis-trans* isomerase b (EPPIb), which contains twelve Phe residues, specifically labeled with each one of the three types of SAIL Phe, and with uniformly ^{13}C-labeled Phe. These SAIL EPPIb samples were prepared according to the method described previously (Kariya et al. 2000; Torizawa et al. 2004; Takeda et al. 2007). The aromatic ring signals of the twelve Phe residues of EPPIb are extremely well-resolved for all SAIL Phe types, and their resonances were readily assigned by correlation to the previously assigned back-bone signals (Kariya et al. 2000), by the methods shown in Fig. 1 (Torizawa et al. 2005). Note that, for the δ- and ε-Phe labeled EPPIb's, the signals due to three out of the twelve Phe residues in the protein were severely broadened under the experimental

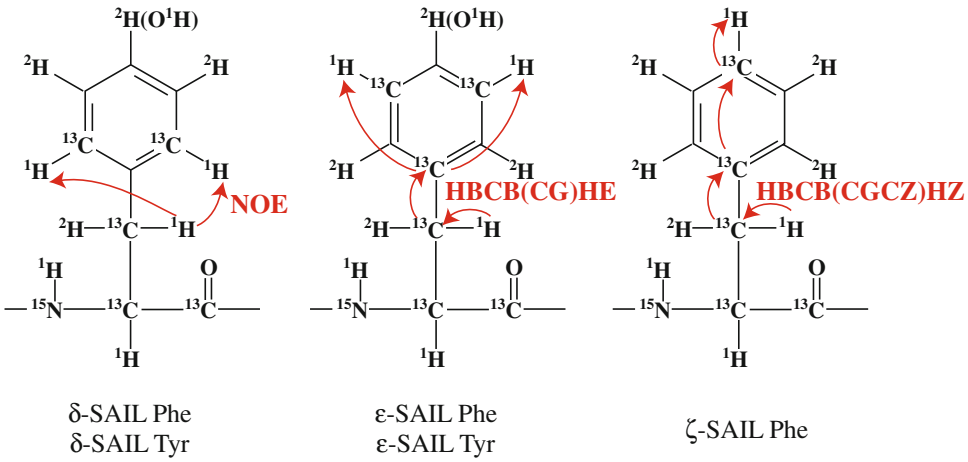


Fig. 1 Chemical structures of various SAIL phenylalanines and tyrosines. The magnetization transfer pathways and their acronyms, which can be used to assign the δ-, ε-, and ζ-signals, are shown by the *arrows*. (All of these SAIL aromatic amino acids are commercially available: http://www.sail-technologies.com)

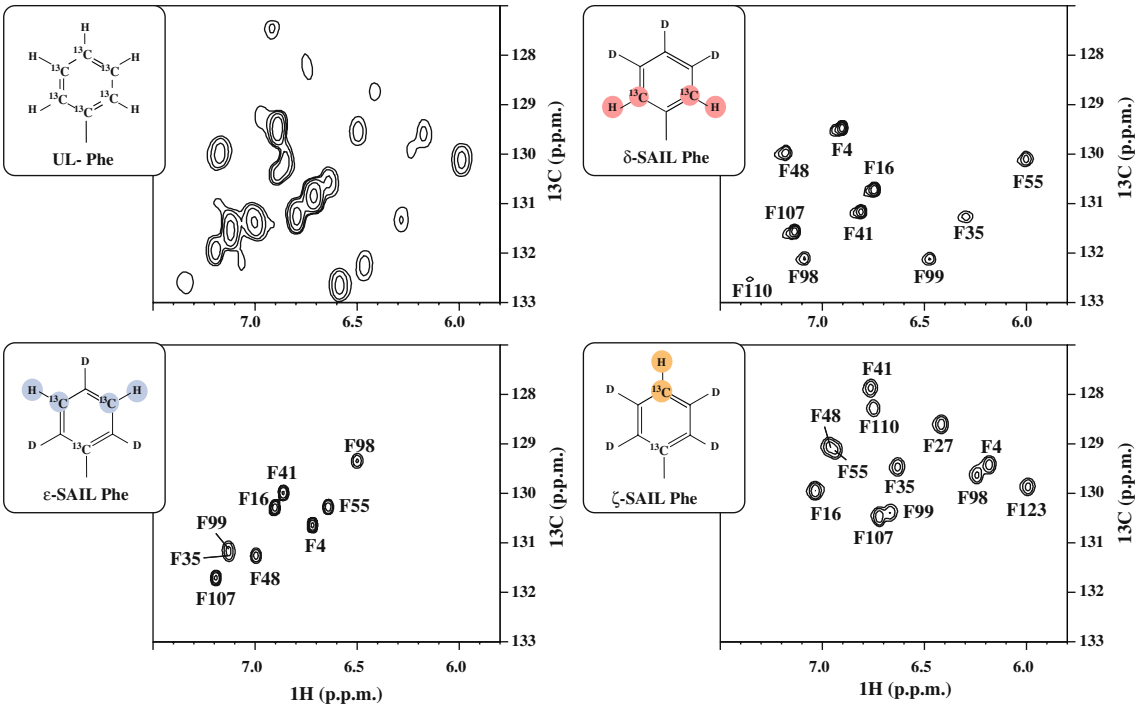


Fig. 2 NMR spectral comparisons of the aromatic regions of EPPIb's selectively labeled with the δ-, ε-, and ζ-SAIL phenylalanines, to that of the uniformly ^{13}C-labeled EPPIb. The HSQC spectra were measured at 30°C, using a DRX600 spectrometer (Bruker) equipped with a TXI cryoprobe. A 280 µL portion of a 0.1–0.2 mM solution of SAIL EPPIb, dissolved in 10 mM sodium phosphate buffer (pH 6.6), containing 100 mM NaCl and 1 mM DTT, in a Shigemi tube was used for NMR measurements. The time points were 1024 (t1) × 256 (t2). EPPIb specifically labeled with **a** uniformly ^{13}C-labeled Phe; **b** δ-SAIL Phe; **c** ε-SAIL Phe; **d** ζ-SAIL. For **a**, the ^{13}C chemical shift encoded in the indirect dimension was acquired in a constant-time manner, whereas it was acquired in a variable-time manner for (**b**), (**c**), and (**d**). The assignments given for each resonance were established by the previously reported pulse sequence (Torizawa et al. 2005) for (**c**) and (**d**)

conditions (30°C, 600 MHz). These residues were identified as Phe-27, -110, and -123, respectively. Their δ- and ε- signals were broadened beyond detection, due to the intermediate ring flipping rates of these three residues at 30°C, except for a single, very weak peak assigned to the δ-^{13}C/^{1}H signal of Phe-110, as indicated in Fig. 2. At an elevated temperature, such as 40°C, this weak δ-^{13}C/^{1}H cross peak became sharper, due to the increased ring flipping rate of Phe-110. The flipping-rate dependent effects were also seen for the other two missing Phe residues, using ε-SAIL Phe/Tyr EPPIb. For example, two very weak and broad signals appeared in the ^{1}H–^{13}C HSQC spectrum at 40°C, and they could not be assigned due to severe line broadening (data not shown). Unfortunately, due to the limited thermal stability of EPPIb, we could not perform measurements at higher temperatures. We also measured ε-SAIL Phe/Tyr and δ-SAIL Phe/Tyr EPPIb at lower temperatures and at 950 MHz. At 0°C, two sets of very weak signals started to appear. It was obvious that the flipping rates of the two Phe residues were too slow to give rise to different chemical shifts for the two δ- and ε -^{13}C–^{1}H pairs (data not shown). Although further analyses have not been performed, these preliminary experiments clearly indicate that the ring flipping rates of Phe-27, -110, and -123 are in the intermediate range, which tends to broaden the ring signals of these residues. On the contrary, all twelve peaks were clearly observed for the ζ-signals in ζ-SAIL Phe labeled EPPIb, since the ζ-^{13}C^{1}H bonds are located along the C_β–C_γ axis, and thus they are not affected by the flipping rates.

Comparative evaluation of various SAIL Phe and Tyr residues for protein structure determinations

With the various types of SAIL Phe/Tyr in hand, we are able to observe and assign all of the ^{13}C–^{1}H pair signals of the aromatic rings extremely efficiently, and thus a variety of applications of those signals are possible. We describe here a simple, yet very important, application among the numerous other possibilities; that is, the use of the assigned aromatic ring signals for protein structure determinations. It would be of practical importance to evaluate the versatile combinations of various SAIL Phe/Tyr for such a purpose. In so doing, we collected NOE data for the three SAIL EPPIb samples, which were labeled with three different

Table 1 Comparison between converged structures calculated from the NOE constraints obtained for SAIL EPPIb's with each of the three combinations of SAIL Phe and SAIL Tyr

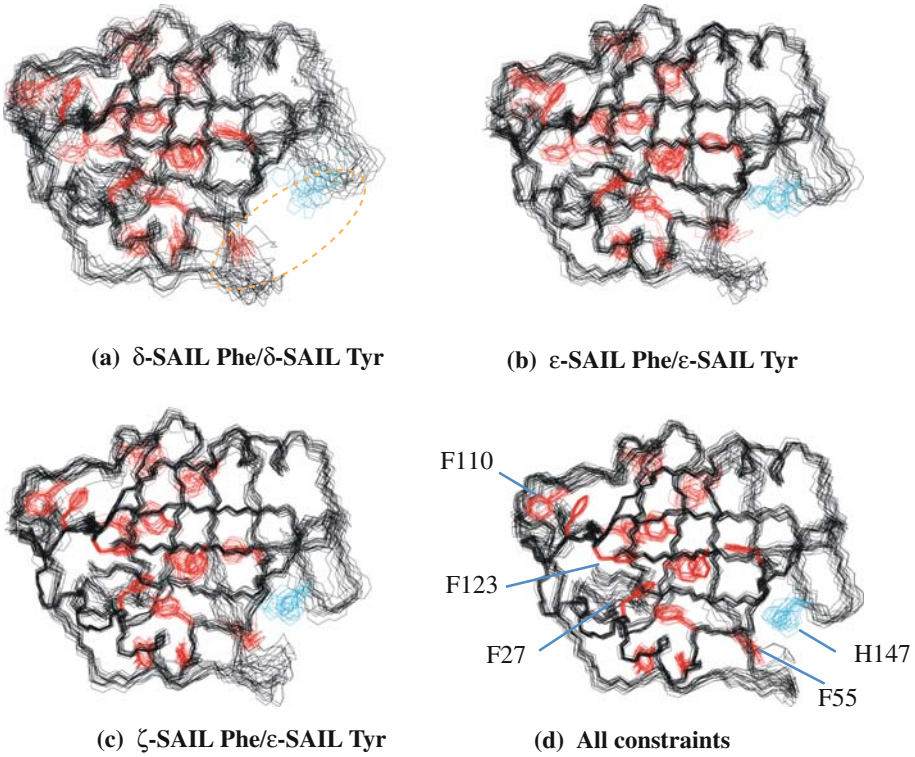
Combination of SAIL Phe and SAIL Tyr	NOEs obtained for each combination[b]				Backbone RMSD (Å)	All heavy atoms (Å)
	$\|i-j\|=0$	$\|i-j\|=1$	$1<\|i-j\|<5$	$\|i-j\|>6$		
δ-SAIL Phe, δ-SAIL Tyr	34 (9)	15 (3)	11 (2)	46 (8)	1.01	1.39
ε-SAIL Phe, ε-SAIL Tyr	3 (1)	6 (0)	11 (2)	60 (10)	0.79	1.17
ζ-SAIL Phe, ε-SAIL Tyr[a]	0	3	6	44	0.60	0.99
All NOE constraints	37 (10)	24 (3)	28 (4)	150 (18)	0.57	0.91

[a] ε-SAIL Tyr was used for this sample, while the number of NOEs shown here refers only to the ζ-Phe based NOEs

[b] Total numbers are shown for each SAIL Phe/Tyr combination, except for the ζ-SAIL Phe, ε-Tyr case, and the NOE numbers from the two Tyr residues are shown in the parentheses

combinations of SAIL Phe/Tyr, while the rest of the amino acids were fully replaced with an identical composition of SAIL amino acids in the three samples. Note that the spectra illustrated in Fig. 2 are those for the samples selectively labeled with each one of the three types of SAIL Phe and uniformly ^{13}C-labeled Phe. As shown in Table 1, the total numbers of the observed NOEs drastically differ for each of the three SAIL Phe/Tyr combinations. Namely, more than one hundred NOE cross peaks were observed for the δ-SAIL Phe/Tyr combination, but the number of NOEs for the ζ-SAIL Phe/ε-SAIL Tyr combination was about fifty. However, a closer look at the NOE statistics reveals that about half of the NOEs acquired for the δ-SAIL Phe/Tyr combination were intra-residue NOEs, which are not very useful distance constraints for structure determinations. On the contrary, almost the same numbers of NOEs among sequentially distant residues were observed for all three combinations (Table 1), and all of the EPPIb structures calculated by CYANA (Güntert 2003) using the three combinations converged well, as shown in Fig. 3. Note that, as mentioned above, the NOEs related to the δ- and ε-signals of Phe-27, -110, and -123 were almost completely missing, and thus the RMSD values for the structures determined for the δ-SAIL Phe/Tyr (Fig. 3a) and ε-SAIL Phe/Tyr (Fig. 3b) combinations are a little worse, as compared to that for the ζ-SAIL Phe/ε-SAIL Tyr (Fig. 3c) combination.

If one takes a closer look at the structures determined by these three combinations, it becomes evident that the convergence of the structures was considerably worse for



(a) δ-SAIL Phe/δ-SAIL Tyr **(b) ε-SAIL Phe/ε-SAIL Tyr**

(c) ζ-SAIL Phe/ε-SAIL Tyr **(d) All constraints**

Fig. 3 Structures of EPPIb determined by the SAIL method, using various combinations of SAIL Phe and Tyr. Twenty overlaid structures from the one hundred structures calculated with the program CYANA (ver. 2.2), using the standard CYANA-simulated annealing schedule with 20,000 torsion angle dynamics steps, for SAIL EPPIb's, labeled with **a** δ-SAIL Phe/Tyr; **b** ε-SAIL Phe/Tyr; and **c** ζ-SAIL Phe and ε-SAIL Tyr, with the rest of the amino acids fully labeled with SAIL amino acids. **d** The structures calculated by the combined use of all NOEs obtained for (**a**), (**b**) and (**c**). Phenylalanine and tyrosine residues are *colored red* and His-147 is *colored blue*

the δ-SAIL Phe/Tyr combination (Fig. 3a) as compared to the other structures, especially around the loop regions shown in the orange circle. This local structural difference is due to a complete lack of NOE constraints between the ring signals of Phe-55 and His-147, which were clearly observed for ε- and ζ-SAIL Phe, but not at all for δ-SAIL Phe. It is also apparent that the lack of NOEs for the δ- and ε-protons of Phe-27, -110, and -123, due to the intermediate ring flipping rates, is clearly reflected in the ill-defined ring positions of these three Phe residues, although the backbone conformation around these residues was defined by the other NOEs. The observation indicates that the structural information obtained for the aromatic ring protons is complimentary, and that accurate protein structures can be determined by the combined use of all aromatic ring-based NOEs. The structures shown in Fig. 3d, which were determined by the simultaneous use of all of the observed NOEs, overlaid well with the crystal structure, with an RMSD of 1.43 Å for the backbone excluding the flexible loop regions (Edwards et al. 1997). Since the structural quality obtained by the ζ-SAIL Phe/ε-SAIL Tyr combination is very similar to that generated by the combined NOE data using all three combinations, the ζ-SAIL Phe/ε-SAIL Tyr combination may be a good choice for structural determinations.

Perspectives of the applications of SAIL aromatic amino acids

We presented the application of various SAIL Phe/Tyr residues for collecting NOE distance constraints, and evaluated its utility in terms of structure determination. The ^{13}C–^{1}H pairs of aromatic ring moieties are considered to be amenable for strong TROSY effects, but in practice, the results reported so far were not quite satisfactory for uniformly ^{13}C-labeled, fully protonated proteins (Pervushin et al. 1998). In each type of SAIL Phe/Tyr, however, all of the ^{13}C and ^{1}H atoms are separated from each other by, at least, either one or two ^{12}C–^{2}H groups, and thus the TROSY effect works perfectly. Although it is not absolutely necessary to use TROSY sequences for small proteins like EPPIb, the line widths of the anti-TROSY peaks observed for the F1-coupled HSQC spectra are broader, as compared to those of the TROSY peaks (data not shown). We recently found that the TROSY components of SAIL Phe remain reasonably sharp, even for protein particles as large as 80 kDa. Therefore, the use of SAIL aromatic amino acids would open up new possibilities to study very large proteins (unpublished). The further optimization of the isotope labeling patterns of SAIL aromatic amino acids is currently in progress.

Acknowledgments This work was supported by Targeted Protein Research Program (MEXT) to M.K. and by a Grant-in-Aid for Young Scientist (B) (21770110) to M.T. We thank Dr. Frank Löhr and Prof. Volker Dötsch, Institute of Biophysical Chemistry and Center of Biological Magnetic Resonance, University of Frankfurt, for their kind help in collecting the NMR data at 950 MHz.

References

Campbell ID, Dobson CM, Moore GR, Perkins SJ, Williams RJP (1976) Temperature dependent molecular motion of a tyrosine residue of ferrocytochrome *C*. FEBS Lett 70:96–100

Edwards KJ, Ollis DL, Dixon NE (1997) Crystal structure of cytoplasmic *Eschericia coli* peptidyl-prolyl isomerase: evidence for decreased mobility of loops upon complexation. J Mol Biol 271:258–265

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Kariya E, Ohki S, Hayano T, Kainosho M (2000) Backbone ^{1}H, ^{13}C, and ^{15}N resonance assignments of an 18.2 kDa protein, *E. coli* peptidyl-prolyl *cis-trans* isomerase b (EPPIb). J Biomol NMR 18:75–76

Okuma K, Ono AM, Tsuchiya S, Oba M, Nishiyama K, Kainosho M, Terauchi T (2009) Asymmetric synthesis of (2S, 3R)- and (2S, 3S)-[2–13C;3–2H] glutamic acid. Tetrahedron Lett 50:1482–1484

Pervushin K, Riek R, Wider G, Wüthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labeled proteins. J Am Chem Soc 120:6394–6400

Takeda M, Ikeya T, Güntert P, Kainosho M (2007) Automated structure determination of proteins with the SAIL-FLYA NMR method. Nat Protoc 2:2896–2902

Takeda M, Chang CK, Ikeya T, Güntert P, Chang YH, Hsu YL, Huang TH, Kainosho M (2008a) Solution structure of the C-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J Mol Biol 380:608–622

Takeda M, Sugimori N, Torizawa T, Terauchi T, Ono AM, Yagi H, Yamaguchi Y, Kato K, Ikeya T, Jee J, Güntert P, Aceti DJ, Markley JL, Kainosho M (2008b) Structure of the putative 32 kDa myrosinase binding protein from Arabidopsis (At3g16450.1) determined by SAIL-NMR. FEBS J 275:5873–5884

Terauchi T, Kobayashi K, Okuma K, Oba M, Nishiyama K, Kainosho M (2008) Stereoselective synthesis of triply isotope-labeled Ser, Cys, and Ala: amino acids for stereoarray isotope labeling technology. Org Lett 10:2785–2787

Torizawa T, Shimizu M, Taoka M, Miyano H, Kainosho M (2004) Efficient production of isotopically labeled proteins by cell-free synthesis: A practical protocol. J Biomol NMR 30:311–325

Torizawa T, Ono AM, Terauchi T, Kainosho M (2005) NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. J Am Chem Soc 127:12620–12626

Wagner G, DeMarco A, Wüthrich K (1976) Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI) I ^{1}H NMR studies. Biophys Struct Mech 2:139–158