Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

1.12 Labeling Techniques

M Takeda, Nagoya University, Nagoya, Japan
M Kainosho, Nagoya University, Nagoya, Japan, and Tokyo Metropolitan University, Tokyo, Japan

Abbreviations

DEALS	deuterium–hydrogen exchange of amide on the line shapes
EPPIb	*Escherichia coli* peptidyl-prolyl *cis–trans* isomerase b
HSQC	heteronuclear single-quantum correlation spectroscopy
MBP	maltodextrin binding protein
NMR	nuclear magnetic resonance
NOESY	NOE spectroscopy
NP	nucleocapsid protein
SAIL	stereo-array isotope labeling
SARS-CoV	SARS coronavirus nucleocapsid protein
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TROSY	transverse relaxation-optimized spectroscopy

Glossary

Heteronuclear single-quantum correlation spectroscopy This experiment provides correlation peaks for two atoms that interact with each other by scalar coupling.

Nuclear Overhauser effect (NOE) This is a through-space interaction between nearby proton atoms. NOE peaks can be used for obtaining information on the spatial distance between the proton atoms.

Stereo-array isotope labeling A method that uses a protein exclusively composed of stereo- and regiospecifically labeled amino acids.

1.12.1 Introduction

Nuclear magnetic resonance (NMR) is a powerful tool for investigating the structures and dynamics of biological molecules. Biologically relevant macromolecules, such as proteins and nucleic acids, contain many protons and their molecular tumbling is slow, leading to the problems of signal overlap and line broadening. In this context, stable isotope labeling of target molecules is essential for investigating biological macromolecules. In the stable isotope labeling approach, the isotopic composition of specific atoms is altered, depending on the intended purpose. In general, the atoms for which the isotopic composition is commonly altered are proton, carbon, and nitrogen atoms. In addition, the use of other kinds of atoms, such as fluorine and oxygen isotopes, is sometimes considered.

Numerous isotope labeling methods are available for proteins and nucleic acids. The goal of isotope labeling can be categorized ultimately into the following two purposes:

- Enhancement of signal intensity for selected NMR peaks (positive labeling)
- Subtraction of unnecessary NMR peaks (negative labeling).

There are useful reviews available on the pre-existing stable isotope labeling methods.[1–6] Therefore, the authors' latest method, the stereo-array isotope labeling (SAIL) method,[7] is the focus of this chapter. The chapter first outlines the ^{2}H and $^{13}C/^{15}N$ labeling of proteins and then describes the SAIL method and its derivatives.

1.12.2 Basic Isotope Labeling Strategies

1.12.2.1 ^{2}H Labeling

Signal overlap and line broadening of NMR resonances are the most fundamental problems in the NMR analysis of large proteins. To address this issue, protein deuteration has long been utilized as a key method.[8–10] The resonance frequency of deuterium (^{2}H) is distinct from that of ^{1}H. Therefore, ^{2}H does not disturb ^{1}H NMR spectra. In addition, the magnetogyric ratio of ^{2}H is ~6.5 times lower than that of ^{1}H, and thus the substitution of ^{1}H to ^{2}H mitigates the dipolar and scalar couplings for the remaining ^{1}H atoms without altering the structure, dynamics, and function of the protein.

In random fractional ^{2}H labeling (the deuteration level ranges from ~50 to 90%, depending on the literature), the protons in protein samples are randomly replaced by deuterons. Thus, each protein molecule in the NMR sample tube has a different ^{2}H labeling pattern. The NMR signals from the residual protons in the random fractionally deuterated proteins yield useful structural information similar to that obtained from the fully protonated proteins, except that the NMR signals are narrower. The usefulness of random fractional deuteration has been demonstrated for a variety of proteins, such as thioredoxin and staphylococcal nuclease.[11,12] LeMaster and Richards achieved the chemical shift assignment for *Escherichia coli* thioredoxin using ~75% random fractional deuteration.[11] This 108-amino acid residue protein was previously considered to be too large for an analysis with ^{1}H–^{1}H homonuclear two-dimensional (2-D) NMR spectra. Their report clearly showed how ^{2}H labeling could alleviate the molecular size limitation for NMR studies. However, the chemical shifts in a deuterated sample are slightly different from those in a fully protonated sample due to deuterium-induced isotope shifts, and thus the residual proton signals were more complex, as described later.

Complete deuteration, termed 'perdeuteration,' further expands the size of proteins amenable to NMR analyses.[13] Although nonexchangeable protons in side chains are no longer observed in a perdeuterated protein, the amide protons yield well-sharpened lines in an H_2O solution due to the mitigated dipole interactions and spin coupling. When perdeuteration is combined with ^{15}N and/or ^{13}C labeling, which is described in the following section, even large proteins can be analyzed by NMR.

The deuterated proteins are commonly produced by growing *E. coli* cells on minimal medium (M9) containing deuterium oxide and either protonated or deuterated carbon sources.[14] Although *E. coli* cells can tolerate D_2O, culturing them in D_2O medium involves a significant reduction in both growth rate and yield due to the deuterium isotope effects. An adaptation of *E. coli* cells to deuterium oxide prior to the main culture is somewhat effective for overcoming this problem. When more deuteration is needed, a deuterated carbon source, such as [$^{2}H_7$] D-glucose, must be used to ensure a high deuteration level.

1.12.2.2 $^{13}C/^{15}N$ Labeling

The most important reason to label protein samples with ^{13}C and ^{15}N is to introduce the heteronuclear NMR dimensions into NMR spectroscopy, which facilitate the analysis of complex proton NMR spectra. The heteronuclear one-bond spin couplings between $^{13}C/^{15}N$ and directly attached ^{1}H are large enough to efficiently transfer nuclear magnetization through these spin couplings, which facilitates a variety of heteronuclear multidimensional experiments.[15] The availability of multidimensional experiments enormously expands the range of NMR study, and these provide a wealth of information on the protein. For example, the [^{1}H,^{15}N] heteronuclear single-quantum correlation spectroscopy (HSQC) spectrum from a uniformly ^{15}N-labeled protein can be used to evaluate the quality of the NMR sample. Therefore, an NMR study of a new protein often starts with the preparation of the uniformly ^{15}N-labeled protein. A good quality [^{1}H,^{15}N] HSQC spectrum promises advanced NMR analysis; otherwise, further optimization of the NMR sample should be performed in a trial-and-error manner. Once a preparation method for an NMR sample with good quality is established, the uniformly ^{13}C- and ^{15}N-labeled protein is prepared, with the aim of the sequential assignment of the backbone and, if required, side chain resonances. For the assignment of the NMR resonances in uniformly ^{13}C- and ^{15}N-labeled proteins, a series of heteronuclear multidimensional experiments have been established, which are now widely and routinely employed.[16,17] The assignment of the NMR resonances enables a wide variety of NMR studies on protein structures, dynamics, and interactions at an atomic resolution, even in the case of relatively

large proteins. For example, the structure determination of a protein up to 25 kDa can be achieved by employing uniform ^{13}C and ^{15}N labeling and the heteronuclear multidimensional NMR method.[18] The emergence of the ^{13}C and ^{15}N labeling approach and multidimensional experiments has enormously advanced NMR investigations of proteins smaller than 25 kDa.

It is noteworthy that the application of ^{13}C labeling is not limited to heteronuclear correlation experiments. An NMR experiment using direct ^{13}C observation may also yield useful structural information, which cannot easily be obtained by the other methods. The utility of the ^{13}C NMR experiments has been well demonstrated through the observation of backbone carbonyl carbons.[1,19–21] In contrast to the ^{1}H detection experiments, the signal dispersion in the ^{13}C chemical shifts and the narrower linewidths are advantages of the ^{13}C direct observation approach for relatively large proteins. The study on *Streptomyces* subtilisin inhibitor (SSI), a 23-kDa homodimer, is a classic example of such an approach. Residue selective ^{13}C labeling was used to observe the carbonyl carbon signal of the functionally important Met residue (Met73–Val74; P1–P1′).[19] To differentiate the three Met carbonyl carbon signals, an SSI sample simultaneously labeled with [1-^{13}C]Met and [^{15}N]Val was prepared. Because the other two Met residues have different N-terminal neighbors, only the Met73 signals showed a 15-Hz one-bond ^{13}C–^{15}N spin coupling. This type of method, which is referred to as selective ^{13}C,^{15}N double labeling, is the prototype of the widely used sequential assignment method using uniformly ^{13}C,^{15}N double-labeled proteins. The original selective ^{13}C,^{15}N double-labeling method by itself is still often used as a reliable assignment method for the carbonyl carbon and amide nitrogen signals of backbone peptides. Another interesting application of carbonyl carbons is the effect of deuterium–hydrogen exchange of amides on the line shapes, known as the deuterium–hydrogen exchange of amide on the line shapes (DEALS) method.[20] This method can be used to investigate the hydrogen exchange rate of backbone amide groups in proteins. When the carbon spectra of proteins dissolved in H_2O and D_2O solutions are compared, the chemical shifts of their backbone carbonyl carbons differ by $\sim$0.1 ppm between them. This difference arises from the proton/deuterium substitution in the amide groups of its own residue and the following residue, which is known as the deuterium isotope shift. In the DEALS method, the backbone carbonyl carbon peaks are observed in a mixture of H_2O and D_2O. In these circumstances, when the lifetime of the amide proton is longer than the inverse of the deuterium isotope shift in the hertz scale, the carbon peaks appear as separate peaks. Conversely, when the exchange rate is rapid, the carbon peaks appear as single averaged peaks. Therefore, slowly exchanging amide protons can readily be identified based on their line shape on the carbon spectrum. This method has an advantage in that the experiment can be conducted in a steady state, which enables the relatively precise estimation of the exchange rate, compared to other proton/deuterium exchange experiments that utilize real-time monitoring. Improvements in sensitivity, using cryogenic probes, have enabled ^{13}C detection in multidimensional protein NMR spectroscopy.[22–24] Further development of the pulse sequences with non-proton detection might allow the establishment of a new strategy of NMR structure determination without ^{1}H–^{1}H nuclear Overhauser effects (NOEs), using fully deuterated and ^{13}C-/^{15}N-labeled proteins.

The uniformly ^{13}C- and/or ^{15}N-labeled proteins are generally produced by growing *E. coli* cells transformed with the target DNA on minimal medium (M9) containing ^{13}C-labeled glucose and/or $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively.[1] The *E. coli* cell can synthesize the 20 amino acids from the isotope-labeled precursors, and then the synthesized amino acids are incorporated into the target proteins.

1.12.2.3 ^{2}H and ^{13}C/^{15}N Labeling

To further expand the applicability of NMR, the aforementioned ^{2}H labeling and ^{13}C/^{15}N labeling can be combined. This strategy is especially effective for analyzing backbone atoms. A suite of NMR experiments for assigning the backbone atoms of ^{2}H/^{13}C/^{15}N-labeled proteins has also been established.[25] Furthermore, transverse relaxation-optimized spectroscopy (TROSY), which utilizes the destructive interference between dipole interactions and chemical shift anisotropy, yields narrower lines for the amide and aromatic resonances of larger proteins.[26,27] With a polarization transfer scheme based on cross-correlated relaxation,[28] observations of the NMR signals of large protein–ligand complexes ($>$100 kDa) can be accomplished.[29] The combined use of random fractional deuteration and uniform ^{13}C /^{15}N labeling has been applied for the structural analysis of proteins.[30] However, it should be kept in mind that random fractional ^{2}H labeling causes some spectral complications due to the presence of several isotopomers.[31,32] The substitution of ^{1}H by ^{2}H induces upfield chemical shift changes of the carbons separated by one to three chemical bonds from the substituted atom. For example, in the case of a methyl group, as many as four isotopomers – namely CD_3, CHD_2, CH_2D, and CH_3 – exist simultaneously. Each isotopomer gives signals at slightly different positions on the [^{1}H,^{13}C] HSQC spectrum, leading to a more complicated NMR spectrum.[32]

1.12.3 Stereo-Array Isotope Labeling Principle

1.12.3.1 The Concept of the SAIL Method

As described previously, deuterium labeling is a key to overcome the size problem of protein NMR studies. The merits of the use of ^{2}H labeling, such as line narrowing and reduced spin diffusion, are quite important when studying large proteins by NMR. The SAIL method utilizes proteins exclusively composed of stereo- and regioselectively labeled amino acids (SAIL amino acids) for NMR studies.[7] In the SAIL amino acids, selected protons are substituted by deuterium, such that each carbon or nitrogen nucleus in the final protein will have at most one ^{1}H nucleus bonded to it. In addition, ^{13}C and ^{15}N are used only where a ^{1}H is bound to it or where it is needed for uninterrupted through-bond assignment pathways. To achieve this, SAIL amino acids are designed as follows (**Figure 1**):

1. Stereoselective replacement of one ^{1}H in a methylene group by ^{2}H.

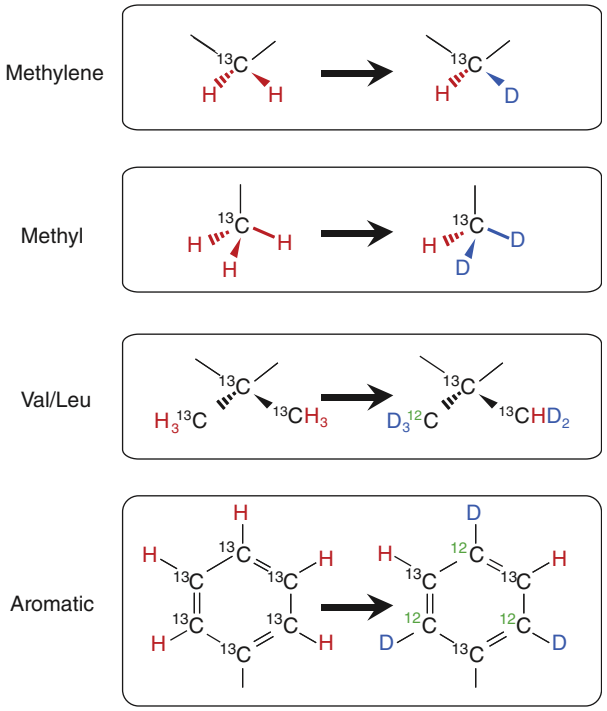

Figure 1 The design concept of the SAIL method. In methylene groups, one of the two methylene protons is stereospecifically substituted by deuteron. In the case of methyl groups, two of the three methyl protons are deuterated. In the case of the prochiral methyl groups in Val and Leu residues, the isotope labeling pattern of one methyl group is $^{13}CHD_2$ and the other is $^{12}CD_3$.

2. Replacement of two ^{1}H in each methyl group by ^{2}H.
3. Stereoselective modification of the prochiral methyl groups of Leu and Val, such that one methyl is $-^{12}C(^{2}H)_3$ and the other is $-^{13}C^{1}H(^{2}H)_2$.
4. Labeling of six-membered aromatic rings by alternating ^{12}C–^{2}H and ^{13}C–^{1}H moieties.

The 20 SAIL amino acids (**Figure 2**) were synthesized based on these design concepts by chemical and enzymatic means. These SAIL amino acids are commercially available from SAIL Technologies (http://www.sail-technologies.com).

The NMR signals from SAIL proteins are sharpened by both decreasing the long-range couplings and eliminating the dipolar relaxation pathways, in comparison to those of UL proteins. The labeling pattern of SAIL amino acids preserves the through-bond connectivity required for backbone and side chain assignments. However, the extensive deuteration of the side chain protons enables efficient transfer of the magnetization. In addition, the stereospecific deuteration in the methylene and prochiral methyl groups eliminates the need for their stereospecific assignment. Thus, a higher level of assignment can readily be achieved by using SAIL proteins compared to UL proteins.

1.12.3.2 Nuclear Overhauser Effect Spectroscopy from SAIL Proteins

The SAIL pattern is also assumed to improve the quality of NOE spectroscopy (NOESY) spectra, thereby leading to higher quality NMR structures. A brief consideration shows that the number of nonexchangeable side chain protons in SAIL proteins is less than half, and the number of expected NOESY cross peaks is decreased by 40–45%, relative to UL proteins. However, most of the additional NOEs from UL proteins either involve fixed (geminal) distances or become redundant in the absence of stereospecific assignments and cause spectral overlap without furnishing independent information. In practice, the stereospecific assignment of methylene and prochiral methyl groups is quite difficult. Indeed, the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu) contains the stereospecific assignments for well below 10% of the methylene protons and the Val/Leu methyl groups. On the other hand, the NOE distance constraints obtained from SAIL proteins are effective in defining the conformations of proteins due to the information from their exclusive stereospecific assignments[33] and the rare signal overlap. In addition, the proton density is decreased in SAIL proteins, which mitigates spin diffusion and thus improves the accuracy of the interproton distance measurements from NOE peaks (**Figure 3**). More detailed theoretical considerations,[7] which take into account the enhanced signal strength and sharper lines in SAIL spectra and the fact that overlap can render peaks unidentifiable, have shown that in practice, with larger molecules, SAIL is expected to increase, rather than decrease, the number of identifiable NOE cross peaks. The expected increase is moderate in regions without overlap but significant in regions with strong overlap. Thus, for larger proteins, SAIL is expected to yield two or more times the number of relevant conformational restraints as uniform labeling.

1.12.3.3 Overlap and Relaxation-Optimized SAIL Patterns

Ideally, the SAIL technique could be applied to the NMR structure determinations of proteins larger than 50 kDa and membrane proteins. To this end, it will presumably be necessary to further optimize the isotope labeling patterns of each SAIL amino acid in order to cope with the extensive crowding and line broadening resulting from the large molecular size. In an overlap and relaxation-optimized version of the SAIL approach, the number of ^{1}H nuclei is reduced further to enable the observation of well-shaped and separated signals, even for proteins larger than 50 kDa. SAIL patterns can be characterized by the percentage of ^{1}H atoms relative to the uniformly labeled protein in an H_2O solution. The original SAIL amino acids retain 64% of all protons and 44% of all side chain protons.[7] An overlap and relaxation-optimized SAIL pattern comprised 53% of all protons and only 28% of all side chain protons.[34] The modified SAIL ^{1}H labeling pattern is essentially a subset of the original SAIL amino acids. For instance, a uniformly ^{13}C-labeled Leu side chain contains four ^{13}C and nine ^{1}H nuclei. The original SAIL pattern reduces the NMR-active nuclei to three ^{13}C and three ^{1}H. In the case of overlap and relaxation-optimized SAIL, the H^{γ} methine proton, which tends to be overlapped and to provide only largely redundant NOEs, is additionally replaced by ^{2}H. C^{γ} can either be replaced by ^{12}C for further reduction of the relaxation or be kept as ^{13}C to enable the assignment of the $^{13}C^{1}H(^{2}H)_2$ methyl group by through-bond experiments. Simulations suggest

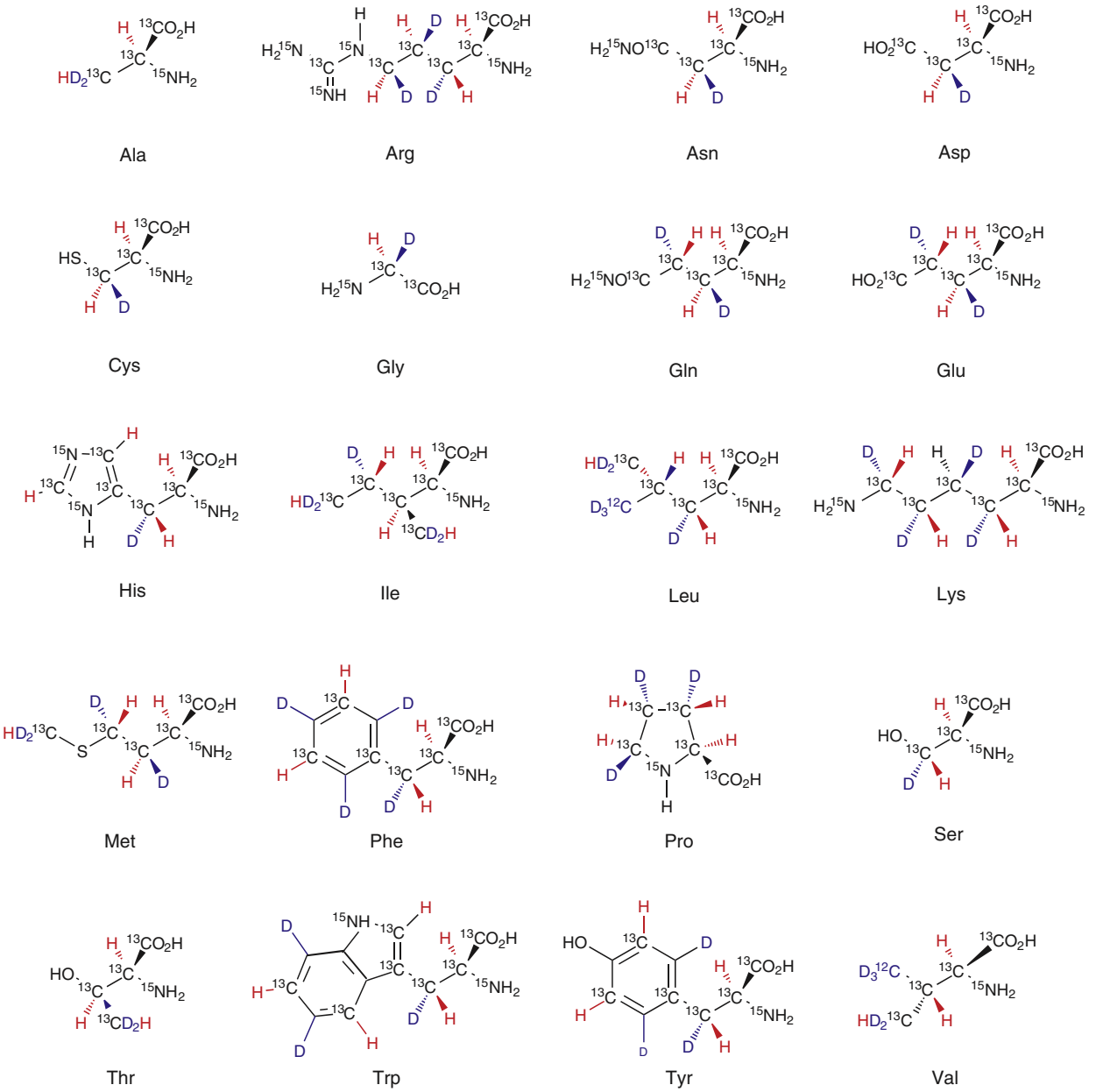

Figure 2 Chemical structures of 20 SAIL amino acids.

that the combined use of optimized SAIL patterns and automated structure calculation algorithms with CYANA has high potential for the automated structure determinations of higher molecular weight proteins as well as membrane proteins.[34] The precision of the structure generated using the overlap and relaxation-optimized modified SAIL pattern was slightly lower than that obtained by uniform labeling and the original SAIL method because the modified SAIL method reduces the number of peaks expected under ideal conditions. However, with the actual experimental data, better results can be expected for this overlap and relaxation-optimized SAIL because it facilitates the analysis of many otherwise broadened or overlapped signals. The overlap and relaxation-optimized SAIL patterns may thus contribute to the determination of high-quality solution structures of proteins in the 30- to 100-kDa range and membrane proteins.

1.12.4 Stereo-Array Isotope Labeling Technologies

1.12.4.1 SAIL Amino Acid Synthesis

The efficient synthesis of versions of all 20 common amino acids that embody the aforementioned SAIL design principles is the cornerstone of the SAIL technology. This has been achieved by chemical and enzymatic reactions using starting materials uniformly labeled with ^{13}C and ^{15}N.[35–40] As previously mentioned, the SAIL amino acids are commercially

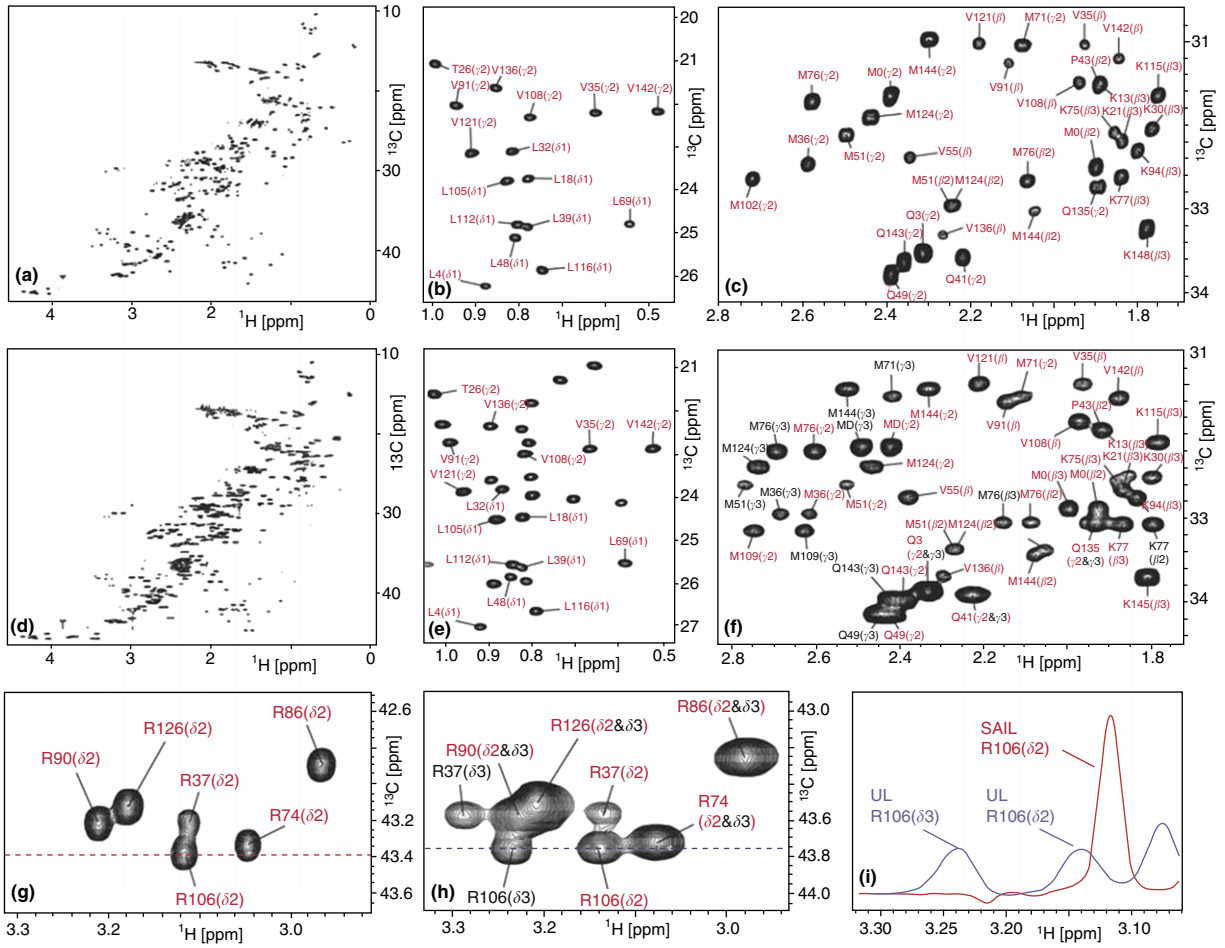

Figure 3 NMR spectra of UL and SAIL proteins. The [^{1}H,^{13}C] constant-time HSQC spectra of CaM. (a) SAIL-CaM, aliphatic region. (b) SAIL-CaM, methyl region. (c) SAIL-CaM, methylene region. (d) UL-CaM, aliphatic region. (e) UL-CaM, methyl region. (f) UL-CaM, methylene region. (g) SAIL-CaM, Arg δ region. (h) UL-CaM, Arg δ region. (i) Cross sections from panels g and h. The spectra for SAIL-CaM and UL-CaM were recorded under identical conditions and scaled for equal noise levels. Assignments are indicated by the one-letter amino acid code, the residue number, and the atom identifier. Reproduced from Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Ono, A. M.; Güntert, P. Optimal isotope labelling for NMR protein structure determinations. *Nature* **2006**, *440*, 52–57.

available from SAIL Technologies, and thus the preparation of SAIL proteins can start with the expression of proteins synthesized with the purchased SAIL amino acids.

1.12.4.2 Cell-Free Protein Synthesis

For the production of proteins exclusively composed of SAIL amino acids, cell-free protein expression methods[41,42] are crucial to use the SAIL amino acids efficiently and to prevent scrambling of the label, which would occur through metabolic pathways present in cell-based protein expression systems.[43,44] In the authors' laboratory, an *E. coli* S30 extract is prepared with modifications to remove the remaining unlabeled amino acids.[45,46] SAIL amino acids are incorporated most efficiently into the target proteins, without isotope scrambling or loss, by the modified *E. coli* cell-free protein expression system.[45,46]

The authors recommend performing both *E. coli* cellular expression and cell-free production of the target protein with uniform ^{15}N labeling for several reasons. First, this allows for a comparison between the NMR spectra obtained from *in vivo* and *in vitro* expression. The [^{1}H,^{15}N] HSQC spectra from both approaches should be compared to detect any possible difference between *in vivo* and *in vitro* expression. It should be kept in mind that N-terminal processing does not necessarily occur in the case of cell-free synthesis.[45] Second, it is less costly and requires less effort to produce proteins from *E. coli* cells than by the cell-free method. Thus, optimizations of the purification protocol and the NMR buffer conditions should be performed using proteins produced by *E. coli*. Third, most proteins that are expressed well in *E. coli* cells can be produced successfully by the cell-free method. Among 44 proteins that were expressed in *E. coli*, 35 (80%) were also found to work well in the cell-free synthesis system. Thus, a protein that is expressed well in *E. coli* cells also has a good chance of being produced well in the cell-free system.

For a successful cell-free reaction, the preparation of an S30 extract with high activity is quite important. Most investigators and commercial vendors of cell-free systems have focused on the level of activity of the extract.[47–51] For example, the activity

depends on the kind of *E. coli* cells used for the preparation. *Escherichia coli* strain A19 yielded a reliable extract because of its lack of RNase I.[49] The strain BL21 Star (DE3) (Invitrogen), lacking RNase E activity, had 1.2–1.4 times higher activity than A19 and was therefore chosen for the extract. The activity also depends on how the extract is prepared and its concentration. The presence of DTT in the S30 extract is necessary for preserving its activity over time. S30 extracts prepared with and without DTT have equivalent activity initially, but the activity of the extract without DTT declines to 80–90% after only 3 months of storage at −80 °C.[45]

In addition to the level of activity, the efficacy of isotope labeled amino acid incorporation is important. When the extract is prepared according to the original procedures, an abundance of unlabeled amino acids remains in the resulting extract. For example, when an S30 extract prepared according to the conventional protocol was used, the incorporation rate of the added amino acids into the target protein was only ~90%. The authors thus introduced a gel filtration step in the preparation of the S30 extract to remove the remaining amino acids. As a result, the incorporation level increased to 96%, without any loss of the activity of the extract. In addition, the concentration of the extract in the cell-free reaction should be optimized in terms of both the expression level and the incorporation rate of added SAIL amino acids. A higher concentration of extract leads to higher yields of protein but also to dilution of the label by the endogenous amino acids in the extract. This problem of label dilution has been quantitatively analyzed.[45]

The extract described previously produced up to 2 mg (*E. coli* peptidyl-prolyl *cis–trans* isomerase b (EPPIb)) protein per milliliter of the reaction solution with optimal concentrations of added amino acids (1 mM each). The protein yield decreases as the levels of added amino acids are reduced, although the labeling efficiency increases. In the case of calmodulin, the optimal concentration for the amino acid mixture was determined to be 1.7 mg ml^{-1}, representing a value slightly below the maximum determined by analyzing protein yields as a function of concentration. Because the labeled amino acids represent the most important cost factor, the mole ratios of the individual amino acids were also adjusted to match the amino acid composition of calmodulin, with the exception that the mole ratio of Gly was increased by a factor of 4. A sample of [^{13}C,^{15}N]-calmodulin, prepared using the dialyzed extract with the optimal level of amino acids, was hydrolyzed into amino acids and analyzed by liquid chromatography–quadrupole mass spectrometry (MS) and capillary electrophoresis MS to investigate the percentage of unlabeled amino acids incorporated into the protein. These were as follows: Gly, 2.2%; Ala, 2.9%; Ser, 3.2%; Pro, 3.2%; Val, 3.7%; Thr, 2.3%; Leu, 3.6%; Asx, 7.5%; Lys, 4.5%; Glx, 0.8%; Met, 1.9%; His, 2.2%; Phe, 1.8%; Arg, 5.7%; and Tyr, 3.0%. When the amount of Gly added was not increased by a factor of 4, the incorporation of unlabeled Gly increased to 12.7%. This is the reason why a higher concentration of Gly was used. If necessary, the levels of other unlabeled amino acids can be decreased by increasing the amounts of the labeled species.

Cell-free protein synthesis systems usually use a large amount of a plasmid containing sequences encoding the target protein, a promoter, and a terminator for T7 RNA polymerase. High-copy plasmids can reduce the cost and time required to produce the template DNA. The pIVEX vector provided by Roche has an advantage over other high-copy plasmids in that it enables the prediction of silent mutants that can improve cell-free synthesis yields.

Table 1 Composition of reaction and dialysis solutions in cell-free reactions for SAIL protein synthesis

Stock solution	*Reaction solution*	*Dialysis solution*
RNase-free water	1120 μl	11 608 μl
1.4 M NH_4OAc	98 μl	392 μl
0.5 M $Mg(OAc)_2$	150 μl	600 μl
SAIL amino acid mixture (25 mM each)	200 μl	800 μl
0.645 M creatine phosphate	400 μl	1600 μl
LM mixture	1250 μl	5000 μl
1 mg ml^{-1} template DNA	100 μl	–
11 mg ml^{-1} T7 RNA polymerase	45 μl	–
40 units μl^{-1} RNase inhibitor	12.5 μl	–
10 mg ml^{-1} creatine kinase	125 μl	–
S30 extract	1500 μl	–
Total volume	5 ml	20 ml

Reproduced from 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*, 311–325.

The preparation of a SAIL protein NMR sample by the *E. coli* cell-free method comprises the following steps:

1. Prepare the reaction solution and the dialysis solution by mixing the components, as shown in **Table 1**. Dissolve the SAIL amino acid mixture in water and then add it to the cell-free reaction solution. If the SAIL amino acids appear to be insoluble in water, warm the mixture up to 60 °C. Use sterile gloves to prevent contamination with RNases. Thaw the frozen S30 extract on ice. Prepare creatine phosphate in RNase-free water just before use. Excess heating of the SAIL amino acids may cause racemization, especially at high pH.
2. Cut the outer tube of the Float-A-Lyzer (Spectra/Por) at an appropriate height such that the inner solution in the tube will be completely immersed in the dialysis solution when the inner membrane apparatus is placed in the outer tube. Pour the dialysis solution into the outer tube. Place the inner membrane apparatus of the Float-A-Lyzer within the outer tube, and pour the reaction solution into the inner membrane. Cover the tube with Parafilm.
3. Shake the tube to facilitate the production of target proteins. The optimal temperature and incubation times should be determined in small-scale cell-free reactions with reaction and dialysis solution volumes of 0.5 and 2.0 ml, respectively, prior to the large-scale reaction.
4. Retrieve the reaction solution and the dialysis solution. If the produced protein has a molecular weight smaller than the molecular weight cutoff of the membrane, check the outer solution for the presence of the protein.
5. Purify the produced protein according to the purification procedures for the target protein. The N terminus of the protein produced by cell-free expression may be

heterogeneous due to incomplete deformylation by peptide deformylase. This problem can be overcome by using a cleavable N-terminal tag.[45]

This approach has been used successfully for the cell-free production of a large number of proteins.[45]

Small-scale reactions are useful for determining whether a target protein is actually synthesized. The recommended composition of reagents in the small-scale reaction is the same as that in the large-scale reaction used for the NMR sample preparation except for the presence of PEG-8000. PEG plays an important role as a stabilizer of the system.[49] In the case of calmodulin, the ratio of protein yields with and without PEG-8000 was 1.6. Even a small amount of PEG can disturb the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) quantitation of the level of target protein production, despite attempts to remove the PEG by acetone or trichloroacetic acid (TCA) precipitation of the protein. For example, after cell-free synthesis and prior to purification, the SDS-PAGE band from acetone- or TCA-precipitated calmodulin was not observed, although it could be observed following purification. For this reason, PEG should be omitted from the small-scale screening reactions.

The yield of proteins from *E. coli* cell-free synthesis reportedly depends on the concentrations of magnesium ions and the template DNA.[52] It is thus worthwhile to optimize these parameters, which would increase the yield of a particular protein in favorable cases. A better yield can also be achieved by introducing a silent mutation into the template DNA. The silent mutation does not alter the amino acid sequence of the resulting protein but leads to changes in the yield, presumably by affecting the formation of secondary structures in the mRNA. For this purpose, the ProteoExpert software (Roche Applied Science; Biomax) was used, although full details on the algorithm are unavailable. One round of calculations provides ten suggested candidates for silent mutations in the region close to the N terminus of the protein. Roche recommends introducing these mutations by polymerase chain reaction (PCR) into linear templates and using small-scale syntheses to monitor the relative yields. The mutated sequence that gives the highest yield is then transferred to a circular template and used for target protein expression.

Designing oligonucleotides with the best silent mutations and introducing them into the circular plasmid by restriction enzymes and ligase avoids possible difficulties and errors in the PCR. For calmodulin, improvements of the protein yields relative to the original sequence were observed for two of the ten predicted silent mutations. In the large-scale reaction performed with the best sequence and 51.1 mg of the amino acid mixture (1.7 mg ml^{-1}), the amount of calmodulin synthesized was 5.2 mg (10 wt %). This was twice the yield of the construct prior to the silent mutagenesis.

The authors have noticed that the positions of the predicted silent mutations are clustered in the region of the cleavable N-terminal tag, and thus the candidates are often identical for different proteins. This means that cloning the target DNA sequence into a limited number of pIVEX vectors with different silent mutations in the N-terminal cleavable tag will ensure that an expression vector with the optimal silent mutation is obtained. Based on this idea, the authors prepared pIVEX vectors encoding an N-terminal cleavable His-tag with a variety of silent mutations.[45] This N-terminal cleavable tag has the same amino acid sequence as that of pET15b, namely a $(His)_6$-tag followed by a thrombin cleavage site. The proteins expressed using this vector can readily be purified by immobilized metal affinity chromatography. Subsequently, the N-terminal tags are cleaved with thrombin. The cleavage is also effective for overcoming the problem of the incomplete N-terminal formyl-Met processing that occurs in the *E. coli* cell-free reaction, as described next.

The [^{1}H,^{15}N] HSQC spectrum of calmodulin synthesized by the cell-free system using the native sequence showed a larger number of peaks than that predicted from the amino acid sequence, although the purified protein was observed as a single band by SDS-PAGE. Doubled peaks were found to arise from the six N-terminal residues and, with smaller chemical shift differences, for eight residues in other regions. The intensities of the superfluous peaks were not always reproducible. N-terminal sequencing and mass spectrometry revealed that the sample consisted of three molecular species: one with N-terminal formyl Met (f-Met_0) (60–90%), one with N-terminal Met0 ($\sim$10%), and one with N-terminal Ala1 (10–40%). Moreover, for the EPPIb protein with 164 amino acid residues, 33 additional peaks were observed in the [^{1}H,^{15}N] HSQC spectrum of the protein prepared by the cell-free system compared to the spectrum of the protein produced in *E. coli* cells.[45]

This problem could be overcome by producing the protein with a cleavable N-terminal tag, which can simultaneously be engineered to improve protein yields, as described previously. The [^{1}H,^{15}N] HSQC spectrum of calmodulin produced in this manner showed no detectable extra peaks.[45] It is noteworthy that the NMR signals from calmodulin, which has a flexible N terminus, are affected by the heterogeneity over a quite extensive region. Larger spectral complications may arise from peptide chain heterogeneity for proteins with a structured N terminus.

In large-scale protein preparations, it is critically important to minimize product losses. The molecular weight cutoff of the membrane used for the cell-free reaction should be 50 kDa because membranes with a smaller molecular weight cutoff reduce the yield of protein.[47] Because this cutoff is larger than most of the proteins that are prepared for NMR spectroscopy, much of the synthesized protein goes into the outer dialysis solution during the reaction. For calmodulin (17 kDa), after the 7-h reaction, the amount of the protein product in the inner reaction solution equaled that in the outer solution, although the protein in the outer chamber was sometimes difficult to detect by SDS-PAGE due to the dilution factor.[45] Hence, the synthesized protein was retrieved from both solutions. The purification protocol must also be optimized for large-scale production. It is frequently advantageous to dialyze the protein solution against the buffer used for the first purification step before purification to remove nucleic acids, salts, DTT, or other contaminants that may interfere with chromatography.

The cell-free protein expression protocol presented in this section can efficiently produce proteins in the quantities required for multidimensional NMR measurements. For

instance, 55 mg of the SAIL amino acid mixture, composed following this protocol, yielded 5.5 mg of purified, soluble 17-kDa calmodulin and 5.3 mg of purified, soluble 41-kDa maltodextrin binding protein (MBP).[7]

1.12.4.3 Residue-Selective SAIL Method

In addition to the full-SAIL labeling method, residue-selective labeling by SAIL amino acid(s) is also a powerful approach. In this case, not only the cell-free expression system but also well-established and thus more robust *in vivo* expression systems can be used for protein production, provided that the added SAIL amino acid is not affected by metabolic scrambling. Compared to *in vitro* expression, however, *in vivo* expression according to the conventional procedure requires a much larger amount of the amino acid to obtain the intended amount in the target protein. One strategy to overcome this problem involves the use of auxotrophic *E. coli* strains. By growing the *E. coli* auxotrophic strain in minimal medium containing a small amount of the target isotope labeled amino acid and a large amount of the 19 other unlabeled amino acids, a reasonable yield of the selectively labeled protein can be obtained:

1. Transform the expression vector into an auxotrophic *E. coli* strain. The authors often use AB2826 (DE3) strains for selective labeling by SAIL aromatic amino acids.
2. Prepare amino acid-containing M9 medium with different amounts of phenylalanine. The basic composition of the medium is listed in **Table 2**. The key point concerning this method is that the amount of the target amino acid can be decreased without reducing the yield. For instance, in the case of phenylalanine, the amount can be decreased to $\sim 10\ \mathrm{mg\ l^{-1}}$ while retaining a reasonable yield.
3. Pick the colonies, and inoculate them into LB medium.
4. Pellet the *E. coli* cells grown on the LB medium, and resuspend them in the minimal medium.
5. Grow the *E. coli* cells to an OD_{600} of 0.6–0.7.
6. Induce the expression by adding IPTG.
7. After a defined amount of time, stop the culture and check the protein expression by SDS-PAGE.

Table 2 Composition of amino acid containing M9 medium

For 1 l, combine the following	
$Na_2HPO_4 \cdot 12\ H_2O$	15.1 g
KH_2PO_4	3.0 g
$^{15}N \cdot NH_4Cl$	1.0 g
NaCl	0.5 g
L-Alanine	400 mg
L-Arginine	400 mg
L-Aspartic acid	250 mg
L-Cysteine	50 mg
L-Glutamic acid	400 mg
L-Glycine	400 mg
L-Histidine	100 mg
L-Isoleucine	100 mg
L-Leucine	100 mg
L-Lysine	150 mg
L-Methionine	50 mg
L-Phenylalanine	50 mg
L-Proline	150 mg
L-Serine	1000 mg
L-Threonine	100 mg
L-Tryptophan	50 mg
L-Tyrosine	100 mg
L-Valine	100 mg
After autoclaving, add the following	
1 M $MgSO_4$	1 ml
0.1 M $CaCl_2$	1 ml
8% thiamine	0.5 ml
20% D-glucose	10 ml

The minimum amount of the SAIL amino acid required for efficient expression varies for different target proteins and labeled amino acids. To reduce the cost, it is desirable to determine the minimum amount of SAIL amino acid required in a small-scale culture. In favorable cases, an incorporation rate of more than 10% of the SAIL amino acid into the target protein can be achieved, which is the same level as that in cell-free reactions.

1.12.5 SAIL NMR Spectroscopy

1.12.5.1 General Considerations

Most heteronuclear multidimensional NMR experiments should benefit from SAIL and can be used with minor modifications for SAIL proteins. When acquiring data for a SAIL protein, deuterium decoupling should be applied during chemical shift encoding on the aliphatic ^{13}C carbon to avoid splitting by the ^{13}C–^{2}H coupling. The [^{1}H,^{13}C] constant time HSQC NMR data sets collected from SAIL and uniformly labeled proteins clearly demonstrate the superiority of the SAIL method (**Figure 3**). The severe signal overlap observed for uniform labeling is alleviated in the spectra from the SAIL proteins. The improvements, which are particularly apparent for larger proteins, result from the reduction in the number of ^{1}H signals as well as from the sharpening of the remaining signals.

The SAIL method also improved sensitivity. Part of the gain arises from the longer ^{1}H and ^{13}C transverse relaxation times resulting from the substitutions of ^{1}H by ^{2}H. The reduced relaxation during magnetization transfer steps in experiments such as constant-time [^{1}H,^{13}C] HSQC leads to an increased signal-to-noise ratio. The fewer long-range couplings result in further signal sharpening. The signal intensities for methylene groups are three to seven times higher with SAIL than with uniform labeling under the same conditions. Although each observed SAIL methyl group contained only one ^{1}H, compared to three equivalent ^{1}H with uniform labeling, comparable signal intensities were observed for the methyl groups as a result of the longer ^{1}H and ^{13}C transverse relaxation times.

1.12.5.2 Assignment of Aromatic Signals Using SAIL Amino Acids

Because aromatic amino acid residues are often located in the hydrophobic core regions of proteins, the extensive collection of NOE restraint data involving aromatic ring signals is

essential for accurate protein structure determination. However, it is often difficult to analyze these spectral regions in uniformly $^{13}C/^{15}N$-labeled proteins, especially for Phe and Tyr residues. In addition to the inherently low sensitivity, the resonances of the δ, ε, and ζ ring carbons of Phe residues and the δ carbons of Tyr residues are prone to overlap with each other in NMR spectra. Such limited chemical shift dispersion and strong $^{13}C–^{13}C$ spin couplings hamper the spectral analyses, especially for Phe residues. Furthermore, exchange broadening due to aromatic ring flipping motions, which have occasionally been reported in the literature,[53,54] makes the NMR signals for the δ and ε atoms of some Phe and Tyr residues even more difficult, or sometimes impossible, to observe. In the SAIL aromatic amino acids, new $^{2}H,^{13}C$ labeling patterns were introduced to make the NMR spectra for aromatic resonances simple and sensitive and to facilitate their spectral assignment by exclusively using through-bond correlations.[39] **Figure 4** illustrates the repertoire of the currently available SAIL Phe and SAIL Tyr amino acids. In the [$^{13}C^{\gamma}$,$^{13}C^{\varepsilon}$,$^{1}H^{\varepsilon}$]-Tyr/Phe pattern (ε-SAIL Tyr/Phe), all nonexchangeable hydrogens in the aromatic rings, except for the ε position, were replaced with ^{2}H, and only the carbons at the γ and ε positions were labeled with ^{13}C. In the other pattern, denoted as [$^{13}C^{\gamma}$,$^{13}C^{\zeta}$,$^{1}H^{\zeta}$]-Phe (ζ-SAIL Phe), the hydrogens at the δ and ε positions of Phe were replaced with ^{2}H, and the carbons at the γ and ζ positions were labeled with ^{13}C. In all cases, the aliphatic moieties were uniformly labeled with ^{13}C and ^{15}N, and the pro-R hydrogen of the β-methylene group was stereospecifically labeled with ^{2}H, thus enabling the unambiguous stereospecific assignment of the prochiral methylene signals. All of the SAIL Phe and Tyr residues lack directly bonded aromatic $^{13}C–^{13}C$ pairs in order 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.

Figure 5 compares the aromatic regions of [$^{1}H,^{13}C$] HSQC spectra of EPPIb, containing uniformly [$^{13}C,^{15}N$]-labeled Phe, δ-SAIL Phe, ε-SAIL Phe, or ζ-SAIL Phe. It is obvious that the severe signal overlap observed for uniformly [$^{13}C,^{15}N$]-labeled Phe is resolved by using SAIL Phe because of the absence of resonances and the sharpening of the remaining resonances. Employing SAIL Phe increased the aromatic signal intensities. One of the reasons for the improvement is the attenuation of dipolar interactions, which was achieved by replacing the hydrogens adjacent to the detected nuclei with ^{2}H. In the case of the uniformly ^{13}C-labeled aromatic ring, a long (17- or 34-ms) constant time evolution period, which causes a reduction of the magnetization, must be used in the [$^{1}H,^{13}C$] HSQC measurement to refocus the $^{13}C–^{13}C$ couplings. In contrast, the SAIL Tyr/Phe labeling patterns do not require constant time evolution in the [$^{1}H,^{13}C$] HSQC measurement.

Unambiguous assignments of the aromatic ring resonances of SAIL Tyr/Phe can readily be achieved by the NOE connectivity between H^{δ} and H^{β} for δ-SAIL Tyr/Phe or by through-bond $^{13}C–^{13}C$ scalar coupling connectivities through $^{13}C^{\gamma}$ for ε-SAIL Tyr/Phe and ζ-SAIL Phe.[39] Two pulse sequences that use the scalar connectivities through a $^{13}C^{\gamma}$ atom were developed: the HBCB(CG)HE experiment with the magnetization transfer pathway H^{β} $(t_1)\rightarrow C^{\beta}$ $(t_2)\rightarrow C^{\gamma}\rightarrow H^{\varepsilon}$ (t_3) and the HBCB(CGCZ)HZ experiment H^{β} $(t_1)\rightarrow C^{\beta}$ $(t_2)\rightarrow C^{\gamma}\rightarrow C^{\zeta}\rightarrow H^{\zeta}$ (t_3). The 3-D HBCB(CG)HE experiment can also be performed in a 2-D version, (HB)CB(CG)HE. These pulse sequences are composed of previously employed building blocks[55] and are similar to the previously reported sequences, $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta})H^{\delta}$ and $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta}C^{\varepsilon})H^{\varepsilon}$.[56] A major difference between the previously published sequences and the sequences for SAIL proteins is the incorporation of H^{β} evolution. This was feasible because the stereoselective deuteration in the β methylene group alleviated the overlap of the H^{β} resonances and prolonged the H^{β} relaxation time. In addition, the slower relaxation of $^{13}C^{\beta}$, due to the deuteration, also enhanced the measurement sensitivity. In the case of HBCB(CG)HE, the magnetization is transferred from C^{γ} to H^{ε} directly via $^{3}J_{C_{\gamma}H_{\varepsilon}}$ (7.9 Hz for Phe and 7.1 Hz for Tyr) after the transfer to C^{γ}. In the HBCB(CGCZ)HZ experiment, the magnetization is transferred from C^{γ} to C^{ζ} via $^{3}J_{C_{\gamma}C_{\zeta}}$ (9.3 Hz), before the final $C^{\zeta}–H^{\zeta}$ one-bond INEPT transfer. The site-specific ^{2}H and ^{13}C labeling patterns made it possible to transfer the magnetization through long-range scalar couplings. Using these experiments, the correlations between the H^{β} and the ring protons could be detected for Phe and Tyr residues in calmodulin and EPPIb, except for those undergoing aromatic flipping-broadening.[39,57]

The $^{13}C–^{1}H$ pairs of aromatic ring moieties are considered to be amenable for strong TROSY effects, but the results

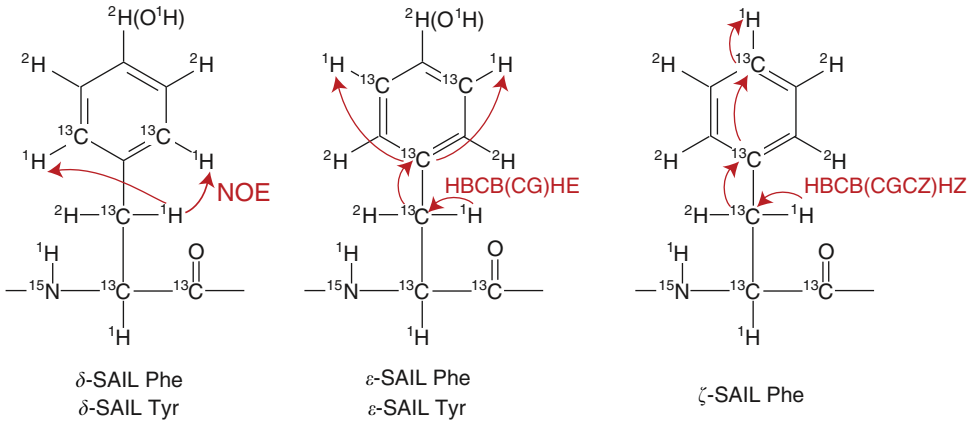

Figure 4 Chemical structures of various SAIL phenylalanines and tyrosines. The magnetization transfer pathways and their acronyms, which can be used to assign the δ, ε, and ζ signals, are shown by the arrows.

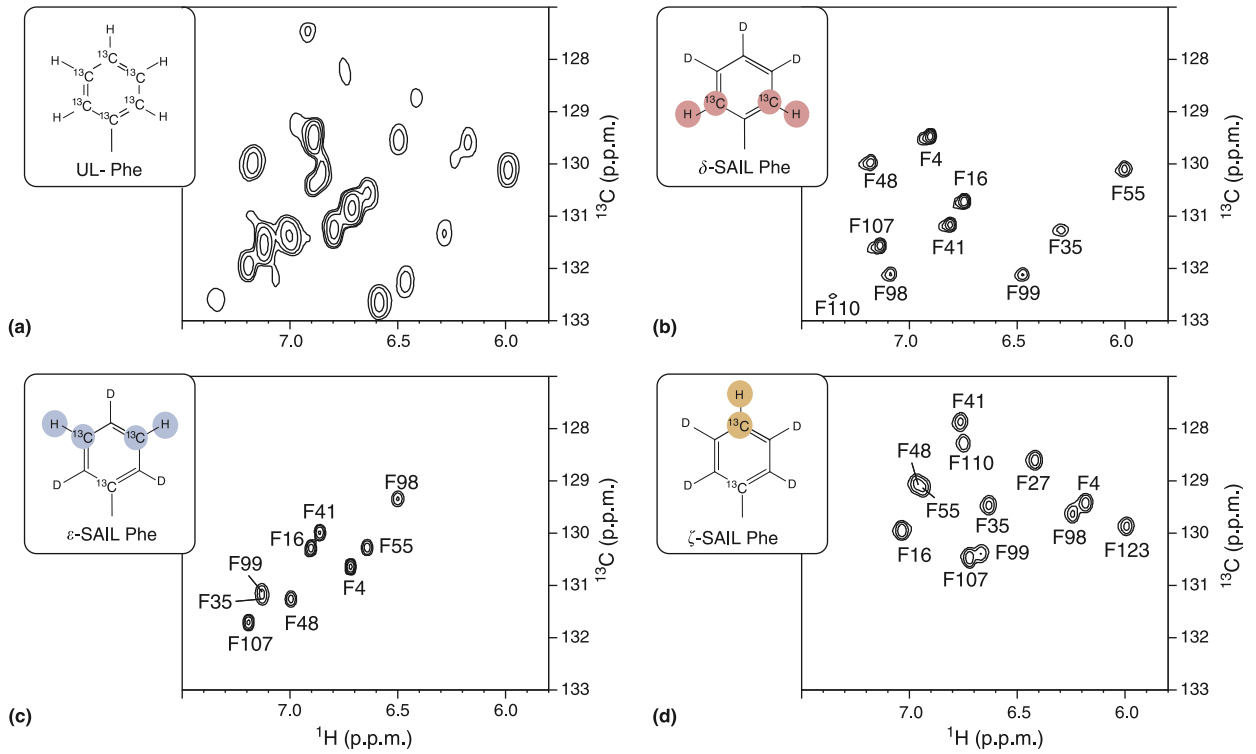

Figure 5 NMR spectral comparisons of the aromatic regions of EPPIb's selectively labeled with the δ-, ε-, and ζ-SAIL phenylalanines to that of the uniformly ^{13}C-labeled EPPIb. The [^{1}H,^{13}C] HSQC spectra were measured at 30 °C using a DRX600 spectrometer (Bruker) equipped with a TCI cryoprobe. A 200-μl portion of a 0.1- to 0.2-mM solution of SAIL EPPIb, dissolved in 10 mM sodium phosphate buffer (pH 6.6), containing 100 m NaCl and 1 mM DTT, in a Shigemi tube was used for NMR measurements. The time points were 1024 (t_1) × 256 (t_2). EPPIb was specifically labeled with (a) uniformly ^{13}C-labeled Phe, (b) δ-SAIL Phe, (c) ε-SAIL Phe, and (d) ζ-SAIL. For panel 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 panels b–d. The assignments given for each resonance were established by the previously reported pulse sequence for panels c and d.[39]

reported thus far are not quite satisfactory for uniformly ^{13}C-labeled, fully protonated proteins.[27] In each type of SAIL Tyr/Phe, however, all of the ^{13}C–^{1}H moieties are separated from each other by at least one ^{12}C–^{2}H group, and thus the TROSY effect works perfectly. The linewidths of the TROSY peaks observed for the F1-coupled [^{1}H,^{13}C] HSQC spectra are narrower than those of the anti-TROSY peaks for the 18-kDa protein EPPIb, and it has been found that the TROSY components of SAIL Phe remain reasonably sharp, even for protein particles as large as 80 kDa. Thus, the use of SAIL aromatic amino acids opens up new possibilities for studying very large proteins.

1.12.5.3 Structural Analysis of SAIL Proteins

The superior quality of SAIL protein NMR spectra and the reduced number of ^{1}H resonances enormously simplifies all types of NMR spectra. In TOCSY-type experiments for side chains, such as HCCH-TOCSY, the efficiency of spin-lock magnetization transfer is improved due to the extensive deuteration attached to the side chains. Compared to UL proteins, a high degree of resonance assignment can readily be accomplished by using SAIL proteins. As described in Section 1.16.3.2, spin relaxation is weakened in SAIL proteins compared to the corresponding UL proteins, which allows long NOE-derived distance constraints to be obtained. Prior knowledge of the stereospecific assignment eliminates the need for using pseudo atoms for methylene and prochiral methyl groups. Distance restraints for the structure calculation were obtained from 3-D ^{15}N- and ^{13}C-edited NOESY spectra[58] and from a 2-D NOESY spectrum for the aromatic region. These spectra were simplified by the presence of fewer signals, as expected. The reduced spin-diffusion maximum NOE intensities for SAIL proteins were typically reached at 1.5–3 times longer mixing times than with uniform labeling (**Figure 6**).

This simplification of the NMR spectra opens up the possibility of the automated analysis of proteins of 10–20 kDa or higher. The automated NOE assignment algorithm of the program CYANA automatically assigns NOESY cross peaks[59,60] on the basis of the chemical shift assignments and lists the NOESY cross peak positions and intensities. The fully automated NMR structure analysis algorithm (FLYA)[61] can solve NMR protein structures by purely computational means using only the primary structure and the NMR spectra as input. As in the classical manual approach, structures are determined by a set of experimental NOE distance restraints without reference to previously existing structures or empirical molecular modeling information. In addition to the 3-D structure of the protein, FLYA yields backbone and side chain chemical shift assignments and cross peak assignments for all spectra. By uniting optimal labeling with full automation, SAIL-based fully automated protein structure analysis (SAIL-FLYA) is

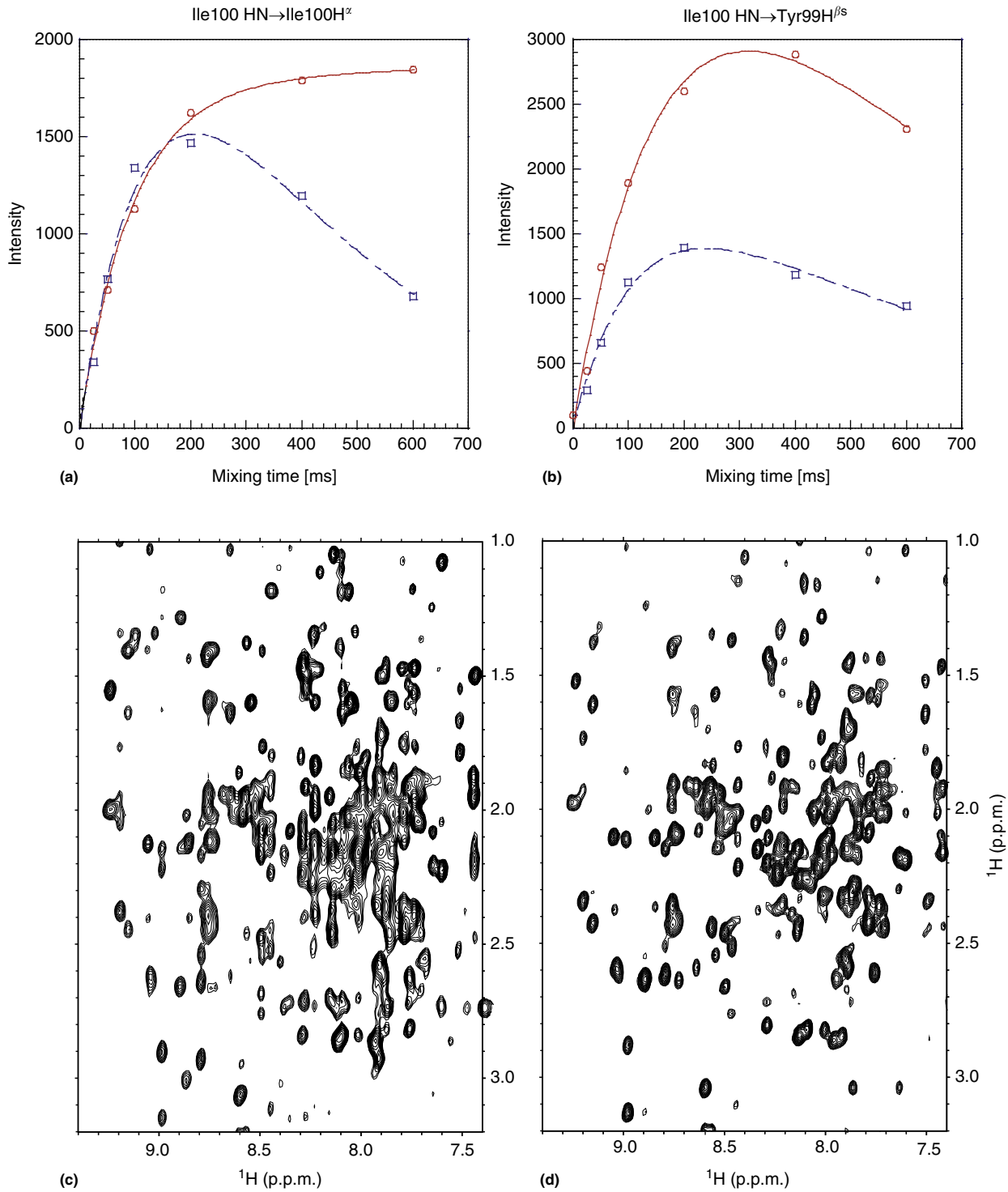

Figure 6 Comparison of NOESY data for SAIL-CaM and UL-CaM. The SAIL-CaM and UL-CaM samples each contained 0.7 mM protein, 5 mM MES-d13 and 10 mM bis-tris-d19 (Cambridge Isotope Laboratories), 5 mM $CaCl_2$, and 0.1 mM NaN_3, pH 6.5. Intensities of the cross peaks Ile100HN-Ile100H$^{\alpha}$ (a) and Ile100HN-Tyr99H$^{\beta s}$ (b) as a function of the mixing time in 3-D NOESY-HSQC experiments performed with SAIL-CaM (red) or UL-CaM (blue). Regions from the 2-D NOESY spectra for UL-CaM (c) and SAIL-CaM (d). The NOESY spectra for SAIL-CaM and UL-CaM were obtained under identical conditions at 37 °C on a Bruker DRX800 spectrometer equipped with a TXI xyz-gradient probe and were scaled to equal noise levels.

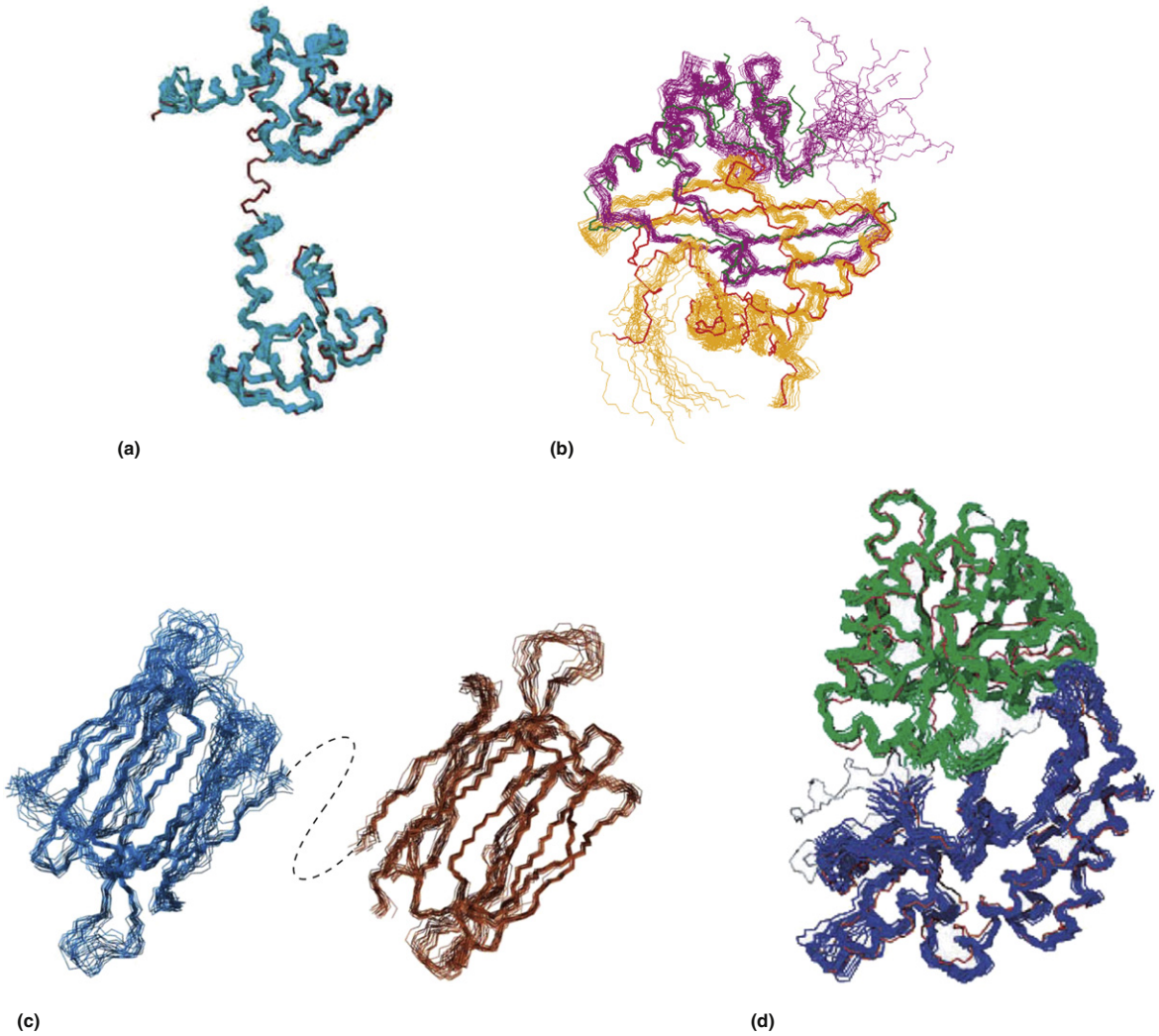

Figure 7 NMR structures of SAIL proteins. (a) SAIL-CaM (backbone in cyan, and Ca^{2+} in white), CaM X-ray structure (red), and three solution conformers of UL-CaM determined from residual dipolar coupling data (blue). (b) SAIL-C-terminal RNA-binding domain of SARS coronavirus nucleocapsid protein. (c) SAIL-At3g16450.1, composed of the N-terminal domain (blue) and the C-terminal domain (red). (d) MBP solution and crystal structures. Backbone of the N-terminal domain (SAIL-MBP in green, and X-ray in red) and the C-terminal domain (SAIL-MBP in blue, and X-ray in red).

designed to solve protein structures rapidly and with enhanced accuracy and size-range applicability.[46,62]

1.12.5.4 Selected Applications of the SAIL NMR Method

The SAIL method has been applied to several proteins (Figure 7). The overall convergence of the SAIL NMR structures is good. In addition, the backbone traces of the obtained NMR structures are quite similar to those of the corresponding crystal structures, verifying the accuracy of the NMR structures. Here, some NMR structures determined by the SAIL method are introduced.

1.12.5.4.1 Calmodulin

Calmodulin is a well-known calcium binding protein. Structures of the calcium-bound form of the 17-kDa calmodulin protein were determined previously by crystallography[63] and NMR.[64]

The structure of calmodulin is composed of two globular domains connected by a flexible linker. The previous NMR structure employed residual dipolar couplings measured mainly for the polypeptide backbone, in addition to conventional NOE distance restraints, to improve the structural quality. The application of the SAIL method to the structure of calmodulin vastly improved the quality of the NMR spectra, especially for the side chains.[7] Based on the side chain information, the NMR structure of the SAIL calmodulin showed good convergence for each domain using only NOE distance constraints and TALOS-derived angle constraints (Figure 7(a)). The structure of SAIL calmodulin is in close agreement with the crystal structure and the RDC-based backbone structure.

1.12.5.4.2 C-terminal dimerization domain of SARS coronavirus nucleocapsid protein

The C-terminal domain (CTD) of the SARS coronavirus (SARS-CoV) nucleocapsid protein (NP) contains a potential

RNA binding region in its N-terminal portion and also serves as a dimerization domain by forming a 28-kDa homodimer. When using uniformly labeled material, the structure determination of the SARS-CoV NP CTD in solution was impeded by the poor quality of the NMR spectra, especially for the aromatic resonances. The SAIL-labeled protein yielded less crowded and better resolved spectra than the uniformly ^{13}C- and ^{15}N-labeled protein. Profound improvement of the aromatic spectra quality was observed for SAIL proteins, which enabled the determination of the homodimeric solution structure.[65] The NMR structure is almost identical to the previously solved crystal structure, except for a disordered putative RNA binding domain at the N terminus (**Figure 7(b)**). The disordered region contains positively charged residues, which were shown to be involved in nucleic acid interactions. Thus, the disorder of the N-terminal RNA binding region is assumed to increase the capture radius of the protein for interacting nucleic acids.[65]

1.12.5.4.3 Putative 32-kDa myrosinase binding protein from Arabidopsis (At3g16450.1)

The product of the At3g16450.1 gene from *Arabidopsis thaliana* is a 32-kDa, 299-residue protein classified as resembling a myrosinase binding protein. These proteins are found in plants as part of a complex with the glucosinolate-degrading enzyme, myrosinase, and are suspected to play a role in the myrosinase-dependent defense against pathogens. Many myrosinase binding proteins and their relatives are composed of repeated homologous sequences with an unknown structure. The size of the protein is larger than that amenable to high-throughput analysis by uniform ^{13}C/^{15}N labeling methods. Nevertheless, the 3-D structure of the At3g16450.1 protein from *Arabidopsis*, which consists of two tandem repeats, was solved with SAIL.[66] NMR data sets collected with the SAIL protein enabled the assignment of the ^{1}H, ^{13}C, and ^{15}N chemical shifts to 95.5% of all atoms, even at the low concentration (0.2 mM) of the protein product. Using NOESY data, the 3-D structure was calculated with the CYANA software package. The structure, the first for a myrosinase binding protein family member, revealed that the At3g16450.1 protein consists of two independent, but similar, lectin-fold domains composed of three β sheets (**Figure 7(c)**).

1.12.5.4.4 Maltodextrin binding protein

The solution structure of the 41-kDa MBP was determined in order to prove that the SAIL method could render large proteins amenable to NMR structure determination.[7] Due to the improved sensitivity and resolution, the signals of SAIL MBP could readily be assigned by established methods to the extent of 94% for SAIL MBP, including more than 90% of the aliphatic and aromatic side chain protons. Many of the shifts that could not be assigned are clustered in the region of residues 229–241, which interacts with the bound cyclodextrin[67] and is assumed to undergo conformational exchange.[68]

A dense network of NOEs, including 949 non-redundant, long-range distance restraints, was established for SAIL MBP. Among the 3818 non-redundant NOE distance restraints, 1879 involve side chain atoms beyond H^{β}. On the basis of the simplified and more quantitative NOESY spectra, detailed structures of SAIL MBP were obtained by means of the combined automated NOE assignment and structure calculation protocol in the program CYANA.[59,60] The solution structure of the 41-kDa SAIL MBP coincides closely with the crystal structure[67] that was determined previously under slightly different conditions (**Figure 7(d)**). The 41-kDa SAIL MBP solution structure has similar precision and accuracy to those of smaller proteins, and the structural statistics are comparable to those commonly found in NMR structure determinations of smaller proteins. Previously,[68] a global fold of MBP was determined by NMR on the basis of the NOEs between the amide and methyl protons, the residual dipolar couplings for the polypeptide backbone, and the hydrogen bond restraints. Although that NMR study determined the global fold of the polypeptide backbone and the conformations of the methyl-containing side chains of valine, leucine, and isoleucine, the approach could not provide direct structural information for the other side chains.

1.12.5.4.5 Escherichia coli peptidyl-prolyl cis–trans isomerase b

EPPIb is an 18.2-kDa protein. The backbone assignment of EPPIb was previously accomplished by using a uniformly ^{13}C,^{15}N double-labeled sample prepared by the cell-free method.[69] EPPIb contains 12 Phe and 3 Tyr residues, some of which form a hydrophobic cluster. Thus, the collection of NOEs involving the aromatic residues is essential for the precise structure determination of this protein. For this purpose, the use of SAIL aromatic residues was found to be very effective. The aromatic ring signals of its 12 Phe residues were well resolved for all SAIL Phe types and could be assigned readily by correlating them to the previously assigned backbone signals.[57] Importantly, for the δ- and ε-SAIL Phe-labeled EPPIb proteins, the signals of Phe27, Phe110, and Phe123 were broadened beyond detection due to the intermediate ring flipping rates of these residues at 30 °C, except for a single, very weak peak assigned to the δ-^{13}C/^{1}H signal of Phe110. At 40 °C, this weak δ-^{13}C/^{1}H cross peak became sharper due to the increased ring flipping rate of Phe110. In contrast, all 12 peaks were clearly observed for the ζ signals in ζ-SAIL Phe-labeled EPPIb. The ζ signals are not affected by ring flipping because the ζ^{13}C–^{1}H moiety lies on the C^{β}–C^{γ} axis.

To further compare the properties of each SAIL Phe/Tyr residue, the authors prepared three SAIL EPPIb samples that were labeled with three different combinations of SAIL Phe/Tyr – namely δ-SAIL Phe/Tyr, ε-SAIL Phe/Tyr, and ζ-SAIL Phe/ε-SAIL Tyr – whereas standard SAIL amino acids were used for the other residues.[57] The number of observed NOESY cross peaks differed significantly between the three SAIL Phe/Tyr combinations. More than 100 NOESY cross peaks involving the aromatic rings of Phe and Tyr were observed for the δ-SAIL Phe/Tyr combination, whereas the corresponding number of NOEs for the ζ-SAIL Phe/ε-SAIL Tyr combination was $\sim$50. However, approximately half of the NOEs acquired for the δ-SAIL Phe/Tyr combination were intraresidual or sequential and thus did not conceivably contribute to defining the overall structure. In contrast, in the case of the ζ-SAIL Phe/ε-SAIL Tyr combination, most of the NOEs were observed between sequentially distant residues. Consequently, the convergence of the resulting structural ensemble calculated with CYANA is much better for the ζ-SAIL Phe/ε-SAIL Tyr combination than

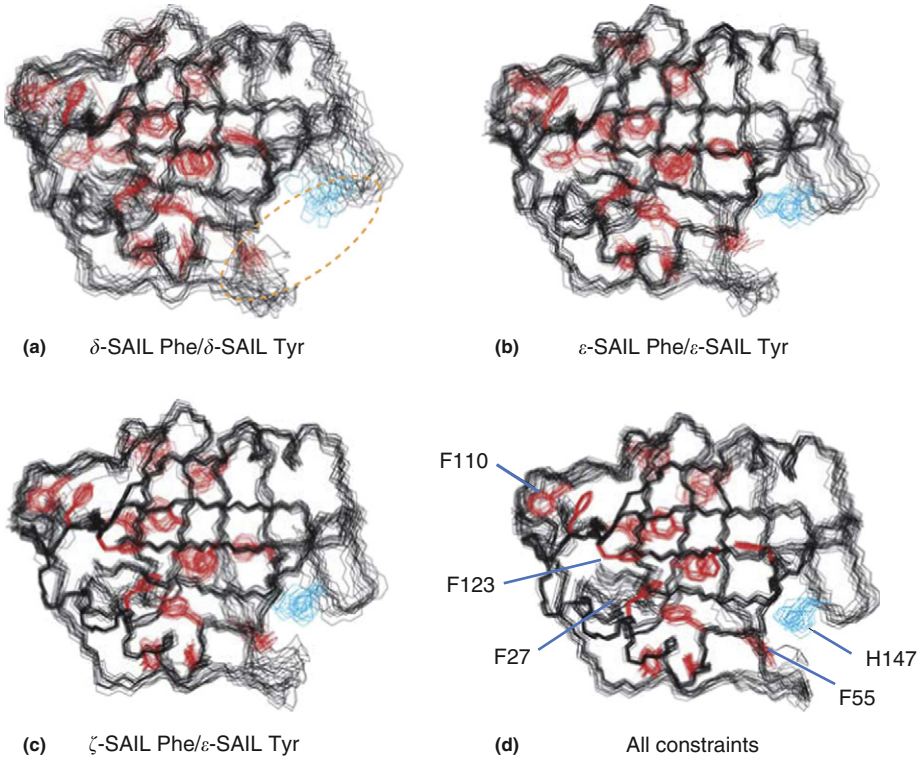

Figure 8 Structures of EPPIb determined by the SAIL method using various combinations of SAIL Phe and Tyr. Twenty overlaid structures calculated by CYANA for SAIL EPPIb's, labeled with (a) δ-SAIL Phe/Tyr, (b) ε-SAIL Phe/Tyr, and (c) ζ-SAIL Phe and ε-SAIL Tyr, with the rest of the amino acids fully labeled with SAIL amino acids. (d) The structures calculated by the combined use of all NOEs obtained for panels (a)–(c). Phenylalanine and tyrosine residues are colored red. His147 is colored cyan.

for the δ-SAIL Phe/Tyr combination (**Figure 8**). The precision of the structure was slightly lower for the δ-SAIL Phe/Tyr combination compared to the other structures, especially for two loop regions on the protein surface. This local structural difference is due to the complete lack of NOE restraints between the rings of Phe55 and His147, which were clearly observed for ε- and ζ-SAIL Phe but not for δ-SAIL Phe. The lack of NOEs for the δ and ε protons of Phe27, Phe110, and Phe123 is due to the intermediate ring flipping rates and is clearly reflected in the poorly defined ring positions of these three Phe residues, although their backbone conformation was defined by other NOEs. The structure determined by the simultaneous use of all NOEs observed with the aforementioned three combinations of SAIL Phe/Tyr was well-defined, with a root mean square deviation against the crystal structure of EPPIb[70] of 1.43 Å for the backbone atoms, excluding the unstructured loop regions. Because 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 structure determinations.

1.12.5.5 Extended DEALS Method

The SAIL method is also a powerful tool for investigations of protein dynamics, as well as the averaged structure. One of the applications is a hydrogen exchange study of Tyr side chain hydroxyl groups.[71] In principle, this method is an extension of the DEALS method, which is used for investigating the hydrogen exchange of backbone amide groups, to study the hydrogen exchange of hydroxyl groups.[20] In the case of Tyr hydroxyl groups, proton/deuterium substitution slightly alters the chemical shifts of the carbon atom at the ζ position in the Tyr ring by ~0.13 ppm due to the proton/deuterium isotope shift effect. In a 1:1 mixture of H_2O and D_2O, the $^{13}C_\zeta$(–OH) and $^{13}C_\zeta$(–OD) peaks are resolved on a carbon spectrum when the residence time of the corresponding hydroxyl proton is longer than the inverse of the deuterium isotope shift effect in hertz; otherwise, they appear as a single averaged peak. However, the observation of such a slight chemical shift is difficult in the case of a uniformly ^{13}C-labeled Tyr ring due to the broad linewidth of the carbon peaks arising from dipole interactions and scalar couplings. To overcome this problem, (2*S*,3*R*)-[β_2,$\varepsilon_{1,2}$-$^2H_3$;0,α,β,ζ-$^{13}C_4$;^{15}N]-Tyr, ζ-SAIL Tyr, was synthesized by SAIL technology. In this Tyr aromatic ring, only the carbon at the ζ position is selectively enriched by ^{13}C, and the ring protons at the ε positions are deuterated (**Figure 9**). This labeling pattern facilitates efficient magnetization transfer between the $^1H^\delta$ protons and the $^{13}C^\zeta$ atoms via three-bond scalar couplings, with a size of 7 or 8 Hz. Therefore, the assignment of the $^{13}C^\zeta$ atoms can readily be performed by the NOE correlation between $^1H^\beta$ and $^1H^\delta$ and subsequently by the HSQC correlations between the $^1H^\delta$ and $^{13}C^\zeta$ atoms. The linewidth of the $^{13}C^\zeta$ peak is sharp enough to observe the slight deuterium isotope shift effect from the hydroxyl group.

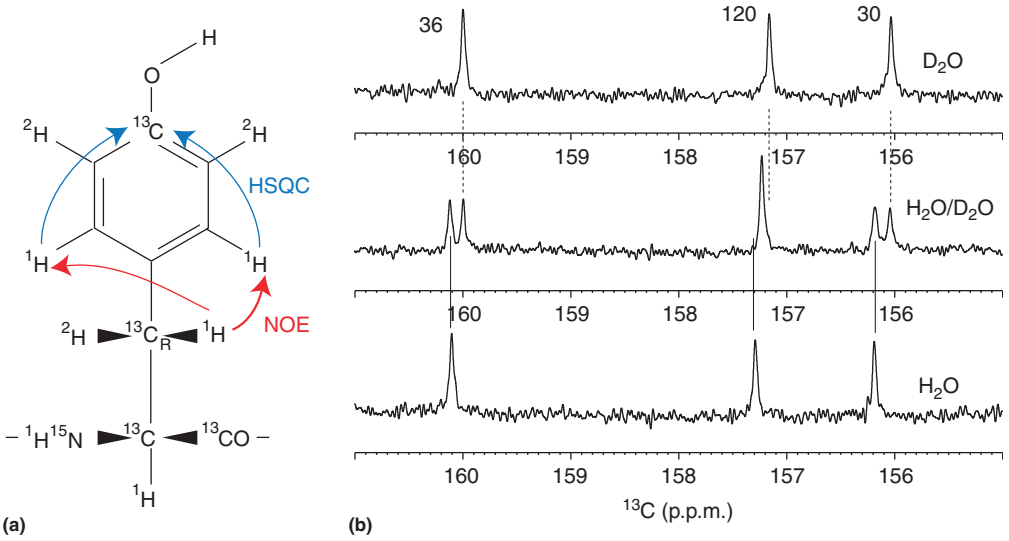

Figure 9 (a) Structure of ζ-SAIL Tyr and the magnetization pathway related to the assignment of the $^{13}C^{\zeta}$ atom. (b) Carbon spectra of EPPIb selectively labeled by ζ-SAIL Tyr in 100% H_2O, 50% H_2O/50% D_2O, and 100% D_2O solutions. These spectra were acquired by using a Bruker DRX600 spectrometer equipped with a TCI cryogenic probe at 40 °C.

In the application of the extended DEALS method to EPPIb, all three Tyr $^{13}C^{\zeta}$ peaks were observed as sharp lines, with a linewidth as small as 7 or 8 Hz at 40 °C. When the sample was dissolved in an H_2O/D_2O mixture, two of the three peaks were doubled, which shows that their hydrogen exchange rate is smaller than the size of the deuterium isotope shift effect (Figure 9). In addition, the exchange rate was quantitatively estimated by performing a ^{13}C-exchange experiment. This extended DEALS method is quite simple to perform and thus will be readily applicable to other cases.

1.12.6 Conclusion and Outlook

The stable isotope labeling technique is now widely used as an essential tool for biological macromolecular NMR spectroscopy. The isotope labeling approach has great potential to address the size problem of NMR, and the combined and sophisticated use of ^{2}H, ^{13}C, and ^{15}N isotopes promises to expand the applicability of NMR for investigating various biological and biophysical problems. The essential concept of the SAIL method is to optimize the isotope labeling patterns of amino acids, and thus the proteins labeled with such amino acids, for various steps of protein NMR spectroscopy, such as data acquisition, spectral analysis, structural determination, relaxation analysis and dynamics, and protein–protein and protein–ligand interactions. This chapter described only a small portion of the numerous possible applications to be explored in the future, for which only NMR spectroscopy can be used. Developments in NMR instrumentation, including superconducting magnets as high as 1 GHz and highly sensitive cryogenic probes for various applications, have made NMR spectroscopy an even more important technique to bridge protein structures and their biological functions. In this context, the further optimization of NMR samples based on the SAIL concept will be the most crucial issue for the NMR spectroscopy of biological macromolecules.

Acknowledgments

Financial support for the development of the SAIL technology was provided by the Core Research for Evolutional Science and Technology program of the Japan Science and Technology Corporation; the Targeted Protein Research Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan; and in part by a Grant-in-Aid for Young Scientists (B) (21770110).

References

[1] Markley, J. L.; Kainosho, M. Stable isotope labeling and resonance assignment in larger proteins. In *Stable Isotope Labeling and Resonance Assignments in Larger Proteins: NMR of Macromolecules*; Roberts, G. C. K., Ed.; Oxford University Press: New York, 1993; pp 101–152.
[2] LeMaster, D. M. *Prog. Nuclear Magnetic Resonance Spectroscopy* **1994**, *26*, 371–419.
[3] Kainosho, M. *Nature Struct. Biol.* **1997**, *4*, 854–857.
[4] Goto, N. K.; Kay, L. E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 585–592.
[5] Lian, L. Y.; Middleton, D. A. *Prog. NMR Spectrosc.* **2001**, *39*, 171–190.
[6] Ohki, S.; Kainosho, M. *Prog. NMR Spectrosc.* **2008**, *53*, 208–226.
[7] Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Ono, A. M.; Güntert, P. *Nature* **2006**, *440*, 52–57.
[8] Crespi, H. L.; Rosenberg, R. M.; Katz, J. J. *Science* **1968**, *161*, 795–796.
[9] Markley, J. L.; Putter, I.; Jardetzky, O. *Science* **1968**, *161*, 1249–1251.
[10] Kalbitzer, H. R.; Leberman, R.; Wittinghofer, A. *FEBS Lett.* **1985**, *180*, 40–42.
[11] LeMaster, D. M.; Richards, F. M. *Biochemistry* **1988**, *27*, 142–150.
[12] Torchia, D. A.; Sparks, S. W.; Bax, A. *J. Am. Chem. Soc.* **1988**, *110*, 2320–2321.
[13] Venters, R. A.; Farmer, B. T., 2nd; Fierke, C. A.; Spicer, L. D. *J. Mol. Biol.* **1996**, *264*, 1101–1116.
[14] Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357–406.
[15] Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J.; Rance, M. *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Academic Press: San Diego, 2006.
[16] Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1990**, *89*, 496–514.
[17] Ikura, M.; Kay, L. E.; Bax, A. *Biochemistry* **1990**, *29*, 4659–4667.
[18] Clore, G. M.; Gronenborn, A. M. *Methods Enzymol.* **1994**, *239*, 349–363.
[19] Kainosho, M.; Tsuji, T. *Biochemistry* **1982**, *21*, 6273–6279.

[20] Kainosho, M.; Nagao, H.; Tsuji, T. *Biochemistry* **1987**, *26*, 1068–1075.
[21] Uchida, K.; Markely, J. L.; Kainosho, M. *Biochemistry* **2005**, *44*, 11811–11820.
[22] Bermel, W.; Bertini, I.; Duma, L.; Felli, I. C.; Emsley, L.; Pierattelli, R.; Vasos, P. R. *Angew. Chem. Int. Ed.* **2005**, *44*, 3089–3092.
[23] Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R. *Angew. Chem. Int. Ed.* **2004**, *43*, 2257–2259.
[24] Pervushin, K.; Eletsky, A. *J. Biomol. NMR* **2003**, *25*, 147–152.
[25] Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655–11666.
[26] Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371.
[27] Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1998**, *120*, 6394–6400.
[28] Riek, R.; Wider, G.; Pervushin, K.; Wüthrich, K. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4918–4923.
[29] Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K. *Nature* **2002**, *418*, 207–211.
[30] Nietlispach, D.; Clowes, R. T.; Broadhurst, R. W.; Ito, Y.; Keeler, J.; Kelly, M.; Ashurst, J.; Oschkinat, H.; Domaille, P. J.; Laue, E. D. *J. Am. Chem. Soc.* **1996**, *118*, 407–415.
[31] Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627–636.
[32] Ollerenshaw, J. E.; Tugarinov, V.; Skrynnikov, N. R.; Kay, L. E. *J. Biomol. NMR* **2005**, *33*, 25–41.
[33] Güntert, P.; Braun, W.; Billeter, M.; Wüthrich, K. *J. Am. Chem. Soc.* **1989**, *111*, 3997–4004.
[34] Ikeya, T.; Terauchi, T.; Güntert, P.; Kainosho, M. *Magn. Reson. Chem.* **2006**, *44*, S152–S157.
[35] Oba, M.; Kobayashi, M.; Oikawa, F.; Nishiyama, K.; Kainosho, M. *J. Organic Chem.* **2001**, *66*, 5919–5922.
[36] Oba, M.; Terauchi, T.; Miyakawa, A.; Kamo, H.; Nishiyama, K. *Tetrahedron Lett.* **1998**, *39*, 1595–1598.
[37] Oba, M.; Terauchi, T.; Miyakawa, A.; Nishiyama, K. *Tetrahedron-Asymmetry* **1999**, *10*, 937–945.
[38] Terauchi, T.; Kobayashi, K.; Okuma, K.; Oba, M.; Nishiyama, K.; Kainosho, M. *Organic Lett.* **2008**, *10*, 2785–2787.
[39] Torizawa, T.; Ono, A. M.; Terauchi, T.; Kainosho, M. *J. Am. Chem. Soc.* **2005**, *127*, 12620–12626.
[40] Okuma, K.; Ono, A. M.; Tsuchiya, S.; Oba, M.; Nishiyama, K.; Kainosho, M.; Terauchi, T. *Tetrahedron Lett.* **2009**, *50*, 1482–1484.
[41] Kigawa, T.; Muto, Y.; Yokoyama, S. *J. Biomol. NMR* **1995**, *6*, 129–134.
[42] Zubay, G. *Annu. Rev. Genet.* **1973**, *7*, 267–287.
[43] Arata, Y.; Kato, K.; Takahashi, H.; Shimada, I. *Methods Enzymol.* **1994**, *239*, 440–464.
[44] McIntosh, L. P.; Dahlquist, F. W. *Q. Rev. Biophys.* **1990**, *23*, 1–38.
[45] Torizawa, T.; Shimizu, M.; Taoka, M.; Miyano, H.; Kainosho, M. *J. Biomol. NMR* **2004**, *30*, 311–325.
[46] Takeda, M.; Ikeya, T.; Güntert, P.; Kainosho, M. *Nature Prot.* **2007**, *2*, 2896–2902.
[47] Kigawa, T.; Yabuki, T.; Yoshida, Y.; Tsutsui, M.; Ito, Y.; Shibata, T.; Yokoyama, S. *FEBS Lett.* **1999**, *442*, 15–19.
[48] Kim, D. M.; Kigawa, T.; Choi, C. Y.; Yokoyama, S. *Eur. J. Biochem.* **1996**, *239*, 881–886.
[49] Kim, D. M.; Swartz, J. R. *Biotechnol. Prog.* **2000**, *16*, 385–390.
[50] Madin, K.; Sawasaki, T.; Ogasawara, T.; Endo, Y. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 559–564.
[51] Spirin, A. S.; Baranov, V. I.; Ryabova, L. A.; Ovodov, S. Y.; Alakhov, Y. B. *Science* **1988**, *242*, 1162–1164.
[52] Kramer, G.; Kudlicki, W.; Hardesty, B. *Cell-Free Coupled Transcription–Translation Systems from Escherichia coli*; Oxford University Press: New York, 1999.
[53] Campbell, I. D.; Dobson, C. M.; Moore, G. R.; Perkins, S. J.; Williams, R. J. *FEBS Lett.* **1976**, *70*, 96–100.
[54] Wagner, G.; DeMarco, A.; Wüthrich, K. *Biophys. Struct. Mech.* **1976**, *2*, 139–158.
[55] Edison, A. S.; Abildgaard, F.; Westler, W. M.; Mooberry, E. S.; Markley, J. L. *Methods Enzymol.* **1994**, *239*, 3–79.
[56] Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1993**, *115*, 11054–11055.
[57] Takeda, M.; Ono, A. M.; Terauchi, T.; Kainosho, M. *J. Biomol. NMR* **2010**, *46*, 45–49.
[58] Marion, D.; Kay, L. E.; Sparks, S. W.; Torchia, D. A.; Bax, A. *J. Am. Chem. Soc.* **1989**, *111*, 1515–1517.
[59] Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209–227.
[60] Güntert, P. *Prog. NMR Spectrosc.* **2003**, *43*, 105–125.
[61] López-Méndez, B.; Güntert, P. *J. Am. Chem. Soc.* **2006**, *128*, 13112–13122.
[62] Ikeya, T.; Takeda, M.; Yoshida, H.; Terauchi, T.; Jee, J.; Kainosho, M. *J. Biomol. NMR* **2009**, *44*, 261–272.
[63] Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177–1192.
[64] Chou, J. J.; Li, S. P.; Klee, C. B.; Bax, A. *Nature Struct. Biol.* **2001**, *8*, 990–997.
[65] Takeda, M.; Chang, C. K.; Ikeya, T.; Güntert, P.; Chang, Y. H.; Hsu, Y. L.; Huang, T. H.; Kainosho, M. *J. Mol. Biol.* **2008**, *380*, 608–622.
[66] Takeda, M.; Sugimori, N.; Torizawa, T.; Terauchi, T.; Ono, A. M.; Yagi, H.; Yamaguchi, Y.; Kato, K.; Ikeya, T.; Jee, J.; Güntert, P.; Aceti, D. J.; Markley, J. L.; Kainosho, M. *FEBS J.* **2008**, *275*, 5873–5884.
[67] Sharff, A. J.; Rodseth, L. E.; Quiocho, F. A. *Biochemistry* **1993**, *32*, 10553–10559.
[68] Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197–212.
[69] Kariya, E.; Ohki, S.; Hayano, T.; Kainosho, M. *J. Biomol. NMR* **2000**, *18*, 75–76.
[70] Edwards, K. J.; Ollis, D. L.; Dixon, N. E. *J. Mol. Biol.* **1997**, *271*, 258–265.
[71] Takeda, M.; Jee, J.; Ono, A. M.; Terauchi, T.; Kainosho, M. *J. Am. Chem. Soc.* **2009**, *131*, 18556–18562.