Quarterly Reviews of Biophysics **42**, 4 (2009), pp. 247–300. © Cambridge University Press 2010
doi:10.1017/S0033583510000016 Printed in the United States of America

SAIL – stereo-array isotope labeling

Masatsune Kainosho[1,2,3]* and Peter Güntert[1,3,4]

[1] Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo, Japan
[2] Graduate School of Science, Nagoya University, Nagoya, Japan
[3] Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University, Frankfurt am Main, Germany
[4] Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt am Main, Germany

Abstract. Optimal stereospecific and regiospecific labeling of proteins with stable isotopes enhances the nuclear magnetic resonance (NMR) method for the determination of the three-dimensional protein structures in solution. Stereo-array isotope labeling (SAIL) offers sharpened lines, spectral simplification without loss of information and the ability to rapidly collect and automatically evaluate the structural restraints required to solve a high-quality solution structure for proteins up to twice as large as before. This review gives an overview of stable isotope labeling methods for NMR spectroscopy with proteins and provides an in-depth treatment of the SAIL technology.

1. Introduction 248

2. Stable isotope labeling strategies 251
2.1 Historical development of stable isotope labeling with proteins 251
2.2 Early deuterium labeling 252
2.3 Random fractional ^{2}H labeling 252
2.4 Uniform ^{2}H labeling 253
2.5 Uniform ^{15}N labeling 253
2.6 Amino acid type-selective ^{13}C and ^{15}N labeling 254
2.7 Uniform ^{13}C and ^{15}N double labeling 254
2.8 ^{12}C-reverse labeling of aromatics in uniform ^{13}C and ^{15}N double labeling 255
2.9 ^{2}H random fractional labeling in uniform ^{13}C and ^{15}N double labeling 255
2.10 Combination of full ^{2}H labeling and uniform ^{13}C and/or ^{15}N labeling 256
2.11 Selective backbone protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling 257
2.12 Amino acid type selective protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling 257
2.13 Methyl group selective protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling 258
2.14 Stereoselective labeling of prochiral methyl groups in Val and Leu residues 259
2.15 Stereospecific ^{2}H labeling 260
2.16 Segmental labeling 260

3. SAIL principles 262
3.1 SAIL design principles 262
3.2 Expected numbers of identifiable NOESY peaks 262
3.3 Overlap and relaxation optimized SAIL patterns 263

* Author for correspondence: Prof. M. Kainosho, Center of Priority Areas, Graduate School of Science and Technology, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, Tokyo 192-0397, Japan.
Tel.: 81-42-677-4873; Fax: 81-42-677-4873; Email: kainosho@nmr.chem.metro-u.ac.jp

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

4. SAIL techniques 265
4.1 SAIL amino acid synthesis 265
4.1.1 Phenylalanine and tyrosine 265
4.1.2 Serine 267
4.1.3 Cysteine 268
4.1.4 Alanine 269
4.1.5 Glutamic acid 269
4.2 Cell-free protein synthesis 270
4.2.1 Validation of *E. coli* expression 271
4.2.2 Construction of cell-free protein synthesis system 272
4.2.3 Preparation of S30 extract 274
4.2.4 Preparation of SAIL proteins by the *E. coli* cell-free method 275
4.2.5 Validation of cell-free synthesis of the protein 276
4.2.6 Improvement of the yield of synthesized protein 276
4.2.7 N-terminal heterogeneity in cell-free protein synthesis 277
4.2.8 Large-scale protein production for NMR samples 277
4.3 SAIL NMR spectroscopy 278
4.3.1 General considerations 278
4.3.2 Assignment of aromatic ring signals 279

5. Automated structure analysis of SAIL proteins 283
5.1 Automated NOESY assignment and structure calculation with CYANA 283
5.2 Fully automated structure analysis with SAIL-FLYA 283

6. SAIL applications 286
6.1 Ubiquitin 286
6.2 Calmodulin 287
6.3 *E. coli* peptidyl-prolyl *cis-trans* isomerase b 287
6.4 C-terminal dimerization domain of SARS coronavirus nucleocapsid protein 289
6.5 Putative 32-kDa myrosinase-binding protein from *Arabidopsis* (At3g16450.1) 289
6.6 Maltodextrin-binding protein 289

7. Conclusions and Outlook 291

8. Acknowledgments 291

9. References 291

1. Introduction

Three-dimensional (3D) structures of proteins and nucleic acids provide insight into fundamental biological processes and play an important role in strategies for the rational design of drugs against human diseases. Methodologies for the determination of 3D protein structures thus play a central role in structural biology. Their development sets a pace for our understanding of life at the molecular level. When J. C. Kendrew and his colleagues beheld the first model of a 3D protein structure, they expressed their amazement in the words (Kendrew *et al.* 1958): "Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is much more complicated than has been predicted by any theory of protein structure."

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

After 50 years, much of this excitement remains. On March 31, 2009, there were 56 751 experimentally determined structures of biological macromolecules available from the Protein Data Bank (PDB), the central repository of 3D structural data of proteins and nucleic acids (Berman *et al.* 2000). This wealth of structures is the fruit of the efforts of a large number of curiosity-driven individual research groups, as well as recent large-scale 'structural genomics' projects that aim at supplementing the knowledge of the linear amino acid sequences of proteins by information on their 3D structure on a genome-wide scale (Chandonia & Brenner, 2006).

More than 99% of the 3D structures in the PDB were solved by X-ray crystallography or NMR spectroscopy, the two principal experimental methods for elucidating protein structures at atomic resolution. The number of new 3D structure depositions in the PDB has been rising nearly exponentially for many years (Fig. 1*a*). The distribution of X-ray or NMR protein structures as a function of the molecular mass (Fig. 1*b*) reveals the 'size problem' of NMR protein structure analysis: NMR has delivered, with few exceptions, only 3D structures of proteins smaller than 25 kDa. This is compounded by the observation from the Biological Magnetic Resonance Data Bank (BMRB) (Ulrich *et al.* 2008) that obtaining nearly complete chemical shift assignments, which are required to solve a high-quality structure by NMR, is difficult for larger proteins (Fig. 1*c*). The reason for this is that spectra of larger polypeptides are more complex, with more signals and consequently more overlap within the same limited range of frequencies. Furthermore, larger proteins tumble slowly in solution, which results in broad, eventually not interpretable peaks in the spectra.

In recent years, NMR spectroscopy has been applied to biomolecular systems of ever increasing size. Transverse relaxation optimized spectroscopy (TROSY) provides sharper lines for amide and aromatic groups in large proteins (Pervushin *et al.* 1997). Polarization transfer schemes based on cross-correlated relaxation were developed for solution NMR with large molecules (Riek *et al.* 1999). Backbone assignments and global folds of a small number of proteins larger than 30 kDa could be derived on the basis of conformational restraints for amide and selected methyl groups (Mueller *et al.* 2000; Tugarinov *et al.* 2005). However, assignments and conformational data for other aliphatic or aromatic groups remain difficult to collect for proteins larger than 25 kDa.

It has long been recognized that labeling with the stable isotopes ^{2}H, ^{13}C and ^{15}N can lead to simpler and more readily interpretable NMR spectra. The two principal strategies are to reduce the number of NMR-active nuclei in a molecule and to introduce additional NMR-active nuclei that resonate at different frequencies. Deuterium (^{2}H) is a spin 1 nucleus that is not detected by most NMR experiments. Replacing hydrogen ^{1}H nuclei by ^{2}H simplifies the spectra by eliminating the corresponding signals and relaxation pathways. On the other hand, the nuclear spin-1/2 isotopes ^{13}C and ^{15}N make the otherwise undetectable carbon and nitrogen atoms visible for NMR. Spectral overlap of the crucial hydrogen (^{1}H) signals can be resolved by spreading the peaks in additional dimensions indexed by the ^{13}C and ^{15}N resonance frequencies.

Many different approaches that rely on the labeling with stable isotopes have been devised and reviewed (Arrowsmith & Wu, 1998; Gardner & Kay, 1998; Goto & Kay, 2000; Kainosho, 1997; Lian & Middleton, 2001; Ohki & Kainosho, 2008). Several of these methods are in widespread use in biomolecular NMR and have been instrumental for extending the original homonuclear NMR structure determination method (Wüthrich, 1986) beyond its size limit of about 100 amino acid residues.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

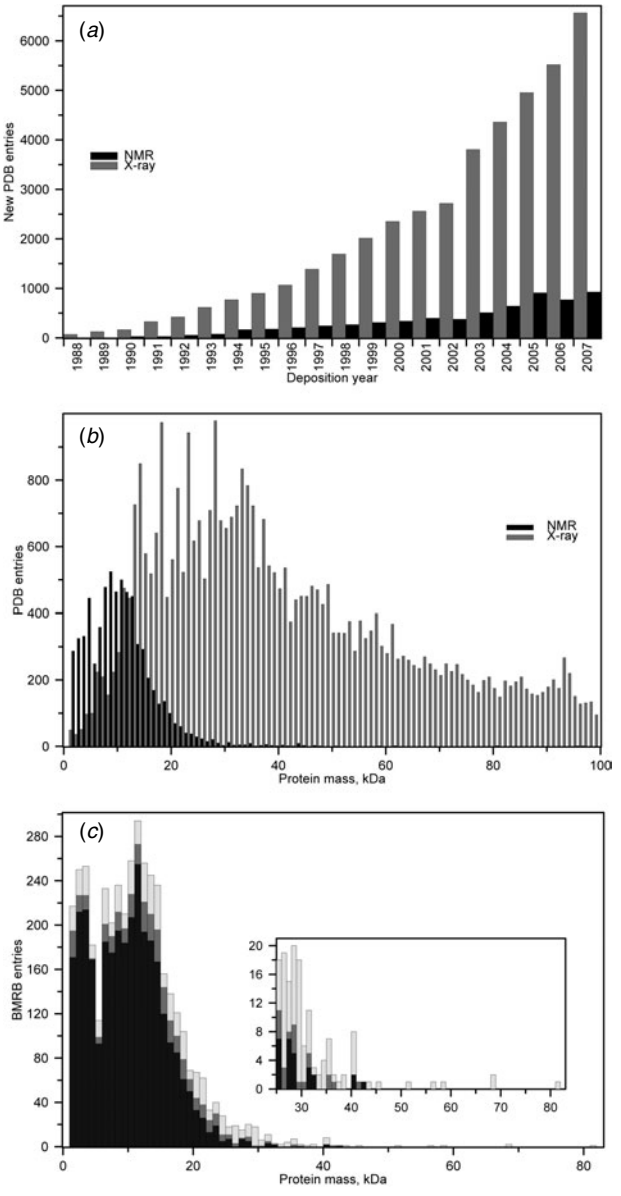


Fig. 1. (*a*) Annual depositions of X-ray and NMR structures to the PDB 1988–2007 (Berman *et al.* 2000). (*b*) Size distribution of X-ray and NMR structures in the PDB of January 2009. The molecular mass was calculated by summing up the mass of all protein atoms for which coordinates are provided. (*c*) Completeness of the assignments of backbone amide ^{1}H and aliphatic ^{1}H chemical shifts in the BMRB of January 2009 (Ulrich *et al.* 2008). Completeness of the chemical shift assignments 10%–30%, black bars; 30%–70%, dark gray bars; 70%–90%, medium gray bars; more than 90%, light grey bars. The molecular mass was calculated by summing up the mass of all residues in the amino acid sequence. In the case of symmetric multimers, one monomer is considered. The inset provides an enlarged representation of the data for proteins greater than 25 kDa.

Nevertheless, before the introduction of stereo-array isotope labeling (SAIL) labeling patterns and the approaches for achieving them have remained suboptimal. Only the SAIL method (Kainosho *et al.* 2006) achieved a complete stereospecific and regiospecific pattern of stable

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

isotopes, which is optimal with regard to the quality and information content of the resulting NMR spectra. This review summarizes earlier labeling schemes for protein NMR in section 2 and gives a detailed overview of the SAIL method in the sections 3–6.

2. Stable isotope labeling strategies

2.1 Historical development of stable isotope labeling with proteins

Even at the dawn of biological NMR spectroscopy, when only one-dimensional (1D) continuous wave experiments were possible, stable isotope labeling techniques were recognized as powerful tools to overcome the difficulties in analyzing the NMR spectra of proteins (Jardetzky & Roberts, 1981). However, even with the various labeling methods developed at the time, 1D NMR spectroscopy could not determine 3D structures of proteins. Rather, NMR was used to study local conformations, conformational changes, interactions, kinetics and dynamics of proteins. Although these experimental results obtained for proteins provided useful information about their functions and characteristics, corresponding 3D structures solved by X-ray crystallography were always needed for detailed discussions. In this sense, NMR was not an independent analytical tool in structural biology at that time. The two-dimensional (2D) NMR methods pioneered by Ernst's and Wüthrich's groups at ETH completely changed this situation (Ernst *et al.* 1987; Wüthrich, 1986). In the mid-1980s, a strategy to determine the 3D structures of proteins by ^{1}H homonuclear 2D NMR was established (Wüthrich, 1986) that promoted NMR as an independent method for 3D structure determinations of biomolecules. Several years later, ^{1}H-^{1}H 2D NMR was extended to 3D and 4D multidimensional experiments for which new stable isotope labeling techniques were a prerequisite. The new method used triple-resonance NMR experiments with ^{13}C- and ^{15}N-uniformly labeled proteins. This progress meant that larger proteins could be analyzed by NMR.

In the past 15 years, numerous new experimental methods in the NMR field, including sample preparation, were proposed. A set of triple resonance experiments facilitates the chemical shift assignment of all ^{1}H, ^{13}C and ^{15}N nuclei in proteins (Cavanagh *et al.* 2006). Transverse relaxation optimized spectroscopy (TROSY) enables the NMR analysis of biomolecules larger than 100 kDa and their complexes (Pervushin *et al.* 1997). Heteronuclear (^{2}H, ^{13}C or ^{15}N) relaxation studies provide information about the dynamics of biomolecules (Kay, 2005). Pulse sequences to measure *J*-coupling constants provide torsion angle information. The direct observation of hydrogen bonds through the measurement of the values of small *J*-coupling constants yields new insights into protein structures and their interactions (Grzesiek *et al.* 2004). Molecules can be weakly aligned to provide structural information in the form of residual dipolar coupling constants, which are independent of NOEs (Prestegard *et al.* 2004). Stable isotope labeling was indispensable for all these experiments. In parallel with such progress, NMR instrumentation has also greatly improved. Higher field magnets have been developed to provide better sensitivity and resolution. Today, 700–800 MHz instruments have become popular and are widely used, and 1 GHz instruments have become commercially available. Probe heads equipped with *xyz*-triple axis pulsed field gradient (PFG) coils achieve water suppression effectively and reduce the amount of phase cycling by choosing coherence transfer pathways in pulse sequences, resulting in shorter measurement times (Bax & Pochapsky, 1992). Cryogenic probes provide unprecedented sensitivity. These technological advances make it possible to study larger and more complex biological systems with NMR.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Several structural genomics and proteomics projects have been pursued around the world. As a result, NMR has become ever more important as a tool for studying protein structures, dynamics and interactions (Montelione *et al.* 2000; Staunton *et al.* 2003; Yee *et al.* 2003; Yokoyama *et al.* 2000). In these large-scale projects, high-throughput productivity is needed without compromising the structural quality. To satisfy these conflicting goals, stable isotope labeling methods and related technologies are continuously being developed further. In this section, we outline various stable isotope labeling techniques that have been developed before the introduction of the optimal isotope labeling by SAIL.

2.2 Early deuterium labeling

The resonance frequency of deuterium (^{2}H) is distinct from that of ^{1}H. Therefore, ^{2}H does not disturb ^{1}H NMR spectra. The replacement of ^{1}H by ^{2}H causes the ^{1}H signal to disappear from the ^{1}H NMR spectra. In this sense, deuterium labeling, or deuteration, is frequently called *negative* labeling. It is commonly used for labile protons that can in most cases be exchanged to ^{2}H simply by dissolving the protein in ^2H_2O instead of ^1H_2O. Also the ^{2}H labeling of non-labile protons was used early on as a powerful tool to simplify ^{1}H 1D NMR spectra (Crespi *et al.* 1968; Markley *et al.* 1968). In one of the first applications (Markley *et al.* 1968) with the protein *Staphylococcal* nuclease, only the amino acid types Trp, Met, $H^{\delta1}/H^{\delta2}$ of Tyr, $H^{\delta2}$ of His, $H^{\beta2}/H^{\beta3}$ of Asn/Asp and $H^{\gamma2}/H^{\gamma3}$ of Gln/Glu were kept protonated, while all others were deuterated. Compared with the fully protonated protein, this ^{2}H-labeled sample yielded strongly simplified 1D ^{1}H-NMR spectra. By monitoring the simplified NMR spectra, the effects of calcium ions and the inhibitor 3′,5′-thymidine diphosphate on the protein structure could be analyzed. However, because of the difficult sample preparation, this stable isotope labeling method did not become popular at the time.

2.3 Random fractional ^{2}H labeling

In addition to simplifying the spectra, ^{2}H has another advantage in ^{1}H 1D- and ^{1}H-^{1}H homonuclear 2D NMR experiments. Since ^{2}H has a significantly lower gyromagnetic ratio than ^{1}H, $\gamma(^2\mathrm{H}) \approx 0{\cdot}15\ \gamma(^1\mathrm{H})$, the presence of ^{2}H reduces the number of ^{1}H-^{1}H dipolar spin relaxation pathways and of ^{1}H-^{1}H scalar couplings. As a result, deuterium labeling increases the spin-spin relaxation time and thus can lead to narrower and more intense NMR signals. A simple strategy, introduced in the 1980s, is random fractional ^{2}H labeling by which 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 summed signal from the protein molecules in the NMR sample yields spectra that are nearly identical to those of the fully protonated sample, except that the peaks are sharper. The clearer ^{1}H 1D NMR spectrum of a randomly deuterated sample was reported for the 43-kDa *Escherichia coli* EF-Tu protein (Kalbitzer *et al.* 1985). In that case, about 90% of the ^{1}H sites were replaced by ^{2}H, dramatically enhancing the resolution in the ^{1}H 1D NMR spectrum without ^{2}H-decoupling. The utility of random ^{2}H labeling in ^{1}H-^{1}H homonuclear 2D NMR was also examined (LeMaster & Richards, 1988; Torchia *et al.* 1988). Optimal results were observed with deuteration levels of 50%–75%. LeMaster and Richards achieved the chemical shift assignment for *E. coli* thioredoxin using about 75% random fractional deuteration (LeMaster & Richards, 1988). This protein of 108 amino acid residues was then considered too large for analysis with ^{1}H-^{1}H homonuclear 2D NMR spectra. Their report

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

clearly showed how ^{2}H labeling could alleviate the molecular size limitation for NMR studies. A complication of random fractional deuteration is that the chemical shifts of the residual protons in random fractional deuterated proteins are slightly different due to the isotope shifts induced by deuteration, as described below.

2.4 Uniform ^{2}H labeling

Uniform ^{2}H labeling, or perdeuteration, was used for investigating protein-protein and protein-ligand complexes by ^{1}H 1D and 2D NMR. In the case of a 1:1 complex, one of the subunits in the complex is completely labeled by ^{2}H, and the other remains fully protonated. Thus, the ^{1}H NMR signals from the protonated subunit can be observed selectively. Comparisons of the NMR data of the protein in the free and bound state provided information about the conformational changes upon complex formation and the structure of a subunit in the complex. For instance, the conformation change of the peptide melittin induced by complex formation with the Ca^{2+}-binding protein calmodulin was monitored using perdeuterated calmodulin and unlabeled melittin (Seeholzer *et al.* 1986).

2.5 Uniform ^{15}N labeling

In the late 1980s, the 2D NMR methods were extended to 3D NMR in order to separate unresolved signals in ^{1}H-^{1}H 2D NMR spectra. To reduce overlap in the ^{1}H-^{1}H plane, the 2D spectra were separated into multiple planes along another indirect dimension. Initially, the new, third axis was frequency labeled with ^{1}H (Griesinger *et al.* 1987a, b; Oschkinat *et al.* 1989). Soon thereafter, Fesik and Zuiderweg realized the advantage of heteronuclear indirect detection for the third axis, using a ^{15}N-labeled tripeptide (Fesik & Zuiderweg, 1988). ^{15}N-edited (same as ^{15}N-separated) COSY, NOESY and TOCSY experiments were introduced (Marion *et al.* 1989) that provided much better peak separation than ^{1}H-^{1}H homonuclear 2D NMR. The concept for these 3D NMR pulse sequences was straightforward: a ^{1}H-^{1}H homonuclear 2D experiment was inserted into the detection period of a [^{1}H,^{15}N]-HMQC. Various modifications of the pulse sequences were introduced. HSQC was employed instead of HMQC. HSQC was followed by the ^{1}H-^{1}H 2D experiments. A sensitivity enhancement detection period was added for small and medium size proteins (Palmer *et al.* 1991b). Pulsed field gradients were included to select coherence transfer pathways and to achieve water suppression (Akke *et al.* 1994; Cavanagh *et al.* 1991; Kay *et al.* 1992a; Palmer *et al.* 1991a, 1992). The advantages and disadvantages of these variations have been discussed (Cavanagh *et al.* 2006).

Uniformly ^{15}N-labeled protein samples are crucial for the NMR based discovery of high-affinity ligands for proteins by the technique of 'SAR by NMR' (Shuker *et al.* 1996) in which small organic molecules that bind to proximal subsites of a protein are identified by the observation of ^{15}N- or ^{1}H-amide chemical shift changes in 2D [^{1}H,^{15}N]-HSQC spectra of the uniformly ^{15}N-labeled target protein, optimized and linked together to produce high-affinity ligands.

Today, uniformly ^{15}N-labeled samples play an important role in structural genomics. A [^{1}H,^{15}N]-HSQC spectrum is measured as an initial screening step for high-throughput structure determination. Since the quality of the [^{1}H,^{15}N]-HSQC spectrum depends on the protein structure, an excellent [^{1}H,^{15}N]-HSQC spectrum promises a successful structure determination (Yee *et al.* 2002). For proteins that show poor [^{1}H,^{15}N]-HSQC spectra the domain boundary selection may be revised, sample conditions may be optimized, or the structure determination may be deferred.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

2.6 Amino acid type-selective ^{13}C and ^{15}N labeling

Amino acid type-selective ^{13}C and ^{15}N double labeling has been used as a powerful tool for NMR signal assignments, the confirmation of signal assignments and the monitoring of specific residues, especially in large or membrane-bound proteins. Sequential resonance assignments are difficult in the simplified spectra recorded for selectively labeled proteins because of the gaps that correspond to the unlabeled residues. However, a straightforward signal assignment strategy was developed using amino acid type-selective double labeling (Kainosho & Tsuji, 1982). They assigned the $^{13}C'$ signals of the three Met residues that occurred in Met-Cys, Met-Val and Met-Asn dipeptides in the *Streptomyces* subtilisin inhibitor by preparing two samples with either [1-^{13}C]-Met and [^{15}N]-Cys, or [1-^{13}C]-Met and [^{15}N]-Val. This method can be applied for large proteins, is still used in current research and has found renewed interest for the study of membrane-bound proteins (Trbovic *et al.* 2005).

2.7 Uniform ^{13}C and ^{15}N double labeling

As an extension of ^{15}N labeling, Bax and co-workers prepared a protein in which all carbon and nitrogen atoms were labeled with ^{13}C and ^{15}N, respectively (Ikura *et al.* 1990; Kay *et al.* 1990). This is called uniform ^{13}C and ^{15}N double labeling. The presence of the ^{1}H, ^{13}C and ^{15}N spins means that almost all sites in the protein are NMR-visible. The heteronuclear one- or two-bond spin-spin couplings are relatively large in the proteins, which allows them to be used for efficient magnetization transfer (Cavanagh *et al.* 2006). Thus, their through-bond correlations can be monitored by multidimensional triple-resonance NMR experiments. Initially, the triple-resonance 3D NMR experiments HNCO, HNCA, HCACO and HN(CO)CA were proposed (Ikura *et al.* 1990). Subsequently, many additional triple-resonance pulse sequences were developed. Today, a set of spectra consisting of HNCA, CBCA(CO)NH, HNCO and HNCACB (or CBCANH) is routinely obtained to establish the sequential connectivities in the protein backbone, including C^{β} (Grzesiek & Bax, 1992a, b; Wittekind & Mueller, 1993). For the side-chain assignments, 3D-HCCH-TOCSY is frequently used (Bax *et al.* 1990; Olejniczak *et al.* 1992). In the HCCH-TOCSY pulse sequence, ^{13}C-^{13}C couplings are used for TOCSY magnetization transfer and the ^{13}C chemical shift is used to separate the signals in the ^{1}H-^{1}H plane. The development of the uniform ^{13}C and ^{15}N double labeling method and the corresponding triple resonance experiments increased the molecular size limit for NMR structure analysis to about 25 kDa.

Uniform ^{13}C and ^{15}N double labeling has also been adapted to study protein complexes. Several labeling approaches have been proposed for protein-peptide, protein-protein and protein-DNA/RNA complexes, as well as symmetric oligomers. The most popular method uses mixing labeled and unlabeled components. For homodimeric proteins, this method generates three kinds of species in the sample solution: labeled-unlabeled (50%), labeled-labeled (25%) and unlabeled-unlabeled (25%) pairs. It is hard to separate these species by biochemical methods, because they have almost identical properties except for the molecular mass. However, these species can readily be differentiated by NMR experiments. The binding surface can be identified by intermolecular NOEs observed with isotope-filtered NMR experiments (Folmer *et al.* 1995; Ikura & Bax, 1992; Lee *et al.* 1994; Melacini, 2000; Ogura *et al.* 1996; Otting & Wüthrich, 1990; Zangger *et al.* 2003; Zwahlen *et al.* 1997), even for the otherwise identical ^{1}H pairs. A similar strategy can be applied with larger, oligomeric protein complex systems. One subunit is uniformly labeled and the others are not. Then, the characteristics of the labeled component in the

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

complex can be studied. For 3D structure determinations of such complexes, computer algorithms have been developed for assigning intermolecular NOEs with consideration of the symmetry (O'Donoghue & Nilges, 1999; Yoshida *et al.* 2008).

2.8 ^{12}C-reverse labeling of aromatics in uniform ^{13}C and ^{15}N double labeling

The peaks of the Phe side-chain appear in a narrow chemical shift range in the ^{13}C dimension, and thus, the signal assignments of $^{13}C^{\delta}$, $^{13}C^{\varepsilon}$ and $^{13}C^{\zeta}$ are frequently difficult. However, the resonance assignments of the ^{13}C and ^{1}H signals are essential to analyze their local environments in protein structures. For this purpose, unlabeled Phe can be incorporated into a uniformly ^{13}C and ^{15}N doubly labeled protein. The strategy is called ^{12}C-reverse labeling. This labeling simplifies the aromatic region in ^{13}C-edited NMR spectra. With regard to the analysis of unlabeled Phe residues, isotope-filtered experiments can be employed. In the isotope-filtered experiments, the longer T_2 value of ^{1}H attached to ^{12}C, as compared with ^{1}H attached to ^{13}C, helps to improve both the sensitivity and resolution. In the application reported for the DNA-binding domain of *Drosophila* heat shock factor, dHSF(33-155), all of the Phe side-chain peaks were completely assigned (Vuister *et al.* 1994). Furthermore, long-range NOEs could be observed, thus allowing a discussion of the local conformation around the reverse-labeled Phe residues. In that report, the $^{3}J_{\alpha\beta}$ coupling constants were also analyzed quantitatively to estimate the χ^1 angles. Such structural restraints can be used as additional constraints together with the information obtained from the uniformly labeled samples. A modification of this method involving the use of Phe residues, in which the ε aromatic positions are ^{13}C-labeled, has also been reported (Wang *et al.* 1999). The same strategy can be applied to Tyr residues. The utility was demonstrated for the Dbl homology domain, with a molecular mass of 24 kDa (Aghazadeh *et al.* 1998), and a 25-kDa anti-apoptotic protein, Bcl-xL (Medek *et al.* 2000).

2.9 ^{2}H random fractional labeling in uniform ^{13}C and ^{15}N double labeling

Multidimensional NMR experiments for uniformly ^{13}C- and ^{15}N-double-labeled proteins with molecular weights larger than 25 kDa often yield poor NMR spectra. One of the problems is the widespread overlap of peaks. Since the number of atoms in proteins increases in parallel with the molecular weight, this is easily understood. The other problem is the severe line broadening caused by the shorter spin-spin relaxation time, T_2. The past decade has seen great progress in overcoming these problems. The breakthrough was the combination of ^{2}H labeling with the development of TROSY pulse sequences that select the narrow line components of ^{1}H-^{15}N or ^{1}H-^{13}C correlation peaks.

The TROSY pulse sequence, developed by Wüthrich's group, yields HSQC-like spectra, but the line width of each peak strongly depends on the external magnetic field strength. Theoretical considerations indicate that the [^{1}H,^{15}N]-TROSY peaks become sharpest at a ^{1}H frequency of about 1 GHz and that a better TROSY effect is seen for larger molecules (Pervushin *et al.* 1997, 1998; Wüthrich, 1998). Today, many triple resonance experiments performed on 800- to 950-MHz NMR spectrometers for larger proteins have been revised based on the TROSY technique (Salzmann *et al.* 1998).

On the other hand, the utility of ^{2}H had been recognized, as described in Section 2.2. The first application to examine the ability of ^{2}H labeling combined with triple resonance multidimensional NMR experiments was reported in the development of a pulse sequence, 4D-HN(COCA)NH, in 1993 (Grzesiek *et al.* 1993). Random fractional ^{2}H labeling at the level of

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

85%, together with uniform ^{13}C and ^{15}N labeling, was introduced into a protein sample, calcineurin B, with a molecular mass of 19·4 kDa. The ^{2}H labeling and decoupling successfully yielded sharp $^{13}C^{\alpha}$ signals. Deuteration narrowed the line width from about 25 to about 10 Hz. Thus, almost all of the sequential scalar coupling connectivities for the 19·4-kDa protein could be observed in the 4D-HN(COCA)NH spectra. Following the initial report, many modifications of triple resonance experiments were proposed, and various applications using this stable isotope labeling strategy have been reported (Garrett *et al.* 1997; Shan *et al.* 1996; Yamazaki *et al.* 1994). Together with this progress, the molecular weights of proteins for which chemical shift assignments can be carried out have increased considerably. A study by Wüthrich's group revealed that the ^{2}H, ^{13}C and ^{15}N-triple labeling method, combined with TROSY-based NMR analysis, can be feasible even for a 900-kDa protein complex (Fiaux *et al.* 2002).

In earlier applications of triple-resonance multidimensional NMR experiments with ^{2}H-decoupling, uniformly random fractional incorporation of ^{2}H into the non-exchangeable ^{1}H sites in proteins was employed. In general, the degree of the ^{2}H labeling affects the quality of the NMR spectra, and thus, higher levels of random uniform ^{2}H labeling give sharper signals. Therefore, highly deuterated proteins are believed to be better for main chain assignments. However, the lack of ^{1}H at non-exchangeable sites is a disadvantage for side-chain analysis, especially for observations of ^{1}H-^{1}H NOE for structure determination. Therefore, 50%–90% uniformly random fractional ^{2}H labeling is often employed for analysis, but this introduces considerable problems related to isotopomers. For example, the methyl groups of Ala, Leu, Ile, Val and Met contain isotopomers, i.e., CH_3, CH_2D, CHD_2 and CD_3. Three of the four isotopomers give NMR signals, and they appear at slightly different chemical shifts, due to the isotope shift. As a consequence, the methyl region shows crowded signals, thus preventing an extensive analysis. However, the methyl group is an interesting probe to monitor protein dynamics. An example is described in Section 2.13. In some cases, the single D in a CH_2D methyl group can also be an attractive target (Kay, 2005). Although an optimized sample preparation method and specific pulse sequences to observe one isotopomer in the sample solution have been reported (Ishima *et al.* 2001), the signal intensity is relatively low, because the actual number of detectable molecules is much less than the total sample concentration.

Recent improvements in sensitivity, using cryogenic probes, have enabled ^{13}C-detection in multidimensional protein NMR spectroscopy (Bermel *et al.* 2005; Bertini *et al.* 2004; Pervushin & Eletsky, 2003). Several experiments based on protonless ^{13}C direct detection have been proposed for heteronuclear assignment, in order to study large proteins, paramagnetic proteins, and natively unfolded proteins (Arnesano *et al.* 2003; Machonkin *et al.* 2002). In contrast to the analogous ^{1}H-detection experiments, the signal dispersion in the ^{13}C chemical shift is an advantage for the analysis of such protein samples. 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 NOEs, using fully deuterated and [^{13}C,^{15}N]-labeled proteins.

2.10 Combination of full ^{2}H labeling and uniform ^{13}C and/or ^{15}N labeling

As described in Section 2.2, a fully deuterated protein can be used to study protein-ligand, protein-protein and protein-membrane complexes. As an extension, ^{15}N was employed to identify the binding interface. Shimada and co-workers developed a new method, named transferred cross-saturation (Takahashi *et al.* 2000). In this method, one molecule of a heterodimeric complex is unlabeled and the other is labeled with ^{2}H and ^{15}N. An RF-pulse excites the aliphatic

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

^{1}H of the unlabeled molecule, and the magnetization is transferred to the neighboring amide ^{15}N in the labeled molecule. Amide ^{1}H-^{15}N correlation signals can then be detected for part of the molecule, suggesting that these amides are close to aliphatic ^{1}H, i.e., located in the interaction site. This technique is a good example of the benefit of combining ^{2}H labeling and a TROSY-based experiment.

2.11 Selective backbone protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling

Protocols for site-specifically protonated proteins with uniform stable isotope labeling by ^{2}H, ^{13}C and ^{15}N have been proposed for backbone resonance assignments (Yamazaki *et al.* 1997). Yamazaki *et al.* developed a biosynthetic approach to prepare proteins with high levels of ^{2}H labeling at all but the $^{13}C^{\alpha}$ positions. They used a commercially available ^{2}H, ^{13}C and ^{15}N labeled amino acid mixture, and the $^{1}H^{\alpha}$ labeling was achieved using a chemical reaction. The amino acid mixture was then added to the H_2O-based culture medium to express the target protein, the C-terminal domain of the α subunit of *E. coli* RNA polymerase, in an *E. coli* system. Furthermore, new pulse sequences were designed for this sample by using $^{13}C^{\alpha}$ magnetization to obtain the correlation through scalar couplings, such as $^{13}C^{\alpha}$-^{15}N and $^{13}C^{\alpha}$-$^{13}C'$. This approach is especially helpful to obtain high quality triple resonance NMR spectra, when the protein sample shows fast amide-proton exchange rates. Yamazaki *et al.* carried out the analysis on NMR data recorded at 10 °C. As judged from the T_2 values of $^{13}C^{\alpha}$, the overall tumbling rate of their sample at 10 °C was estimated to be similar to that of a 30- to 40-kDa protein at 30–40 °C. Thus, this method also has the potential to facilitate the analysis of large proteins.

Another backbone-labeling strategy was tested for the Phe, Val and Leu residues in a 26-kDa glycoprotein, the β subunit of human chorionic gonadotropin (Coughlin *et al.* 1999). In this case, chemically synthesized Phe, Val and Leu were used. Their backbone atoms were uniformly labeled with ^{13}C and ^{15}N. In addition, all of the proton sites of the amino acids were randomly replaced by ^{2}H, at a 50% level. These labeled Phe, Val and Leu amino acids and the other unlabeled amino acids were added to the growth medium for the sample preparation. The protein expression was achieved in mammalian cells. The use of such a protein, in which only the backbone atoms are isotope-labeled with ^{13}C and ^{15}N, results in the elimination of the undesirable one bond $^{13}C^{\alpha}$-$^{13}C^{\beta}$ scalar coupling and improves both the sensitivity and resolution of the triple-resonance NMR experiments. This stable isotope labeling strategy has the benefits of generating optimal resolution and increasing the sensitivity for all residue types. Moreover, the absence of $^{13}C^{\beta}$ provides an opportunity to measure the residual dipolar coupling constants between $^{13}C'$-$^{13}C^{\alpha}$ and $^{13}C^{\alpha}$-^{1}H (Ding & Gronenborn, 2004), which can be included in the restraints for structure calculation.

Alternative approaches using uniformly ^{13}C- and ^{15}N-labeled proteins have been proposed to remove the splitting due to the one bond $^{13}C^{\alpha}$-$^{13}C^{\beta}$ coupling, such as constant-time experiments (Santoro & King, 1992; Shan *et al.* 1996) and band-selective $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ decoupling experiments (Matsuo *et al.* 1996a, b). However, the former causes a significant loss in sensitivity. The latter does not achieve complete $^{13}C^{\beta}$ decoupling, because the $^{13}C^{\beta}$ signals of Ser and Thr are shifted downfield, as compared with others.

2.12 Amino acid type selective protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling

To reduce the spectral complexity, alternative labeling methods were proposed, using ^{2}H, ^{13}C and ^{15}N. For example, a ^{2}H, ^{13}C and ^{15}N-triple labeled protein, in which only the sites to be

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

monitored are protonated, was prepared. Amino acid type selective protonation is one of the methods in this category. In one such labeling strategy, the protein sample was expressed in minimal medium containing 95% D_2O, ^{2}H-labeled glucose, $(^{15}NH_4)_2SO_4$ and ^{1}H/^{13}C/^{15}N-labeled Ile, Leu, Val amino acids (Metzler *et al.* 1996). Using the sample prepared with this method, high-quality [^{1}H,^{13}C]-HSQC and NOESY spectra were observed for the hydrophobic residues. The aliphatic-aliphatic NOEs can also be observed in ^{13}C-edited NOESY spectra. These NOEs, combined with the $^1H^N$-$^1H^N$ NOEs obtained by ^{15}N-edited NOESY, can be used as input for a structure calculation. In a similar approach for the ras p21 protein, with a molecular mass of 19 kDa, protonated Ile, Leu, Val, Phe and Tyr were used (Smith *et al.* 1996). The advantages of these methods were discussed, but the limited distance information yielded only global fold information for the sample proteins. It is difficult to separate all the methyl peaks in large proteins. Aromatic peaks are also prone to spectral overlapping because the Tyr and Phe side-chains appear in a narrow chemical shift range in the ^{13}C dimension, as described in Section 2.8.

Recently, Ito *et al.* reported a new sample preparation method for the *E. coli* biosynthesis of deuterated proteins with selective protonation of the aromatic rings of Phe, Tyr and Trp (Rajesh *et al.* 2003). Specific protonation of the aromatic rings was achieved with complete deuteration at the $^{13}C^{\alpha/\beta}$ positions. The labeling was accomplished by using shikimate-supplemented D_2O medium. Co-expression of a shikimate transporter in prototrophic bacteria resulted in protonation levels of 62%–79%. The NMR spectra for yeast ubiquitin hydrolase 1, composed of 236 residues, were observed to demonstrate the method. The purpose of this labeling is to obtain the NOE distance restraints for the aromatic rings.

2.13 Methyl group selective protonation in uniform ^{2}H, ^{13}C and ^{15}N labeling

Kay and co-workers developed a method for selective methyl group protonation (Gardner & Kay, 1997; Goto *et al.* 1999; Rosen *et al.* 1996). A stable isotope labeled supplement added to the *E. coli* growth medium facilitates the selective labeling. The labeled chemical compound acts as a precursor of Ala, Val, Leu and Ile. The proteins prepared by this method are uniformly ^{2}H-, ^{13}C- and ^{15}N-labeled, except for the methyl groups of Ala, Val and Leu and the γ^2 methyl site of Ile that remain highly protonated. Protonation levels of about 80% for Leu, 60% for Ile (γ^2) and Val, and 40% for Ala were reported. The NMR spectra revealed some protonation at the C^β positions of Asp/Asn and the C^γ positions of Gln/Glu, Arg and Pro. An improved protocol by Kay *et al.* labels only the Leu δ, Val γ and Ile δ^1 methyl groups. In contrast to the previous method, a high labeling level of about 90% was achieved, and isotopomers, i.e., $^{13}CH_2D$ and $^{13}CHD_2$, did not appear.

The high level of deuteration in uniformly ^{13}C and ^{15}N-labeled proteins makes it easy to assign the backbone ^{1}H, ^{15}N, $^{13}C'$ and $^{13}C^\alpha$ signals and the side-chain $^{13}C^\beta$ signals of the 370-residue protein MBP (41 kDa) (Gardner *et al.* 1998). Kay's group also demonstrated that assignments can be obtained for a 723-residue (82 kDa) single polypeptide multidomain protein, malate synthase G (Tugarinov *et al.* 2005). Furthermore, they obtained $^1H^N$-$^1H^N$, $^1H^N$-methyl and methyl-methyl NOEs, backbone dihedral angle restraints based on ^{15}N, $^{13}C^\alpha$ and $^{13}C'$ chemical shift values, ^{1}H-^{15}N residual dipolar coupling constants and chemical shift anisotropy (CSA) values. On the basis of these data, they calculated the 3D structure of malate synthase G. The same labeling strategy enabled the structure determination of the integral membrane protein OmpX from *E. coli* reconstituted in 60-kDa DHPC micelles (Fernández *et al.* 2004).

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

The precision and the accuracy of protein structures solved by this method depend on the number of protonated methyl groups and their spatial distribution in the protein structure. The 3D structures generated by this method may be less precise than X-ray crystal structures, and NMR can at present not compete with X-ray crystallography for structure determinations of large proteins. However, the potential of NMR for the structure analysis of large proteins in solution was clearly indicated.

Three labeling strategies for side-chain methyl groups in high molecular weight proteins using $^{13}CH_3$, $^{13}CH_2D$ and $^{13}CHD_2$ methyl isotopomers were compared (Ollerenshaw *et al.* 2005). For each labeling scheme, ^{1}H-^{13}C correlation pulse sequences that give optimal resolution and sensitivity were identified and tested experimentally on three highly deuterated samples of malate synthase G with $^{13}CH_3$, $^{13}CH_2D$ and $^{13}CHD_2$ labeling in the Ile $\delta 1$ positions. Although excellent resolution can be obtained with any of the isotopomers, the $^{13}CH_3$ labeling strategy benefitted from a more than twice higher sensitivity than either $^{13}CH_2D$ or $^{13}CHD_2$ labeling.

Recently, this methyl labeling method was applied to the 20S proteasome, with a molecular mass of 670 kDa. Kay's group observed clear NMR spectra and analyzed the dynamics of the 20S proteasome (Sprangers & Kay, 2007).

2.14 Stereoselective labeling of prochiral methyl groups in Val and Leu residues

The stereospecific assignment of prochiral methyl groups in Val and Leu residues is important to determine protein structures with high accuracy. Normally, the NMR signals of the two isopropyl methyl groups have different chemical shifts, as a consequence of their different magnetic environments. For a detailed discussion of their local conformation, stereospecific assignments are essential. Furthermore, when stereospecific assignments are unknown pseudoatom corrections should be applied to the NOE upper distance bounds from these groups during the structure calculation. This tends to result in less well-defined structures. Stereospecific assignments of the prochiral methyl groups avoid this loss of the structural quality.

The first approach used to assign the diastereotopic methyl groups in Val and Leu was carried out by biosynthetic fractional ^{13}C labeling (Neri *et al.* 1989). This method was based on the known stereoselective biosynthetic pathways for these amino acids from glucose. The isopropyl group is composed of a two-carbon fragment from one pyruvate unit, while the second methyl group is transferred from another pyruvate unit. This reaction occurs stereoselectively, and the migrating methyl group becomes *pro*-S in both Val and Leu; i.e., they are γ^2CH_3 and δ^2CH_3, respectively. Fractional ^{13}C labeling of the methyl groups occurs when a mixture of ^{13}C-glucose and ^{12}C-glucose is used as the carbon source. In the ^{1}H-^{13}C correlation spectra, the ^{13}C-attached proton shows a $^{1}J_{CH}$ splitting, thus clearly indicating the stereospecific assignment.

In a similar way, the simultaneous unlabeling of selected amino acid residues and the fractional ^{13}C labeling of the remaining amino acids was achieved by growing the host microorganism in minimal medium containing a mixture of 10% [$^{13}C_6$]-glucose and 90% [$^{12}C_6$]-glucose, along with the unlabeled forms of the amino acids that were to remain unlabeled (Atreya & Chary, 2001). Since the methyl region in ^{1}H-^{13}C 2D spectra is crowded, this method can simplify the NMR spectra and enable the stereospecific assignment of Val and Leu methyl side-chains.

Block labeling is an alternative to the stereospecific assignment of the prochiral methyl groups in Val and Leu residues. First, Val and Leu amino acids, in which one of the methyl carbons is stereoselectively labeled with ^{13}C and the other is not, are either synthesized chemically or obtained by microbial fermentation. Next, these amino acids are incorporated into proteins. The

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

stereochemistry of the amino acids is retained in these labeled residues, and constant-time [^{1}H,^{13}C]-HSQC spectra clearly indicate the chirality. The utility of this simple strategy was shown by the analysis of cystatin A (Tate *et al.* 1995), and the signal assignment of a 10-kDa functional domain of phosphoprotein CPI-17, which has 13 Leu and 6 Val residues, was accomplished by this method (Ohki *et al.* 2001).

2.15 Stereospecific ^{2}H labeling

In a similar manner as the block labeling described above, a method for the stereospecific assignment of the methyl groups in Val and Leu residues was achieved with ^{2}H labeling (Ostler *et al.* 1993). In this method, the stereoselectively ^{2}H-labeled compound was supplied in the medium to obtain the protein by bacterial synthesis. In the first application (2*S*,4*R*)[5,5,5-$^{2}H_3$] leucine was added to the medium to express a protein by *Lactobacillus casei*. The absence of the signals in the ^{1}H-^{1}H homonuclear 2D NMR spectra clearly showed the stereospecific assignments.

For the stereospecific assignment of the two H^{α} protons in Gly, LeMaster *et al.* developed a stereospecific ^{2}H-labeling method (LeMaster *et al.* 1994). They prepared a protein, *Staphylococcal* nuclease H124L, in which the Gly residues were labeled by [$^{2}H_R$,2-^{13}C]Gly and [$^{2}H_S$,2-^{13}C]Gly. Isotope shifts made the chemical shift of each Gly $^{13}C^{\alpha}$ different for the two labeling patterns, and two peaks were observed in [^{1}H,^{13}C]-HSQC spectra. Furthermore, each peak splits into two peaks in the ^{1}H dimension when ^{13}C was not decoupled, and the ^{1}J scalar coupling constants are slightly different. Based on these values, the conformation of Gly was discussed. In general, HNHA or HMQC-*J*-type experiments using uniformly ^{15}N-labeled samples cannot determine the dihedral angles for Gly residues, and thus, this labeling provides some advantage for studying the local conformation around these residues.

2.16 Segmental labeling

Segmental labeling is an attractive method for preparing NMR samples. The uniqueness of this method is that one or several parts of the sample protein can be labeled. Large proteins often consist of multiple domains. This stable isotope labeling method is useful for studying the structure, dynamics and interactions of a domain of interest in a large, intact protein. Segmental labeling is based on peptide splicing reactions (Duan *et al.* 1997; Klabunde *et al.* 1998) carried out by inteins. This is a post-translational chemical modification (Hirata *et al.* 1990; Kane *et al.* 1990).

For the first demonstration of segmental labeling, Yamazaki *et al.* employed an intein, PI-*Pfu*I, from *Pyrococcus furiosus* (Otomo *et al.* 1999b; Yamazaki *et al.* 1998). The key idea of their labeling strategy was that PI-*Pfu*I can be split into two fragments such that their mixture still maintains the function of the intact intein. Exploiting this interesting property, they prepared two kinds of protein-precursor constructs separately. One was the N-terminal half of the sample protein conjugated to the N-terminal half of the intein. The other was the C-terminal half of the intein with the C-terminal half of the sample protein. One of them was expressed in *E. coli* grown in uniformly ^{15}N-labeled medium, and the other was prepared in unlabeled culture medium. Both fragments were expressed as inclusion bodies. The samples were mixed in denaturing reagents and refolded to generate the active intein, which could accomplish the splicing reaction to ligate the target polypeptides. The sample was subjected to heat shock at 70 °C to initiate the protein splicing reaction. After the splicing reaction, they obtained the segmentally labeled RNA

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

polymerase subunit α. They recorded the [^{1}H,^{15}N]-HSQC spectra of three kinds of samples, i.e., N-terminal half labeled, C-terminal half labeled and fully labeled. The sum of the [^{1}H,^{15}N]-HSQC spectra for the segmental labeled samples was identical to that of the intact protein, except for the inserted residues. Since the splicing reaction inserted five extra amino acid residues into the ligation site, the junction site should be selected carefully to retain the entire protein structure and the full biological activity. Suitable ligation sites are usually found in flexible loop regions on the protein surface or in interdomain linkers, where they can be identified by enzymatic digestion experiments. In an extension of the original method two kinds of inteins were employed to demonstrate the stable isotope labeling of a central region of a protein (Otomo *et al.* 1999a). Maltose-binding protein (MBP) with an isotope-labeled, central segment flanked by two unlabeled elements was successfully prepared with about 90% yield for the intein-mediated ligation. Segmental labeling was also applied to the 52-kDa F1-ATPase β subunit motor protein (Kobayashi *et al.* 2008; Yagi *et al.* 2004). Four kinds of segmentally labeled samples were prepared and analyzed by NMR, yielding about 90% of the backbone NH, $^{13}C^{\alpha}$, $^{13}C'$ and $^{13}C^{\beta}$ signals of this 473 residue protein. This segmental labeling strategy is a powerful tool for analyzing domains in a large protein, although the heating step for splicing and the insertion of extra amino acids may limit its application.

Using a similar concept, a versatile segmental labeling method was proposed that has the advantages that the splicing reaction occurs under gentle conditions *in vitro* and does not introduce extra residues at the ligation site (Iwai & Plückthun, 1999; Xu *et al.* 1999). The N-terminal fragment of the sample protein is expressed with an intein plus chitin-binding domain at its C-terminal end. The C-terminal fragment, in which the N-terminus is a Cys residue, was also prepared. One of them was isotopically labeled and the other was not. The N-terminal fusion protein was trappped in an affinity column via the chitin-binding domain, and all the splicing reactions were carried out in the column. Finally, the segmentally labeled product was eluted from the column. This method was tested to connect two independently folded domains, SH2 and SH3, from Abelson protein tyrosine kinase. A yield of 70% was reported for the reaction.

Recently, Züger and Iwai developed an elegant technique for the intein-based biosynthetic incorporation of unlabeled polypeptide segments into isotopically labeled proteins (Iwai *et al.* 2006; Züger & Iwai, 2005). In their method, protein splicing is automatically performed *in vivo*. The preparation of the precursor fragments and the ligation experiments *in vitro* are not necessary, and the purified protein can be subjected to NMR experiments without removal of the tag. This is a great advantage for sample preparation. A dual expression system allows the sequential expression of two precursor products. The key idea of this strategy is that the two fragments are expressed in a single cell by independent plasmids with different promoters: one is controlled by IPTG and the other is controlled by L-arabinose. Thus, the time when a given fragment was overexpressed in the cell could be controlled. First, IPTG (or L-arabinose) is used to start the expression of one of the precursors in the labeled (or unlabeled) medium. Then, the cells were collected and transferred to the unlabeled (or labeled) medium, where the expression of the other fragment was initiated by the addition of L-arabinose (or IPTG). When the two kinds of fragments were expressed in cells, the protein splicing reaction occurred automatically. Using this technique, isotope-labeled proteins fused with an unlabeled chitin-binding domain at the C-terminal end were prepared (Züger & Iwai, 2005). The chitin-binding domain is frequently employed to simplify protein purification steps, but it also has the function to stabilize the fused proteins. Thus, their method provides an opportunity for the successful preparation and NMR study of unstable proteins, such as natively unfolded samples.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Recently, the segmental isotopic labeling of a central domain in a multidomain protein was achieved by protein *trans*-splicing using only the well characterized *Npu*DnaE intein for three-fragment ligation without any refolding steps (Busche *et al.* 2009). This simple and robust scheme for labeling a central fragment of a protein could become a valuable tool for elucidating structure-function relationships of a domain or a region in an intact protein by NMR spectroscopy.

Presently, more than 100 inteins have been discovered. Only few of them have yet been employed for NMR sample preparation. The use of inteins is an active field of research and an increasingly important technique for N- and/or C-terminal modification, protein stabilization, purification and ligation (Brenzel *et al.* 2006; Burz *et al.* 2006; Cowburn *et al.* 2004; Iwai & Plückthun, 1999; Iwai *et al.* 2006; Muona *et al.* 2008; Romanelli *et al.* 2004; Skrisovska & Allain, 2008; Vitali *et al.* 2006; Williams *et al.* 2005; Williams *et al.* 2002; Zhao *et al.* 2008; Züger & Iwai, 2005).

3. SAIL principles

3.1 SAIL design principles

The basic strategy of SAIL is to incorporate deuterium into the amino acids constituting a protein such that each carbon or nitrogen nucleus in the final protein will have at most one ^{1}H nucleus bonded to it, the remaining hydrogen atoms having been replaced by ^{2}H. 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. This is achieved by preparing amino acids with the following features (Fig. 2):

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

SAIL versions of the 20 common amino acids (Fig. 3) were synthesized based on these design concepts by chemical and enzymatic syntheses as described in Section 4.1 below.

The patterns of replacement are designed to provide the most information consistent with spectral simplification and isotopic dilution and constitute a general approach to overcoming the limitations of deuteration. This labeling pattern preserves through-bond connectivity information needed for backbone and side-chain assignments, eliminates the need for stereospecific assignments, simplifies measurements of couplings and removes the most serious sources of spin diffusion so as to improve the accuracy of inter-proton distance measurements. Lines are sharpened both by decreasing long-range couplings and by eliminating dipolar relaxation pathways. The methyl and methylene labeling patterns simplify the analysis of the motional properties of the side-chains from relaxation measurements (Kay *et al.* 1992b). The aromatic ring labeling strategy removes one-bond aromatic ^{13}C–^{13}C couplings, which often complicate spectra.

3.2 Expected numbers of identifiable NOESY peaks

The SAIL technology improves NMR spectra because it decreases the number of signals without losing sensitivity for the resonances of interest. The number of non-exchangeable side-chain protons, which are prone to overlap but essential for defining the side-chain conformations, is

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

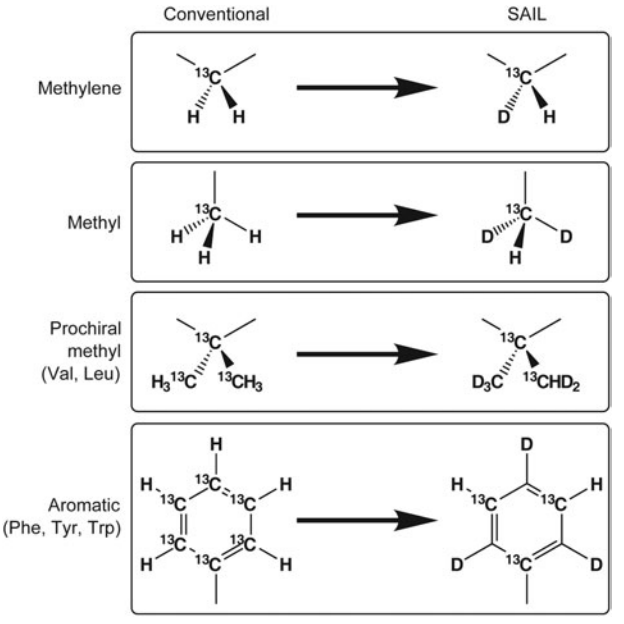


Fig. 2. Design concepts embodied in the SAIL amino acids (Kainosho *et al.* 2006).

reduced to less than half, and the number of expected NOESY cross-peaks is decreased by 40%–45%. Most of the additional NOEs from uniformly labeled proteins either involve fixed (geminal) distances or become redundant in the absence of stereospecific assignments and thus contribute to spectral overlap without furnishing independent information. In principle, stereospecific assignments could allow for slightly more precise structures in the case of uniform labeling (Güntert *et al.* 1989). However, only a limited number of stereospecific assignments can be made in practice. The BioMagResBank reports stereospecific assignments only for well below 10% of the methylene protons and Val/Leu methyl groups. Therefore, in general the expected number of non-redundant, structurally relevant NOE restraints is retained with SAIL, even if the effect of the better signal-to-noise ratio with SAIL is neglected. These numbers are corroborated by the experimental findings (see below).

More detailed theoretical considerations (Kainosho *et al.* 2006), which take into account the enhanced signal strength and sharper lines in SAIL spectra and the fact that overlap can render peaks unidentifiable, show 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, and thus for larger proteins, for which SAIL is expected to yield twice or more times the number of relevant conformational restraints than uniform labeling.

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. For this, it will presumably be necessary to optimize the isotope labeling patterns further, in order to cope with the extensive crowding and line-broadening that is characteristic of the spectra of such proteins. In an overlap and

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

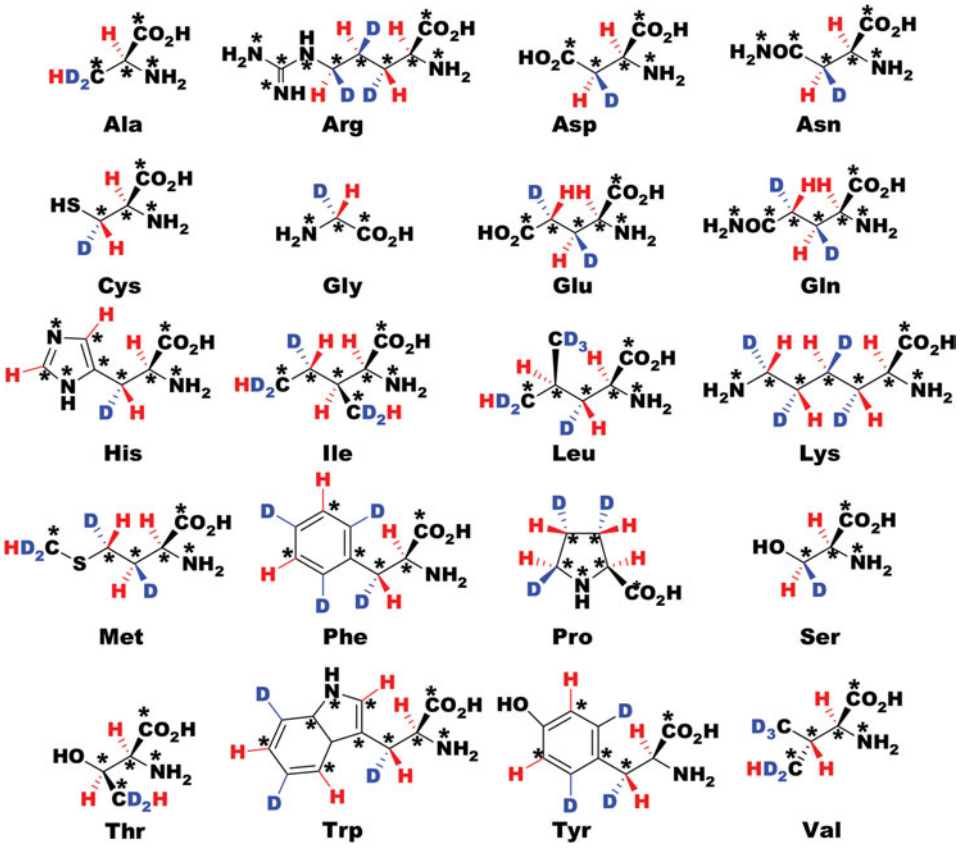


Fig. 3. Chemical structures of the SAIL amino acids (Kainosho *et al.* 2006).

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 beyond 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 method retains 64% of all protons, and 44% of all side-chain protons (Kainosho *et al.* 2006). An overlap and relaxation optimized SAIL pattern (Ikeya *et al.* 2006) comprised 53% of all protons, and only 28% of all side-chain protons. The modified SAIL ^{1}H labeling pattern is essentially a subset of standard SAIL. 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 (Fig. 3). 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 be replaced by ^{12}C for a 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. The only instance in which modified SAIL is not a subset of SAIL occurs with the aromatic ring of Phe, which has ^{1}H exclusively in the two ε-positions in SAIL, but only at the H^{ζ}-position in modified SAIL.

Simulations suggested that the combined use of optimized SAIL patterns and automated structure calculation algorithms with CYANA has high potential for the automated structure

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

determinations of higher molecular weight proteins as well as membrane proteins (Ikeya *et al.* 2006). The precision of the structure using the new, overlap and relaxation optimized modified SAIL pattern was slightly lower than that obtained by uniform labeling and the original SAIL method because modified SAIL reduces the number of peaks expected under ideal conditions. However, with the actual experimental data, better results can be expected for overlap and relaxation optimized SAIL because it facilitates the analysis of many otherwise broadened or overlapped signals. 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 size range and membrane proteins.

4. SAIL techniques

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 that are commercially available (Oba *et al.* 1998, 1999, 2001; Terauchi *et al.* 2008; Torizawa *et al.* 2005). The approach proved amenable to scale-up, enabling the preparation of significant quantities of these amino acids. Here, the syntheses of the SAIL amino acids Phe, Tyr, Ser, Cys and Ala are described as examples.

4.1.1 Phenylalanine and tyrosine

Stereoselective syntheses of phenylalanine labeled with deuterium on the H^{δ}, H^{ε} and H^{ξ} positions had been reported previously (Nishiyama *et al.* 1994). However, these synthetic routes were not suitable for the synthesis of these amino acids because of the difficulty in the regioselective incorporation of ^{13}C in the aromatic ring. Instead, SAIL phenylalanine and tyrosine can be synthesized efficiently using a hydroxyl benzoic acid derivative as key intermediate, starting from various types of ^{13}C-labeled acetone and ethyl malonate (Torizawa *et al.* 2005). The synthesis pathways are shown in Fig. 4.

To synthesize SAIL Tyr, 4*H*-pyran-4-one **4** was obtained from **3** by a modification of a reported procedure (Murray III & Williams, 1958; Riegel & Zwilgmeyer, 1943). Then, 4*H*-pyran-4-one **4** was subjected to a condensation/aromatization reaction to give the key intermediate **5** (Lang *et al.* 2002). For the synthesis of Tyr **1a** (Scheme 1), the hydroxyl group of **5** was methylated and the ester was reduced by $LiAlD_4$. The following oxidation by PDC yielded **6** (Beyer *et al.* 1998). Then, **6** was converted into the dehydrotyrosine derivative **7** by a condensation with labeled Gly derivatives (Erlenmeyer, 1893; Ojima *et al.* 1989). Compound **7** was subjected to hydrogenation in the presence of (*S*,*S*)-Et-DuPhos-Rh as a catalyst in MeOH (Burk *et al.* 1993) and was then deprotected to yield Tyr **1a** (Li *et al.* 1993).

Phe **1b** can be synthesized by the deoxygenation of the hydroxyl group of the key intermediate **5**. Thus, a reductive deuteration of the hydroxyl group of **5** was carried out, using recently reported methods to place a deuterium at the H_{ξ} position (Nishiyama *et al.* 1994; Viswanatha & Hruby, 1980). Then, as in the method for the synthesis of Tyr, **9** was converted into Phe **1b**.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

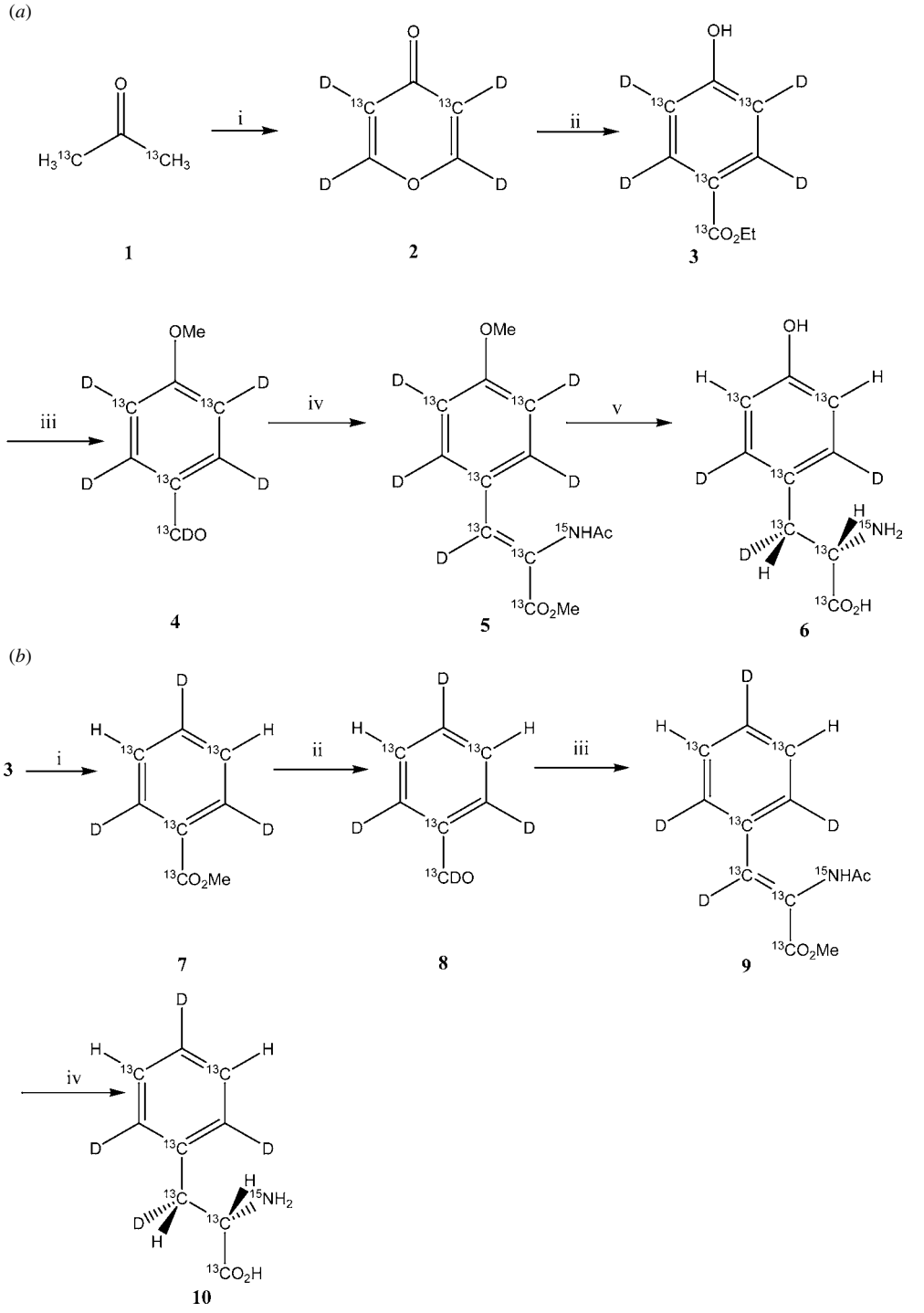


Fig. 4. Synthesis of aromatic SAIL amino acids (Torizawa *et al.* 2005). (*a*) SAIL tyrosine. Reagents and conditions: (i) 1, $(EtOCO)_2$, NaOEt; 2, conc. DCl; 3, Cu; (ii) 1, [$^{13}C_3$]-ethyl malonate, *t*-BuOK/BuOD; (iii) 1, MeI, acetone; 2, $LiAlD_4$; 3, PDC; (iv) 1, [$^{13}C_2$;^{15}N]-NAc-glycine, Ac_2O, AcONa; 2, MeOH, Et_3N; (v) 1, (*S*,*S*)-Et-DuPhos-Rh, H_2; 2, 1 N-HCl; 3, HBr-AcOH. (*b*) SAIL phenylalanine. Reagents and conditions: (i) 1, HCl, Δ; 2, Tet-Cl, K_2CO_3; 3, H_2, Pd/C; (ii) 1, $LiAlD_4$; 2, PDC; (iii) 1, [$^{13}C^2$;^{15}N]-NAc-Glycine, Ac_2O, AcONa; 2, MeOH, Et_3N; (iv) 1, (*S*,*S*)-Et-DuPhos-Rh, H_2; 2, N-HCl.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

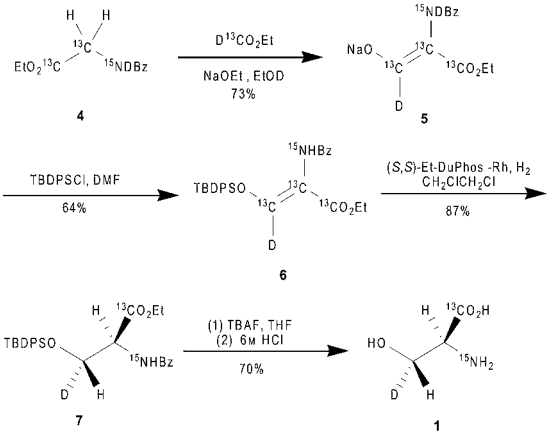


Fig. 5. Synthesis of SAIL serine (Terauchi *et al.* 2008).

4.1.2 Serine

There are many other examples of methods used for the synthesis of isotope-labeled serine, cysteine and alanine. The synthesis of a racemic mixture of serine labeled with deuterium at the β-position, via the reduction of the dehydroserine derivative, was reported more than 20 years ago (Aberhart & Russell, 1984; Kainosho & Ajisaka, 1975). Gani and co-workers used the Baeyer-Villiger oxidation as a key reaction for the enantioselective synthesis of deuterium-labeled serine (Gani & Young, 1983). However, the enantiopurity of the deuterated serine was insufficient for the SAIL method. Therefore, the synthesis of SAIL serine (Terauchi *et al.* 2008) uses catalytic asymmetric hydrogenation for the chiral introduction of a deuterium atom at the β-position, using the more easily accessible [1,2-$^{13}C_2$;^{15}N;2,2-$^{2}H_2$]glycine and ethyl [^{13}C,^{2}H]formate as starting materials.

Before starting the synthesis of SAIL serine, the hydrogenation of the dehydroserine derivatives was optimized using various non-labeled dehydroserine derivatives in the presence of (+)-1,2-bis((2*S*,5*S*)-2,5-diethylphospholano)benzene(cyclooctadiene)rhodium (I) tetrafluoroborate ((*S*,*S*)-Et-DuPHOS-Rh) (Burk, 1991; Burk *et al.* 1998) as the catalyst. The optical yield of the α-position during the hydrogenation was determined from serine. When the acyl group was used to protect the enol group, the deoxygenation occurred at the β-position, and the optical yields at the α-position were not sufficient. In order to inhibit the deoxygenation, silyl enol ether was applied to the substrate. The *t*-butyldiphenylsilyl (TBDPS) and thexyldimethylsilyl (TDS) derivatives of the protective group did not exhibit deoxygenation; however, TBDMS gave rise to a complex mixture, due to its instability. The optical purities based on the positions of the TBDPS and TDS derivatives were 99% ee and 92% ee, respectively. In order to check the enantiopurity of the β-position by deuteration, the catalytic deuteration of the TBDPS derivatives was examined using deuterium gas instead of hydrogen gas, followed by deprotection, to give α,β-dideuterium-labeled serine. In the 300-MHz ^{1}H NMR spectrum of α,β-dideuterium-labeled serine the signals for the α-proton and the 3S proton had disappeared due to *cis* addition to the *re* face, indicating that the stereochemistry at the β-position is an *R* configuration (Terauchi *et al.* 2008).

The synthesis (Fig. 5) begins with the condensation of ethyl [1,2,3-^{13}C;^{15}N;3-^{2}H]hippurate (**4**) with ethyl [^{13}C;^{2}H]formate, followed by the silylation of the enolate **5**, to give exclusively the *Z*

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

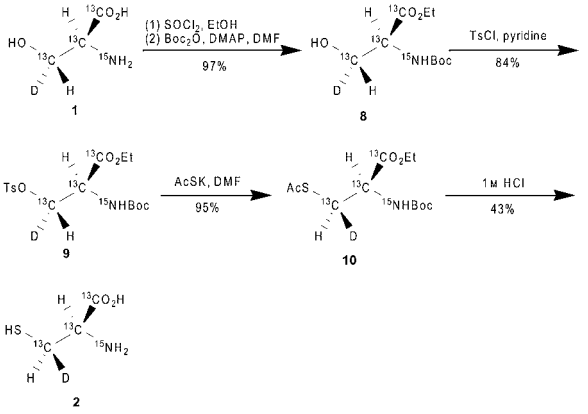


Fig. 6. Synthesis of SAIL cysteine (Terauchi *et al.* 2008).

isomer of the dehydroserine derivative **6** in 64% yield. The obtained derivative **6** was examined using asymmetric catalytic hydrogenation, followed by deprotection to give SAIL serine **1**. The enantiopurity based on the α-position was determined to be 95% ee by a high performance liquid chromatography (HPLC) analysis using a chiral stationary column (SUMICHIRAL OA-6100).

4.1.3 Cysteine

A technique of stereoselective labeling with deuterium at the β-position of cysteine was developed starting with the enzymatic reaction of fumaric acid to produce malic acid, which would form (2*S*,3*S*)-[3-^{2}H]cystine (Axelsson *et al.* 1991). Also the preparation of (2*S*,3*R*)-[3-^{2}H]serine and (2*R*,2′*R*,3*S*, ′*S*)-[3,3′-$^{2}H_2$]cystine starting from (2*R*)-serine were described (Oba *et al.* 2006). Although these methods would be sufficiently affordable for application to both SAIL serine and SAIL cysteine with minor variations, they could not be applied to the SAIL method because of difficulties in the procurement of materials and low yield. As an alternative, the synthesis of SAIL cysteine using SAIL serine 2,6,7-trioxabicyclo[2,2,2]octyl ester was considered (Oba *et al.* 2006). However, this method suffers from poor reproducibility, and similar to SAIL serine starting from [1,2,3-^{13}C;^{15}N]serine, it is difficult to increase the scale of the production capacity.

Instead, SAIL cysteine can be obtained from the abovementioned SAIL serine by converting the hydroxyl group of serine into a thiol group, by an S_N2-type displacement (Terauchi *et al.* 2008). Following a scheme developed by Arnold *et al.* (Arnold *et al.* 1985, 1988; Ramer *et al.* 1986; Williams, 1989), the synthesis of SAIL cystine ((2*R*,2′*R*,3*S*,3*S*)-[1,1′,2,2′,3,3′-$^{13}C_6$,2,2′-$^{15}N_2$; 3,3′-$^{2}H_2$]cystine) could be accomplished with a 35% overall yield, although potentially risky chemicals, such as dimethyl azodicarboxylate, were used. To avoid potentially hazardous reaction materials, a more easily accessible means to synthesize the stereoisomer at the β-position of SAIL cysteine was also examined (Fig. 6). The (2*S*,3*R*)-*N*-benzoyl serine ethyl ester **7** was converted to tosylate and then treated with potassium thioacetate to give the (2*R*,3*S*)-cysteine derivative. Although the displacement at the β-position by the thioacetate anion proceeded with inversion, unfortunately, the yield was low, due to by-product formation. However, changing the protection to use the *t*-butoxycarbonyl group prevented the by-product formation. Thus, derivative **8**,

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

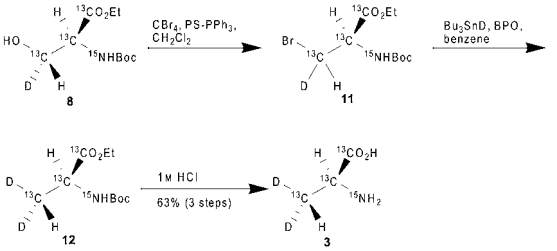


Fig. 7. Synthesis of SAIL alanine (Terauchi *et al.* 2008).

derived from SAIL serine **1**, was subjected to tosylation followed by thioacetylation by potassium thioacetate, to yield the deuterium-labeled cysteine derivative **10** (Fig. 6). The deprotection of compound **10** was accomplished by refluxing with 1 M HCl, and the resulting SAIL cysteine hydrochloride **2** ((2*R*,3*R*)-[1,2,3-$^{13}C_3$;2-^{15}N;3′-^{2}H] cysteine) was purified by ion-exchange column chromatography on a DOWEX 50WX8 column with 43% yield. The enantiopurity of the α-position of amino acid **2** was determined to be 95% ee, by an HPLC analysis using a chiral stationary column (DAICEL CROWNPAK CR+).

4.1.4 Alanine

SAIL alanine **3** is dideuterated at the β-position, in order to increase the signal intensity (Fig. 3). This labeling pattern is optimized for larger proteins. The synthesis of SAIL alanine (Terauchi *et al.* 2008) employs deoxygenation of the hydroxyl group of the SAIL serine derivative **8** (Fig. 7). Derivative **8** was treated with polymer-supported triphenylphosphine and carbon tetrabromide for conversion to the bromoalanine derivative **11**, after which derivative **11** was reduced with tributyltin deuteride, to give derivative **12**. The deprotection of compound **12** was carried out by refluxing with 1 M HCl, and the resulting alanine hydrochloride was subjected to ion-exchange column chromatography, using a DOWEX 50WX8 column, to give SAIL alanine **3** ((2*S*)-[1,2,3-$^{13}C_3$;2-^{15}N;3,3-$^{2}H_2$]alanine) in 63% yield from derivative **8**. The enantiopurity at the α-position was determined to be 93% ee by an HPLC analysis using a chiral stationary column (SUMICHIRAL OA-6100).

4.1.5 Glutamic acid

Although several groups had reported the stereoselective synthesis of glutamic acid derivatives with stereospecific deuterium labeling, these synthetic routes were not suitable for regioselective ^{13}C labeling coupled with deuterium labeling because the starting ^{13}C-labeled materials were unavailable. However, (2*S*,3*S*)- and (2*S*,3*R*)-[2-^{13}C;3-^{2}H]glutamic acid can be obtained from commercially available materials by a straightforward synthesis via the asymmetric hydrogenation of a 2,3-didehydroamino acid derivative using an α-oxoglutarate or a dehydroserine derivative as the starting material (Okuma *et al.* 2009). This report describes a simple method for the synthesis of [2-^{13}C;3-^{2}H] glutamic acid, routed through ornithine as the key intermediate. Asymmetric hydrogenation of the [2-^{13}C;3-^{2}H] 2,3-didehydroornithine derivative, which is derived from readily available [2-^{13}C] glycine and β-alanine, is carried out to yield an ornithine derivative labeled with deuterium at the β-position. This is stereoselectively followed by conversion of the derivative into the target [2-^{13}C;3-^{2}H] glutamic acids by oxidation of the δ-position.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

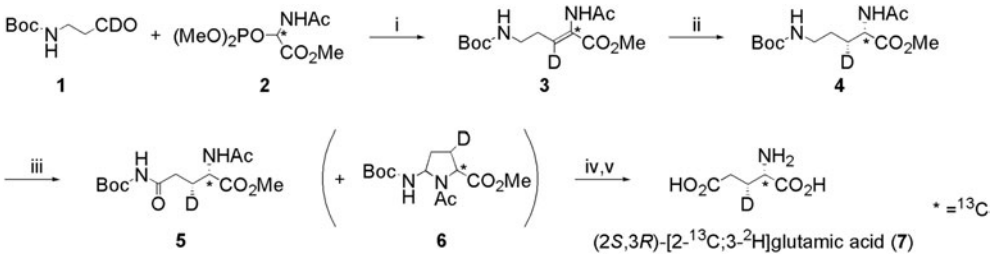


Fig. 8. Synthesis of SAIL glutamic acid (Okuma *et al.* 2009). Reagents and conditions: (i) DBU, CH_2Cl_2, −25 °C, 71%; (ii) (*S*,*S*)-Et-DuPHOS-Rh, benzene, H_2 (0·4 MPa), 84%; (iii) RuO_2, $NaIO_4$, ethyl acetate-H_2O, 38%; (iv) 2 M HCl reflux; (v) DOWEX 50WX8, 70% (2 steps).

As shown in Fig. 8, the synthesis of (2*S*,3*R*)-[2-^{13}C;3-^{2}H] glutamic acid (**7**) started with the preparation of the stable isotope-labeled dehydroornithine derivative (**3**) using the Horner-Wadsworth-Emmons reaction. The starting N-Boc-3-amino[1-^{2}H]propionaldehyde (**1**) was prepared by the reduction of the Weinreb amide of N-Boc-β-alanine with lithium aluminum deuteride (LiAlD4) and was then condensed with phosphoryl[2-^{13}C]glycine ester (2) in the presence of 1,8-diazabicyclo[5·4·0]undec-7-ene (DBU) to give 2,3-didehydro[2-^{13}C;3-^{2}H]-ornithine (3) in 71% yield. Highly *Z*-selective olefination was observed when the reaction was conducted at −25 °C. Asymmetric hydrogenation of the obtained dehydroornithine (**3**) was carried out under medium pressure (0·4 MPa) of hydrogen in the presence of (+)-1,2-bis[(2*S*,5*S*)-2,5-diethylphospholano]benzene(cyclooctadiene)rhodium(I) trifluoromethanesulfonate [(*S*,*S*)-Et-DuPHOS-Rh], because the DuPHOS family of catalysts is highly efficient for the hydrogenation of β-monosubstituted acetamidoacrylates, yielding a variety of amino acids with high enantioselectivities. A DuPHOS-Rh catalyst was also employed for the asymmetric hydrogenation of dehydroserine derivatives (Terauchi *et al.* 2008). The stereochemistry at the α and β positions of [2-13C;3-2H]ornithine (4) was determined after the final glutamic acid was obtained.

The conversion of ornithine (**4**) to glutamine (**5**) was performed using a catalytic amount of ruthenium dioxide and an excess of sodium periodate in a two-phase system of ethyl acetate and water. However, [2-^{13}C;3-^{2}H]glutamine (**5**) was obtained in only 38% yield because of the considerable side reaction yielding the cyclic compound **6**. Hydrolysis of **5** was performed by refluxing it in 2 M HCl, and subsequent ion exchange treatment of the obtained glutamic acid hydrochloride with DOWEX 50WX8 afforded (2*S*,3*R*)-[2-^{13}C;3-^{2}H]glutamic acid (**7**) in 70% yield. The enantiopurity based on the α-position was determined to be 99% ee by HPLC analysis using a chiral column (MCIGEL CRS10W). This stereochemical outcome can be attributed to the exclusive *cis*-addition of hydrogen to the α-re-face of deuterated (*Z*)-2,3-didehydroornithine (**3**), promoted by the (*S*,*S*)-Et-DuPHOS-Rh catalyst.

4.2 Cell-free protein synthesis

Cell-free protein expression methods (Kigawa *et al.* 1995; Zubay, 1973) are crucial to make efficient use of the SAIL amino acids and to prevent scrambling of the label, which would occur through metabolic pathways present in cell-based protein expression systems (Arata *et al.* 1994; McIntosh & Dahlquist, 1990).

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Cell extracts from several organisms have been used for the purpose of cell-free protein synthesis (Clemens & Prujin, 1999; Kramer *et al.* 1999). Proteins that cannot be produced from cells because they are toxic to the host cells can be made from cell-free systems (Chrunyk *et al.* 1993; Henrich *et al.* 1982). In addition, proteins that tend to be digested by intracellular proteases (Goff & Goldberg, 1987; Maurizi, 1987) or to aggregate in inclusion bodies (Chrunyk *et al.* 1993) in cells can sometimes be produced from cell-free extracts. With cell-free systems, proteins can be made from PCR-amplified DNA in as high yields as from plasmid DNA (Wu *et al.* 2007). Labeled amino acids can be incorporated efficiently and with minimal scrambling of the label, because the amino acid metabolic activity of cell-free extracts is low (Kigawa *et al.* 1995). Cell-free systems are now considered as alternatives to conventional *in vivo* expression. Commercial products provide convenience and improved protein yields (Betton, 2003). Protein synthesis can also be carried out by the 'PURE' system, which was reconstituted from purified components of the translation machinery (Shimizu *et al.* 2001).

Cell-free systems offer particular advantages for preparing samples for NMR spectroscopy. One can take advantage of the inactive amino acid metabolism for stable isotope labeling, particularly for selective labeling (Guignard *et al.* 2002; Kigawa *et al.* 1995, 1999; Klammt *et al.* 2004; Shi *et al.* 2004). A good cell-free system for NMR spectroscopy is capable of producing milligram-scale samples of labeled protein in a robust, reproducible fashion at reasonable cost. Commercial cell-free systems, including the PURE system, however, are relatively expensive.

Efficient incorporation of the SAIL amino acids into the protein of interest without metabolic scrambling can be achieved by an *E. coli* cell-free protein synthesis system (Fig. 9) that was optimized for the preparation of labeled NMR samples (Takeda *et al.* 2007; Torizawa *et al.* 2004) and overcomes two major problems with conventional *E. coli* cell-free systems: the production of protein mixtures with heterogeneous N-termini and the dilution of added labeled amino acids by endogenous amino acids of natural abundance from the extract.

4.2.1 Validation of *E. coli* expression

One of the advantages of cell-free systems is that they can sometimes produce proteins that cannot be expressed by *in vivo* methods (Chrunyk *et al.* 1993; Goff & Goldberg, 1987; Henrich *et al.* 1982; Maurizi, 1987). With the exception of such cases, it is recommendable to carry out both *E. coli* cellular expression and cell-free production of the target protein with uniform ^{15}N labeling, for the following reasons: (i) This allows for a comparison between the NMR spectra obtained from *in vivo* and *in vitro* expression. [^{1}H,^{15}N]-HSQC spectra from both approaches should be compared carefully to detect any possible difference between *in vivo* and *in vitro* expression. (ii) It is less costly and requires less effort to produce proteins from *E. coli* cells than by the cell-free method. Thus, protein prepared by the former method can be used to work out the purification protocols and to search for optimal solution conditions before preparing the protein by cell-free methods. (iii) Most proteins that express well in *E. coli* cells can be produced successfully by the cell-free method. Of 44 proteins that could be expressed in *E. coli*, 35 (80%) were found to work well also in the cell-free synthesis system (Fig. 10). Thus, a protein that has been shown to express well in *E. coli* cells has a good chance of being produced well in the cell-free system. It is recommendable to carry out both *E. coli* cellular expression and cell-free production of the target protein with uniform ^{15}N labeling. This allows for a comparison between the NMR spectra obtained from *in vivo* and *in vitro* expression. [^{1}H,^{15}N]-HSQC spectra

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

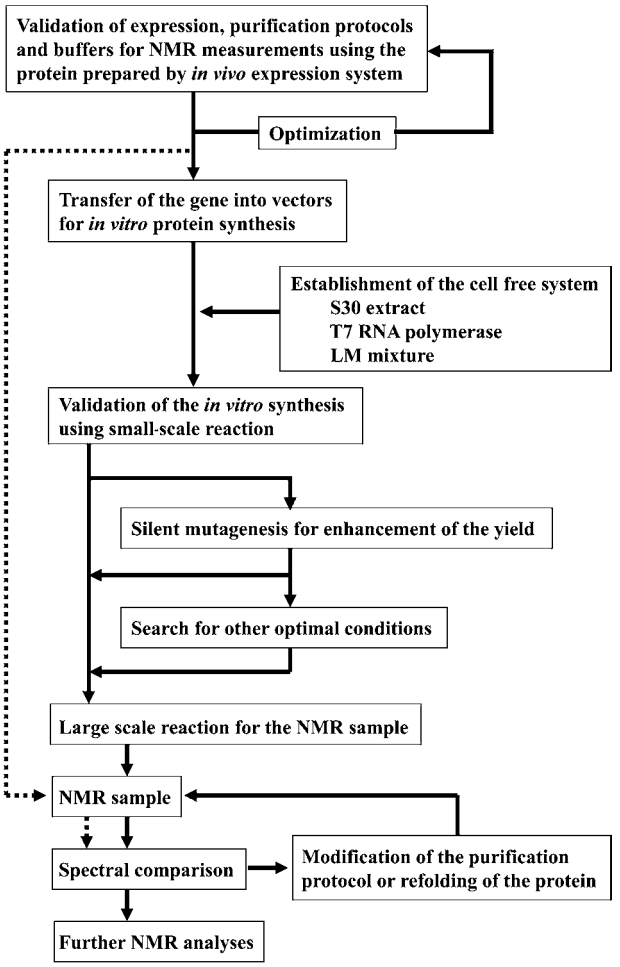


Fig. 9. Flow chart of the cell-free protein synthesis strategy used for preparing SAIL protein samples for NMR analysis (Torizawa *et al.* 2004).

from both approaches should be compared carefully to detect any possible differences between *in vivo* and *in vitro* expression.

4.2.2 Construction of cell-free protein synthesis system

The preparation of S30 extract is critical to the success of the system. As most investigators and commercial vendors of cell-free systems have emphasized, the most important issue is the level of activity of the extract (Kigawa *et al.* 1999; Kim *et al.* 1996; Kim & Swartz, 2000; Madin *et al.* 2000; Spirin *et al.* 1988). *E. coli* strain A19 yielded a reliable extract because of its lack of RNase I (Kim & Swartz, 2000). The strain BL21 Star (DE3) lacking the activity of RNase E was shown to have a 1·2–1·4 times higher activity than A19 and was therefore chosen for the extract. The activity is strongly dependent on how the extract is prepared and on its concentration.

The concentration of the extract should be optimized to maximize the amount of protein (Kigawa *et al.* 1999). For labeled NMR samples, a higher concentration of extract not only leads

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

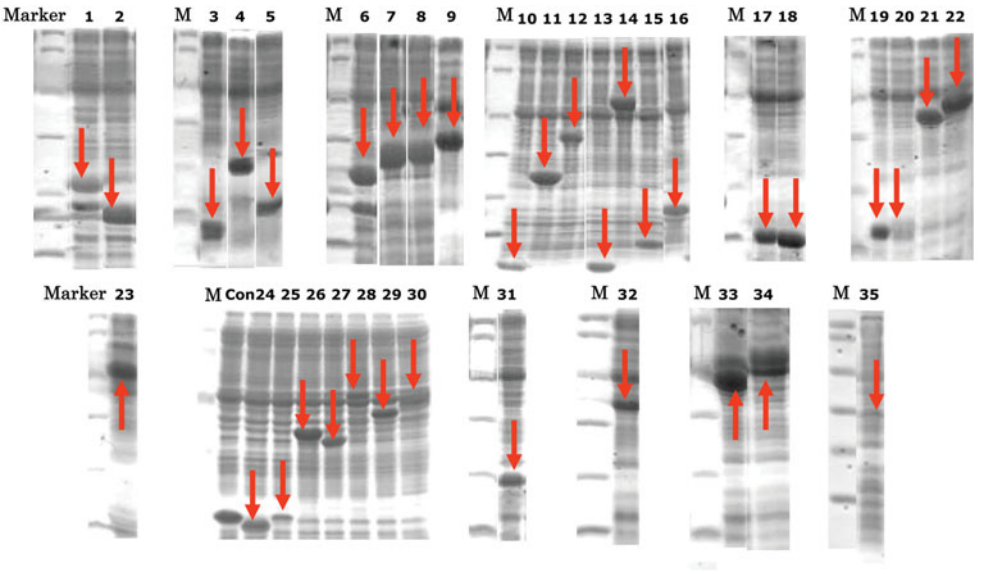


Fig. 10. SDS-PAGE analyses of proteins synthesized by the small scale cell-free system for preparing labeled proteins (Torizawa *et al.* 2004): 1, EPPIa; 2, EPPIb; 3, *E. coli* parvulin; 4, human parvulin; 5, human parvulin without WW domain; 6, DSBa; 7, DSBc; 8, DSBg; 9, DSBe; 10, drk SH3 domain; 11, CAT1; 12, Calcineurin B; 13, α-spectrin SH3 domain; 14, alkaline phosphatise; 15, SSI; 16, CPI17(35–120); 17, FKBP12.6; 18, GroES; 19, SecB; 20, SecE; 21, leader peptidase; 22, OmpA1; 23, IP3 receptor; 24, RuvC; 25, *T. martina* RuvC; 26, At5g45600; 27, At5g24650; 28, At4g20010; 29, At2g02570; 30, At1g18800; 31, calmodulin; 32, At3g16450; 33, MBP; 34. β-subunit; 35, GFPuv. 'M' denotes molecular weight markers. The synthesized proteins are indicated by arrows. Both soluble and insoluble fractions were loaded onto the gel.

to higher yields of protein, but also to a dilution of the label by endogenous amino acids in the extract. This problem of label dilution has been analyzed (Torizawa *et al.* 2004). The labeling efficiency for [^{2}H,^{13}C,^{15}N]-calmodulin produced from a conventional cell-free extract prepared without gel filtration was only about 90%. By introducing a gel filtration step, the incorporation level increased to 96% without loss of the activity of the extract.

The extract described in Section 4.2.3 produced up to 2 mg (EPPIb) protein per milliliter of the reaction solution under optimal concentrations of added amino acids (1 mM each). The yield of protein decreases as the levels of added amino acids are increased, 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 by analyzing protein yields as a function of concentration and by choosing a value slightly below the maximum. Because the labeled amino acids represent the most important cost factor, the mole ratios of the individual amino acids were 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 LC-MS/MS and CE-MS in order to investigate the percentage of unlabeled amino acids incorporated into the protein. These were: 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 level of incorporation of unlabeled Gly increased to 12·7%. This is the reason

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

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 coding for the target protein, a promoter and a terminator for T7 RNA polymerase. High-copy-number plasmids can reduce the cost and time required to produce the template DNA. The pIVEX vector provided by Roche has the advantage over other high-copy plasmids that it enables the prediction of silent mutants that can improve cell-free synthesis yields as discussed in Section 4.2.6.

4.2.3 Preparation of S30 extract

The preparation of a S30 extract is based on earlier work (Davanloo *et al.* 1984; Kigawa *et al.* 1999; Pratt, 1984; Zubay, 1973) and optimized for the *in vitro* synthesis of SAIL proteins (Takeda *et al.* 2007), since commercially available S30 extracts often contain non-labeled amino acids, which results in the incorporation of non-labeled amino acids into the protein (Torizawa *et al.* 2004). It comprises the following steps:

1. Inoculate stock of *E. coli* (A19, BL21 Star etc.) into 10 ml of LB medium in a 50-ml tube and grow the cells overnight at 37 °C with shaking.
2. Inoculate 10 ml of the culture medium into 1 l of incomplete rich medium in a 2-l flask.
3. Grow the cells at 37 °C with shaking to an OD_{650} of 0·7. The growth rate of the cells in the culture correlates with the activity of the resulting extract.
4. Centrifuge the cells (5000 *g*, 4 °C, 10 min) and wash them with 200 ml of ice-cold S30 buffer containing 0·05% 2-mercaptoethanol by suspending moderately three times. Foaming of the suspension must be avoided. Can be left overnight in the centrifuge tube at – 80 °C.
5. Gently resuspend the cell pellet with 200 ml of ice-cold S30 buffer containing 0·05% (v/v) 2-mercaptoethanol by slurrying. Centrifuge them (5000 *g*, 4 °C, 10 min) and weigh the *E. coli* pellets. Resuspend the pellet in 1·27 ml of S30 buffer per gram of *E. coli*.
6. Disrupt the cells with the French press at 20 000 psi (1400 kg cm^{-2}). Add 30 μl of 1 M dithiothreitol (DTT) to the lysate. Centrifuge (30 000 *g*, 4 °C, 30 min) using DEPC-treated/autoclaved centrifuge tubes. Carefully remove approximately 1·4 ml of the supernatant per gram of *E. coli* without mixing with the precipitate.
7. Transfer the supernatant to the unused DEPC-treated/autoclaved tubes. Centrifuge them (30 000 *g*, 4 °C, 30 min) and remove approximately 1·0 ml of the supernatant per gram of *E. coli* into a 50-ml tube.
8. Shake the tube at 37 °C for 80 min.
9. Dialyze the solution at 4 °C for 45 min against 2 l of S30 buffer using a dialysis tube with molecular weight cutoff (MWCO) of 6000–8000. Allow a little air into the tube to float. Repeat the dialysis twice. Centrifuge (15 000 *g*, 10 min, 4 °C). The dialysis tube should be treated with DEPC-containing water and then rinsed with RNase-free water before use.
10. Fill an open column (Econo-column chromatography column 2·5 × 20 cm) with Sephadex G25 resin uniformly and set the column vertically in cold space. Attach an Econo-column funnel to the top end of the column. Pour the 500 ml of S30 buffer from the funnel into the column. Note, that the uniformity of the filled resin affects resolution of gel filtration.
11. Apply the supernatant from step 9 to a column that was pre-equilibrated at step 10 at 4 °C. After loading the supernatant, continue to supply the funnel with the S30 buffer to maintain

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Table 1. *Composition of reaction and dialysis solutions in cell-free reactions for SAIL protein synthesis (Takeda et al. 2007)*

Stock solution	Reaction solution	Dialysis solution
RNase-free water (μl)	1120	11 608
1·4 M NH_4OAc (μl)	98	392
0·5 M $Mg(OAc)_2$ (μl)	150	600
SAIL amino acid mixture (25 mM each) (μl)	200	800
0·645 M creatine phosphate (μl)	400	1600
LM mixture* (μl)	1250	5000
1 mg/ml template DNA (μl)	100	–
11 mg/ml T7 RNA polymerase (μl)	45	–
40 units/μl RNase inhibitor (μl)	12·5	–
10 mg/ml creatine kinase (μl)	125	–
S30 extract (μl)	1500	–
Total volume (ml)	5	20

*Low molecular weight mixture. For 200 ml, combine 22 ml 2 M HEPES-KOH pH 7·5, 33·4 ml 6 M K(OAc), 210 mg DTT, 530 mg ATP, 338 mg CTP, 335 mg GTP, 310 mg UTP, 172 mg cAMP, 28 mg folinic acid, 140 mg tRNA, 64 ml 50% (w/v) PEG-8000, and RNase-free water (Takeda *et al.* 2007).

the flow in the column. When the first fraction reached to the bottom, start to collect 1·4 times the volume of the applied extract. Determine the first fraction judging from its colour and turbidity. The fraction to be collected looks yellow.

12. Dialyse the eluate at 4 °C for 70 to 80 min against 700 ml of an equal weight mixture of PEG-8000 and S30 buffer. Before use, the PEG-S30 buffer at 4 °C should always be stirred to avoid PEG deposition. Adjust the dialysis time so as to concentrate the extract up to 0·86 times the volume. Dialyze it at 4 °C for 60 min against 2 l of S30 buffer.
13. Transfer the extract to 1·5-ml tubes. Freeze the tubes with liquid nitrogen. Store them at −80 °C.

Instruments and glassware used in preparing S30 extracts and in cell-free synthesis reactions were washed with autoclaved water containing 0·5% (v/v) DEPC (denoted as DEPC-treated water) and subsequently autoclaved if possible.

4.2.4 Preparation of SAIL proteins by the *E. coli* cell-free method

The preparation of a SAIL protein NMR sample by the *E. coli* cell-free method comprises the following steps (Takeda *et al.* 2007):

1. Prepare the reaction solution and the dialysis solution by mixing the components as shown in Table 1. Dissolve SAIL amino acid mixture in water and then add it to the cell-free reaction solution. If SAIL amino acids appear insoluble in water, warm it up to 60 °C. Use sanitary gloves to prevent contamination with RNases. Thaw the frozen S30 extract on ice. Prepare creatine phosphate in RNase-free water just before use. An excess of heating to SAIL amino acid may cause a racemization especially at high pH.
2. Cut outer tube of the Float-A-Lyzer at an appropriate height such that inner solution in tube is completely soaked to the dialysis solution when an inner membrane apparatus is set to the outer tube. Pour the dialysis solution into the outer tube. Place the inner membrane apparatus

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

of the Float-A-Lyzer into the outer tube and pour the reaction solution into the inner membrane. Cover the tube with Parafilm.

3. Shake the tube to allow for production of target proteins. The optimal temperature and incubation times should be determined in small-scale cell-free reactions with volumes of the reaction and dialysis solutions of 0·5 and 2·0 ml, respectively, before the large-scale reaction.
4. Retrieve the reaction solution and the dialysis solution. If the produced protein has a molecular weight smaller than MWCO of the membrane, check the outer solution for presence of the protein.
5. Purify the produced protein according to purification procedures of 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 can be overcome by using a cleavable N-terminal tag (Torizawa *et al.* 2004).

The approach has been used successfully for the cell-free production of a large number of proteins (Torizawa *et al.* 2004). SDS-PAGE analyses of many of the proteins synthesized by the cell-free system are shown in Fig. 10.

4.2.5 Validation of cell-free synthesis of the protein

Small-scale reactions are useful for determining whether a target protein is synthesized or not. The recommended composition of reagents in the small scale reaction is the same as in the large-scale reaction used for preparation of the NMR sample, with the important difference of PEG-8000. PEG plays an important role as a stabilizer of the system (Kim & Swartz, 2000). In the case of calmodulin, the ratio of protein yields with and without PEG-8000 was 1·6 (Torizawa *et al.* 2004). Even a small amount of PEG can disturb the SDS-PAGE quantitation of the level of production of a target protein, despite attempts to remove PEG by acetone or TCA precipitation of protein. For example, after cell-free synthesis and before 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 small-scale screening reactions.

4.2.6 Improvement of the yield of synthesized protein

It has been reported that the yield of proteins from *E. coli* cell-free synthesis is dependent on the concentrations of magnesium ions and template DNA (Kramer *et al.* 1999). Table 1 shows the composition of the reaction and dialysis solutions that was nearly optimal for the cell-free synthesis of many proteins (Takeda *et al.* 2007). This can be improved to increase the yield of a particular protein. A better yield can also be achieved by optimizing the sequence of the template DNA. The ProteoExpert software (Roche Applied Science; Biomax) was used to predict potentially yield-increasing silent mutants that do not alter the amino acid sequence of the resulting protein. This empirical prediction is based, in part, on calculations of the secondary structure of the mRNA including the pIVEX vector and the target gene, although full details on the algorithm are unavailable. One round of calculations provides 10 suggested candidates for silent mutations in the region close to the N-terminus of the protein. Roche recommends to introduce these mutations by PCR into linear templates and to use small-scale syntheses to check the relative yields. The mutated sequence that gives the highest yield is then transferred to a circular template and used for large-scale reactions.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Designing oligonucleotides with the best silent mutations and introducing them into the circular plasmid by restriction enzymes and ligase avoids possible difficulties and errors of the PCR reaction. For calmodulin, the protein yields with 2 of the 10 plasmids incorporating silent mutations were much higher than with the original plasmid. In the large-scale reaction carried out with the best sequence and 51·1 mg of amino acid mixture (1·7 mg/ml), the amount of calmodulin synthesized was 5·2 mg (10 wt%). This was twice the yield from the construct before the silent mutagenesis.

The silent mutations in the DNA sequences of a cleavable N-terminal tag are often identical for different proteins. Therefore, it is in general unnecessary to prepare all of the predicted tag DNAs for each protein carrying the tag. This reduces the cost and time to prepare all of the 10 (20 including the complimentary strands) DNAs each time. Two kinds of tags with a variety of the silent mutations were prepared (Torizawa *et al.* 2004). One tag has the same amino acid sequence as that of pET15b, namely, a $(His)_6$-tag followed by a thrombin cleavage site; the other is identical except for the absence of the $(His)_6$-tag. The former is used for easy purification of proteins by immobilized metal affinity chromatography, and the latter is used in cases where the protein of interest can be purified easily by other means. Both of the tags can be cleaved with thrombin. The cleavage can generate another important advantage as described in the next section. Constructs with a cleavable tag are preferred, provided that protein production is adequate, because they will ensure a homogeneous N-terminus of the protein as described in the next section.

4.2.7 N-terminal heterogeneity in cell-free protein synthesis

The $[^1H,^{15}N]$-HSQC spectrum of calmodulin synthesized by the cell-free system using the native sequence showed more peaks than predicted from the sequence, although the purified protein was observed as a single band by SDS-PAGE (Torizawa *et al.* 2004). Doubled peaks were found to arise from the six N-terminal residues and, with smaller chemical shift differences, for eight residues in other regions. Intensities of the superfluous peaks were not always completely reproducible. N-terminal sequencing and mass spectrometry showed that the sample consisted of three molecular species: one with N-terminal formyl Met (f-Met_0) (60%–90%), one with N-terminal Met 0 (~10%) and one with N-terminal Ala 1 (10%–40%). Also for the protein EPPIb with 164 amino acid residues, 33 additional peaks were observed in the $[^1H,^{15}N]$-HSQC spectrum of the protein prepared by the cell-free system compared with the spectrum of the protein produced in *E. coli* cells (Torizawa *et al.* 2004).

This problem could be overcome by producing the protein with a cleavable N-terminal tag, which can simultaneously be engineered to improve protein yields. The $[^1H,^{15}N]$-HSQC spectrum of calmodulin produced in this way showed no detectable extra peaks (Torizawa *et al.* 2004). It is noteworthy that NMR signals from calmodulin, which has a flexible N-terminus, are affected by the heterogeneity and that the region affected is quite extensive. Larger spectral complications may arise from heterogeneity of the peptide chain for proteins with a structured N-terminus.

4.2.8 Large-scale protein production for NMR samples

In large-scale protein preparations, it is critical to minimize losses of the product. The MWCO of the membrane used for the cell-free reaction should be 50 kDa, because membranes with a smaller MWCO reduce the yield of protein (Kigawa *et al.* 1999). Because this cutoff is larger than

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

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 protein product in the inner reaction solution equalled that in the outer solution, although the protein in the outer chamber sometimes could not be detected well by SDS-PAGE owing to the dilution factor (Torizawa *et al.* 2004). 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 before the first chromatography step, either to remove salts before binding to an ion-exchange column or to remove nucleic acids or other contaminants that may interfere with chromatography.

It may be better to dialyze the solutions against the buffer used for the first purification step. Although a protein with a $(His)_6$-tag can be purified easily by Ni affinity chromatography, DTT in the system sometimes disturbs the purification if chromatography is adopted as the first step. 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 is reduced to 80%–90% after only 3 months of storage at $-80\,^{\circ}C$ (Torizawa *et al.* 2004).

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 SAIL amino acid mixture, composed following this protocol, yielded 5·5 mg of purified, soluble 17 kDa calmodulin, or 5·3 mg of purified, soluble 41 kDa maltodextrin-binding protein (MBP) (Kainosho *et al.* 2006).

4.3 SAIL NMR spectroscopy

4.3.1 General considerations

Virtually all types of biomolecular NMR experiments benefit from SAIL and can be used with minimal changes. When collecting and analyzing data for a SAIL protein, deuterium decoupling should be applied during ^{13}C evolution times to avoid splitting by the ^{13}C-^{2}H coupling, and what are methyl and methylene groups in uniformly labeled proteins should be treated as methine groups. Constant-time [^{1}H-^{13}C]-HSQC NMR data sets collected from SAIL and uniformly labeled proteins clearly demonstrate the superiority of the SAIL method (Figs 11 and 12). Each ^{1}H-^{13}C pair in the protein is associated with a peak in the spectrum. Severe signal overlap in the case of uniform labeling is alleviated in the spectrum from the SAIL proteins. The improvements, which are particularly apparent for larger proteins (Fig. 12), result from the reduction in the number of ^{1}H signals as well as from sharpening of the remaining signals.

The SAIL method also improved sensitivity. Part of the gain arises from longer ^{1}H and ^{13}C transverse relaxation times resulting from substitutions of ^{1}H by ^{2}H. Reduced relaxation during magnetization transfer steps in experiments such as constant-time [^{1}H-^{13}C]-HSQC leads to increased signal-to-noise ratio. Reduced long-range couplings result in a further sharpening of signals. The signal intensities for methylene groups are three to seven times higher with SAIL than with uniform labeling under the same conditions. Improvements are more pronounced for the 41-kDa maltodextrin-binding protein MBP protein (Fig. 12*e*) than for the 17-kDa calmodulin (Fig. 11*i*). Although each observed SAIL methyl group contained only one ^{1}H as compared with three equivalent ^{1}H with uniform labeling, comparable signal intensities were observed for methyl groups as a result of the longer ^{1}H and ^{13}C transverse relaxation times. All peaks benefited

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

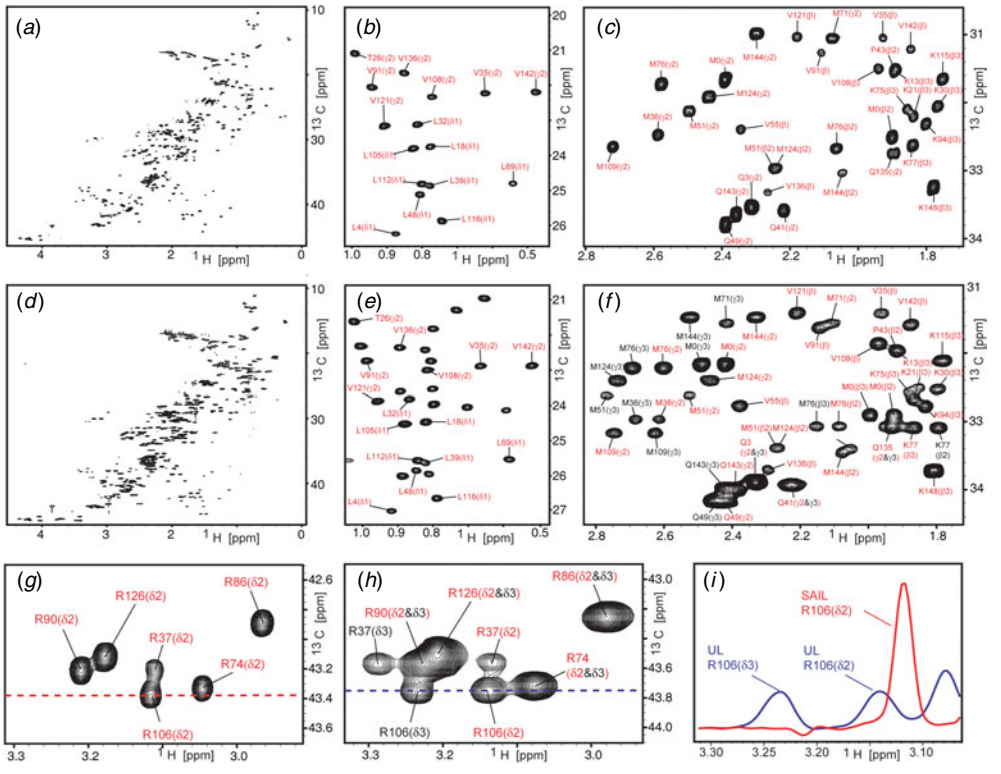


Fig. 11. Constant-time [^{13}C,^{1}H]-HSQC spectra of CaM (Kainosho *et al.* 2006). (*a*) SAIL-CaM, aliphatic region. (*b*) SAIL-CaM, methyl region. (*c*) SAIL-CaM, methylene region. (*d*) UL-CaM, aliphatic region. (*e*) UL-CaM, methyl region. (*f*) UL-CaM, methylene region. (*g*) SAIL-CaM, Arg δ region. (*h*) UL-CaM, Arg δ region. (*i*) Cross-sections from *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 one-letter amino acid code, residue number and atom identifier. Assignments of UL-CaM are as reported previously (Ikura *et al.* 1991).

from improved sensitivity and resolution and allowed the signals of SAIL-calmodulin and SAIL-MBP to be assigned readily by established methods. Side-chain assignments were determined from the analysis of two data sets: HCCH-TOCSY spectra provided connectivities among all side-chain signals and HCCH-COSY spectra were used to identify the spin systems.

4.3.2 Assignment of aromatic ring signals

Since 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. Slow ring flipping motions, which have occasionally been reported in the literature (Campbell *et al.* 1976; Wagner *et al.* 1976), make the NMR signals for the δ and ε

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

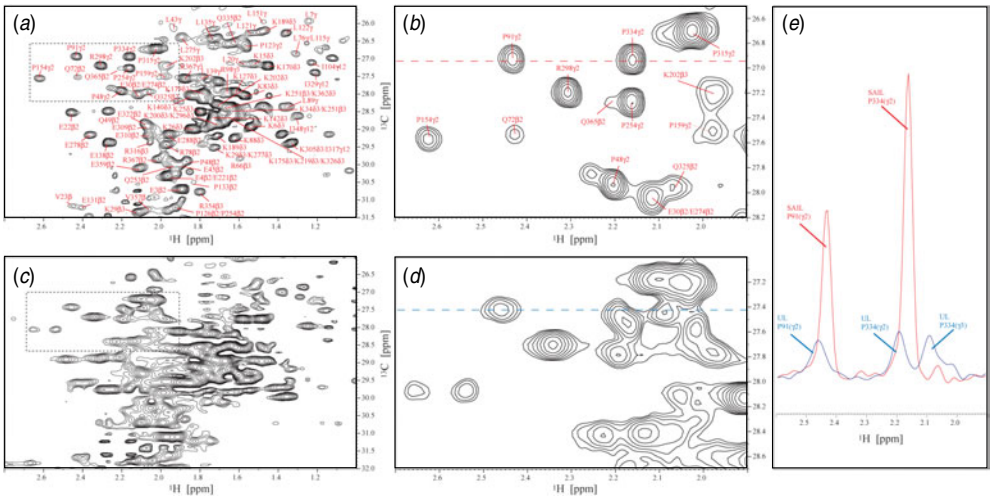


Fig. 12. Constant-time [^{13}C,^{1}H]-HSQC spectra of MBP (Kainosho *et al.* 2006). (*a*) Aliphatic region of CDH groups in SAIL-MBP. (*b*) Enlargement of the rectangular region marked in a. Assignments are indicated with one-letter amino acid code, residue number and atom identifiers. (*c*) and (*d*) Corresponding regions for UL-MBP. (*e*) Cross-sections taken at the positions indicated in *b* and *d*. The spectra for SAIL-MBP and UL-MBP were recorded under identical conditions and scaled for equal noise levels.

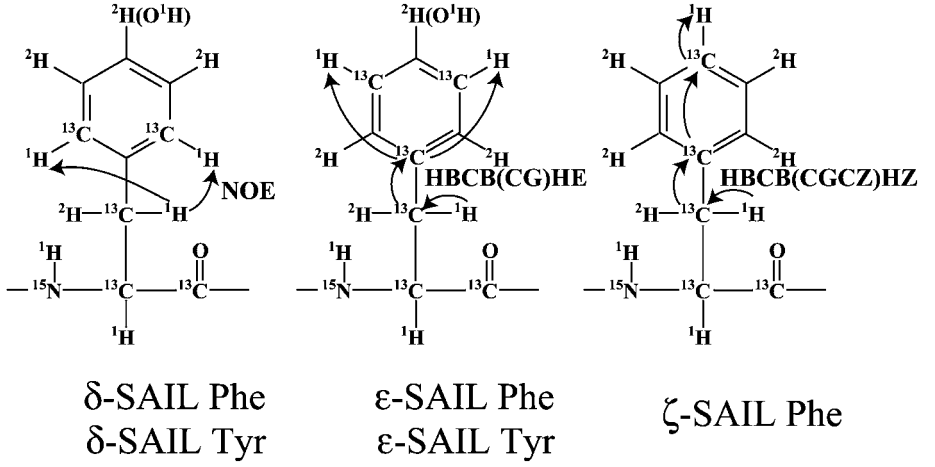


Fig. 13. Stable isotope labeling patterns for Phe and Tyr and magnetization transfer pathways for HBCBCGHE and HBCBCGCZHZ measurements. (*a*) $^{13}C^{\gamma}$-$^{13}C^{\varepsilon}$-$^{1}H^{\varepsilon}$ pattern for [H^{β}-$^{13}C^{\beta}$-$^{13}C^{\gamma}$-$^{13}C^{\varepsilon}$-$^{1}H^{\varepsilon}$]-Tyr/Phe. (*b*) $^{13}C^{\gamma}$-$^{13}C^{\zeta}$-$^{1}H^{\zeta}$ pattern for [H^{β}-$^{13}C^{\beta}$-$^{13}C^{\gamma}$-$^{13}C^{\zeta}$-$^{1}H^{\zeta}$]-Phe. ^{12}C atoms are not shown in the figures. Arrows indicate the magnetization transfer pathways. In both cases, the transfer started from $^{1}H^{\beta}$.

atoms of some Phe and Tyr residues even more difficult, or sometimes impossible, to observe (Takeda *et al.* 2008a).

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 using exclusively through-bond correlations (Torizawa *et al.* 2005). Fig. 13 illustrates the repertoire of the currently available SAIL Phe and SAIL Tyr amino acids. In the

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

[$^{13}C^{\gamma}$,$^{13}C^{\varepsilon}$,$^{1}H^{\varepsilon}$]-Tyr/Phe pattern (ε-SAIL Tyr/Phe), all non-exchangeable 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 (δ-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.

Fig. 14 compares the aromatic regions of [^{1}H,^{13}C]-HSQC spectra of calmodulin containing either uniformly [^{13}C,^{15}N]-labeled Tyr/Phe, ε-SAIL Tyr/Phe or δ-SAIL Phe. It is obvious that severe signal overlap observed for uniformly [^{13}C,^{15}N]-labeled Tyr/Phe is resolved by using SAIL Tyr/Phe because of the absence of resonances and the sharpening of the remaining resonances. Sensitivity improvements were also observed. Employing SAIL Tyr/Phe increased the aromatic signal intensities on average 13-fold for Phe ε, 26-fold for Phe ζ and 4-fold for Tyr over those of the uniformly ^{13}C-labeled rings. One of the reasons for the improvement is the attenuation of dipolar interactions that was achieved by replacing the hydrogens adjacent to the detected nuclei with ^{2}H. In the case of the uniformly ^{13}C-labeled aromatic residues, 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. The SAIL Tyr/Phe labeling patterns do not require constant time evolution in the [^{1}H,^{13}C]-HSQC measurement. TROSY with constant-time evolution has been developed for the sensitive detection of aromatic resonances in uniformly ^{13}C-labeled proteins (Pervushin *et al.* 1998). By altering the evolution period to remove the constant time, this measurement technique would allow more sensitive detection of the aromatic resonances of the novel labeling patterns in larger proteins.

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}, or, preferably, by through-bond ^{13}C-^{13}C scalar coupling connectivities (Torizawa *et al.* 2005), even for proteins containing many aromatic amino acids. Two pulse sequences for the assignment of aromatic resonances 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 with H^{β} (t_1)$\rightarrow C^{\beta}$ (t_2)$\rightarrow C^{\gamma}\rightarrow C^{\zeta}\rightarrow H^{\zeta}$ (t_3). The 3D HBCB(CG)HE experiment can also performed in a 2D version, (HB)CB(CG)HE. These pulse sequences are composed of previously employed building blocks (Edison *et al.* 1994). The most similar previously reported pulse sequences are either $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta})H^{\delta}$ or $(H^{\beta})C^{\beta}(C^{\gamma}C^{\delta}C^{\varepsilon})H^{\varepsilon}$ (Yamazaki *et al.* 1993). A major difference between the previously published sequences and the sequences for SAIL proteins is the incorporation of H^{β} evolution. This was made possible 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. ^{2}H decoupling must be applied during the ^{13}C evolution time to avoid splitting by the ^{13}C-^{2}H coupling. In the case of HBCB(CG)HE, the magnetization is transferred from C^{γ} to detect H^{ε} directly via $^{3}J_{C\gamma H\varepsilon}$ (7·9 Hz for Phe, 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^{ζ}-H^{ζ} 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

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

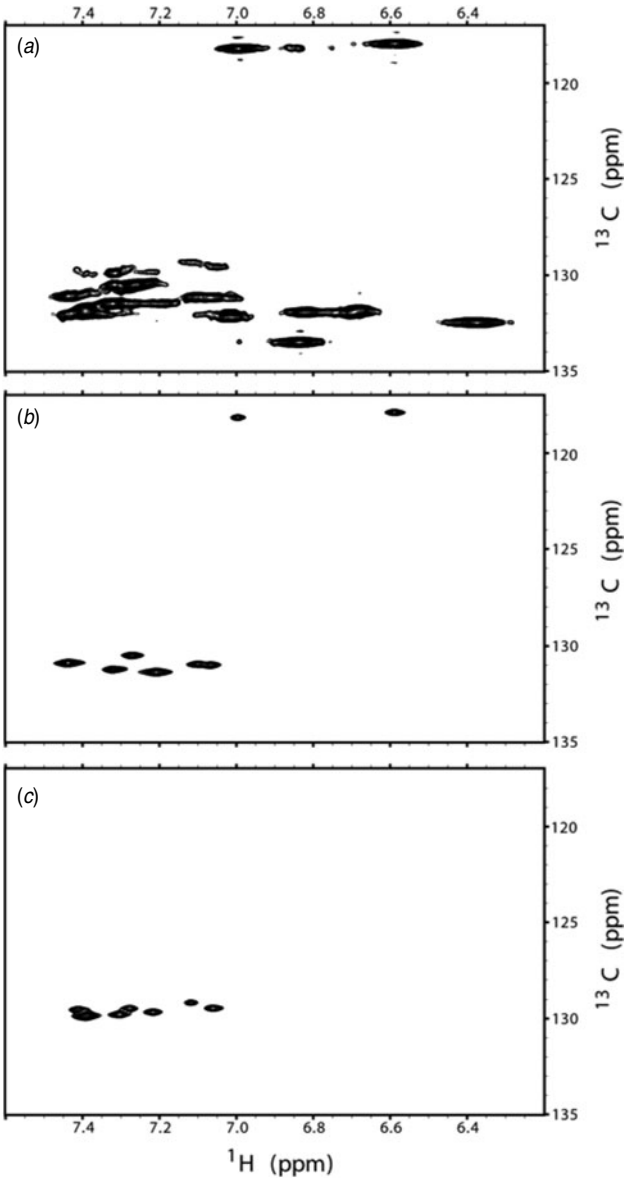


Fig. 14. ^{1}H-^{13}C correlation spectra of calmodulin (Torizawa *et al.* 2005). (*a*) Constant time HSQC spectrum of [uniformly ^{13}C,^{15}N]-Phe and Tyr selectively labeled calmodulin. (*b*) HSQC spectrum of [$^{13}C^{\gamma}$-$^{13}C^{\varepsilon}$-$^{1}H^{\varepsilon}$]-Tyr and Phe selectively labeled calmodulin. (*c*) HSQC spectrum of [$^{13}C^{\gamma}$-$^{13}C^{\zeta}$-$^{1}H^{\zeta}$]-Phe selectively labeled calmodulin. In (*a*), the ^{13}C region between 130 and 135 ppm contains the δ, ε and ζ resonances of Phe, and the δ resonances of Tyr. The corresponding regions in (*b*) and (*c*) contain only the ε or ζ resonances of Phe. Near 118 ppm in (*a*) and (*b*), the ε resonances of Tyr are detected. The small chemical shift differences between (*a*) and (*b*) or (*a*) and (*c*) are caused by isotope shifts.

correlations between the H^{β} and the ring protons could be detected for all eight Phe and two Tyr residues in calmodulin (Torizawa *et al.* 2005).

The ^{13}C-^{1}H pairs of aromatic ring moieties are considered to be amenable for strong TROSY effects, but the results reported so far were not quite satisfactory for uniformly ^{13}C-labeled, fully

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

protonated proteins (Pervushin *et al.* 1998). In each type of SAIL Phe/Tyr, 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 line widths of the TROSY peaks observed for the F1-coupled HSQC spectra are narrower than those of the anti-TROSY peaks already for the 18-kDa protein EPPIb (Takeda *et al.* 2009), and it was recently 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 to study very large proteins.

5. Automated structure analysis of SAIL proteins

5.1 Automated NOESY assignment and structure calculation with CYANA

The superior quality of SAIL protein NMR spectra and the reduced number of ^{1}H resonances render NOESY spectra of SAIL proteins particularly amenable to automated analysis. To this end, the NMR structure calculation program CYANA was supplemented with a specific SAIL amino acid residue library that reflects the $^{1}H/^{2}H$ labeling patterns of the SAIL amino acids and their complete stereospecific assignments. The automated NOE assignment algorithm of the program CYANA assigns automatically NOESY cross-peaks (Herrmann *et al.* 2002) on the basis of the chemical shift assignments and lists of NOESY cross-peak positions and intensities. The overall probability for the correctness of possible NOE assignments is calculated as the product of three probabilities that reflect the agreement between the chemical shift values and the peak position, the consistency with a preliminary 3D structure (Güntert *et al.* 1993) and network anchoring (Herrmann *et al.* 2002), i.e., the extent of embedding in the network formed by other NOEs. Restraints with multiple possible assignments are represented by ambiguous distance restraints (Nilges, 1995). Seven cycles of combined automated NOE assignment and structure calculation by simulated annealing in torsion angle space (Güntert *et al.* 1991a, b, 1997; Güntert & Wüthrich, 1991) and a final structure calculation using only unambiguously assigned distance restraints are performed. Constraint combination (Herrmann *et al.* 2002) is applied in the first two cycles to all NOE distance restraints spanning at least three residues in order to minimize distortions of the structures by erroneous distance restraints that may result from spurious entries in the peak lists and/or incorrect chemical shift assignments.

5.2 Fully automated structure analysis with SAIL-FLYA

The SAIL technology supports high-throughput protein structure determination without loss of structural quality. Until recently it was not possible to determine an NMR protein structure without manual analysis of the data and chemical shift assignments, although many computational approaches have been introduced either to support the interactive analysis by visualization and book-keeping or to provide automation for specific parts of an NMR structure determination. Automated procedures are now widely used for the assignment of NOE distance restraints and the structure calculations. Fully automated NMR structure determination (Güntert, 2009) is more demanding than automating individual parts of NMR structure analysis because the cumulative effect of imperfections at successive steps can easily render the overall process unsuccessful. For example, it has been demonstrated that reliable automated NOE assignment and structure calculation requires around 90% completeness of the chemical shift assignment

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

(Herrmann *et al.* 2002; Jee & Güntert, 2003), which is not straightforward to achieve by unattended automated peak picking and automated resonance assignments. Present systems designed to handle the whole process therefore generally require certain human interventions (Gronwald & Kalbitzer, 2004; Huang *et al.* 2005). The interactive validation of peaks and assignments, however, still constitutes a time-consuming obstacle for efficient NMR protein structure determination.

The fully automated NMR structure analysis algorithm (FLYA) (López-Méndez & Güntert, 2006) can solve NMR protein structures by purely computational means using only the primary structure and NMR spectra as input. As in the classical manual approach, structures are determined by a set of experimental NOE distance restraints without reference to already existing structures or empirical molecular modeling information. In addition to the 3D structure of the protein, FLYA yields backbone and side-chain chemical shift assignments and cross-peak assignments for all spectra. Uniting optimal labeling with full automation, SAIL-based fully automated protein structure analysis (SAIL-FLYA) is designed to solve protein structures rapidly and with enhanced accuracy and size range of applicability. The key objectives of the SAIL-FLYA system are:

1. High speed: A complete 3D structure of a protein within hours after the completion of the NMR experiments.
2. Large proteins: High-quality NMR structures of proteins with molecular masses above 25 kDa, which are largely outside the reach of conventional NMR structure determination, can be solved due to the reduced line width, smaller number of signals, and improved signal-to-noise ratio with SAIL.
3. High structure quality: The higher signal-to-noise ratio of SAIL spectra, the smaller numbers of ^{1}H shifts and the absence of ambiguities for diastereotopic pairs enable structures of higher quality than by conventional, manual or semi-automated methods applied to uniformly labeled proteins.
4. Objectivity: NMR structures are determined by a purely computational procedure without subjective, manual interventions and will thus be fully reproducible.
5. Usable by non-NMR researchers: Due to its fully automated nature the method can be applied by people without highly specialized training in NMR spectroscopy.

The FLYA algorithm (Fig. 15) uses as input data only the protein sequence and multidimensional NMR spectra. Any combination of commonly used heteronuclear and homonuclear 2D, 3D, and 4D NMR spectra can be used as input for the FLYA algorithm, provided that it affords sufficient information for the assignment of the backbone and side-chain chemical shifts and for the collection of conformational restraints. Seven, purely computational steps are applied:

1. Peak picking
2. Peak filtering
3. Initial chemical shift assignment, yielding an ensemble of chemical shift values for every spin
4. Chemical shift consolidation, resulting in a consensus chemical shift value for every spin
5. NOESY assignment
6. Structure calculation
7. Energy refinement in explicit solvent

The complete procedure runs without human intervention, driven by the NMR structure calculation program CYANA.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

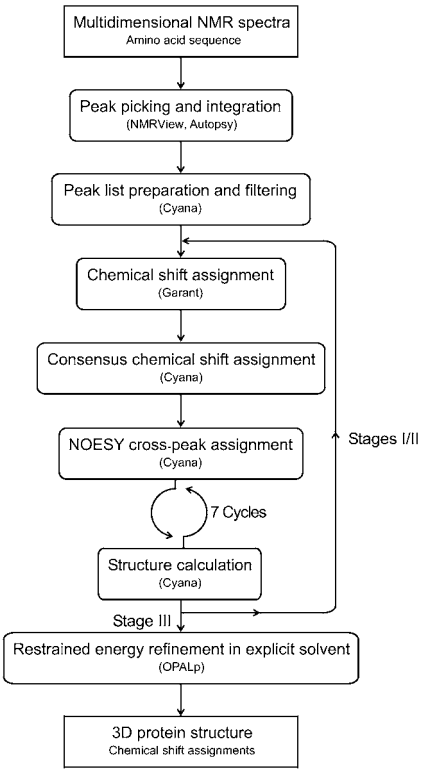


Fig. 15. Flowchart of the fully automated structure determination algorithm FLYA (López-Méndez & Güntert, 2006).

Peaks are identified in the multidimensional NMR spectra using the automated peak picking algorithm of NMRView (Johnson, 2004) or AUTOPSY (Koradi *et al.* 1998). Peak integrals for NOESY cross-peaks are determined simultaneously. Since no manual corrections are applied, the resulting raw peak lists may contain, in addition to the entries representing true signals, a significant number of artifacts (see Figs 2 and 4 of López-Méndez & Güntert, 2006). The following steps of the fully automated structure determination algorithm can tolerate the presence of such artifacts, as long as the majority of the true peaks have been identified.

Based on the peak positions, and in the case of NOESY spectra, peak volumes, peak lists are prepared by CYANA (Güntert, 2003; Güntert *et al.* 1997). Depending on the spectra, the preparation may include unfolding aliased signals, systematic correction of chemical shift referencing and removal of peaks near the diagonal or water lines. The peak lists resulting from this step remain invariable throughout the rest of the procedure. An ensemble of initial chemical shift assignments is obtained by multiple runs of a modified version of the GARANT algorithm (Bartels *et al.* 1996; Bartels *et al.* 1997) with different seed values for the random number generator (Malmodin *et al.* 2003). The original GARANT algorithm was modified for new spectrum types and for the treatment of NOESY spectra when 3D structures are available. In analogy to NMR structure calculation in which not a single structure but an ensemble of conformers is calculated using identical input data but different randomized start conformers, the initial chemical shift assignment produces an ensemble rather than a single chemical shift value for each

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

^{1}H, ^{13}C and ^{15}N nucleus. The peak position tolerance is typically set to 0·03 ppm for the ^{1}H dimensions and to 0·4 ppm for the ^{13}C and ^{15}N dimensions. These initial chemical shift assignments are consolidated by CYANA into a single consensus chemical shift list. The most highly populated chemical shift value in the ensemble is computed for each ^{1}H, ^{13}C and ^{15}N spin and selected as the consensus chemical shift value that will be used for the subsequent automated assignment of NOESY peaks. The consensus chemical shift for a given nucleus is the value ω that maximizes the function $\mu(\omega) = \sum_j \exp\left(-(\omega-\omega_j)^2/2\Delta\omega^2\right)$, where the sum runs over all chemical shift values ω_j for the given nucleus in the ensemble of initial chemical shift assignments and $\Delta\omega$ denotes the aforementioned chemical shift tolerance. NOESY cross-peaks are assigned automatically on the basis of the consensus chemical shift assignments and the same peak lists and chemical shift tolerance values used already for the chemical shift assignment, as described above.

A complete FLYA calculation comprises three stages. In the first stage, the chemical shifts and protein structures are generated *de novo* (stage I). In the next stages (stages II and III), the structures generated by the preceding stage are used as additional input for the determination of chemical shift assignments. Stages II and III are particularly important for aromatic residues and other resonances whose assignment rely on through-space NOESY information. At the end of the third stage, the 20 final CYANA conformers with the lowest target function values are subjected to restrained energy minimization in explicit solvent against the AMBER force field (Cornell *et al.* 1995) using the program OPALp (Koradi *et al.* 2000; Luginbühl *et al.* 1996). The complete procedure is driven by the NMR structure calculation program CYANA, which is also used for parallelization of all time-consuming steps. The performance of the FLYA algorithm can be monitored at different steps of the procedure by quality measures that can be computed without referring to external reference assignments or structures (López-Méndez & Güntert, 2006).

Automated signal identification can be achieved with higher reliability for the fewer, sharper and more intense peaks of SAIL proteins. The danger of making erroneous assignments decreases with the number of nuclei and peaks to assign, and less spin diffusion allows NOEs to be interpreted more quantitatively. As a result of the superior quality of the SAIL NMR spectra, reliable fully automated analysis of the NMR spectra and structure calculation are possible using fewer input spectra than with conventional uniformly ^{13}C/^{15}N-labeled proteins. The purely computational SAIL-FLYA method is thus suitable to substitute all manual spectra analysis and overcomes a major efficiency limitation of the NMR method for protein structure determination.

6. SAIL applications

6.1 Ubiquitin

With SAIL-FLYA, fully automated analyses of the NMR spectra and structure calculations of the 7-kDa protein ubiquitin were possible from far fewer input spectra than with conventional uniformly ^{13}C/^{15}N-labeled proteins (Ikeya *et al.* 2009). FLYA calculations with SAIL ubiquitin, using a single 3D 'through-bond' spectrum (and 2D HSQC spectra) in addition to the ^{13}C- and ^{15}N-edited NOESY spectra for conformational restraints, yielded structures with an accuracy of 0·83–1·15 Å for the backbone RMSD to the conventionally determined solution structure of SAIL ubiquitin.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

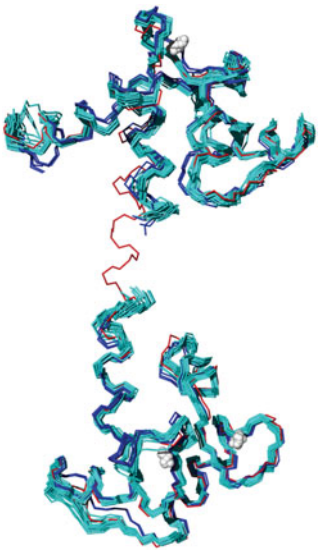


Fig. 16. SAIL-calmodulin solution structure (Kainosho *et al.* 2006) (backbone in cyan, Ca^{2+} in white), calmodulin X-ray structure (Chattopadhyaya *et al.* 1992) (red), and three solution conformers of uniformly labeled calmodulin determined from residual dipolar coupling data (Chou *et al.* 2001) (blue). Superpositions of the calmodulin solution conformers onto the X-ray structures were performed separately for the two flexibly connected domains.

6.2 Calmodulin

Structures of the calcium-bound form of the 17-kDa protein calmodulin were determined previously by crystallography (Chattopadhyaya *et al.* 1992) and NMR. The NMR structure (Chou *et al.* 2001) was based on residual dipolar couplings (RDCs), measured mainly for the polypeptide backbone, and on the crystal structure. The SAIL-calmodulin NMR structure, the first one for calcium-bound calmodulin based exclusively on NOEs, provided detailed information on the side-chain conformations. It is in close agreement with the crystal structure and the RDC-based backbone structure (Fig. 16).

6.3 *E. coli* peptidyl-prolyl *cis-trans* isomerase b

E. coli peptidyl-prolyl *cis-trans* isomerase b (EPPIb) is an 18·2-kDa protein, for which the backbone assignment was achieved using a uniformly ^{13}C- and ^{15}N-double-labeled sample prepared by the cell-free method (Kariya *et al.* 2000). The structure in solution was determined by the SAIL method, and the NOEs related to the 12 Phe and 1 Tyr aromatic rings were carefully analyzed using various types of SAIL Phe/Tyr residues shown in Fig. 13 (Takeda *et al.* 2009). Since the δ, ε, and ζ proton signals of Phe/Tyr residues give rise to complementary NOE restraints, the concomitant use of various types of SAIL-Phe and SAIL-Tyr can result in more accurate protein structures than those obtained with conventional uniformly ^{13}C, ^{15}N-labeled proteins.

The use of various SAIL Phe and Tyr residues simplified the aromatic ring signals of EPPIb. 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 (Kariya *et al.* 2000). For the δ- and ε-SAIL Phe-labeled EPPIb proteins, the signals of Phe 27, Phe 110 and Phe 123 were broadened beyond detection due to intermediate ring flipping rates of these

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

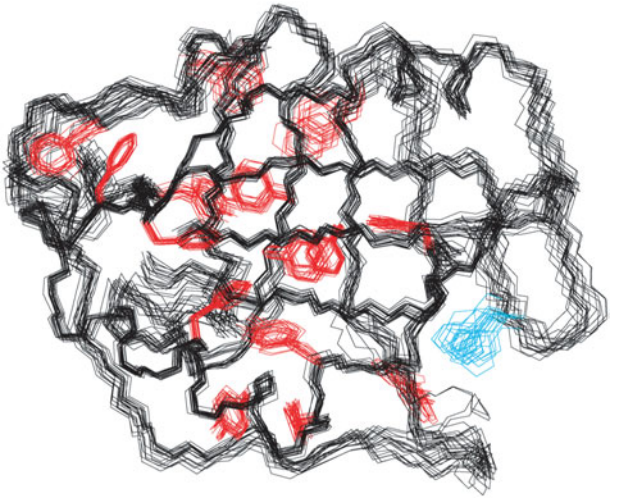


Fig. 17. Structure of *E. coli* peptidyl-prolyl *cis-trans* isomerase b (EPPIb) determined by the combined use of all NOEs from SAIL EPPIb labeled with δ-SAIL Phe/Tyr, ε-SAIL Phe/Tyr, and ζ-SAIL Phe and ε-SAIL Tyr. Standard SAIL amino acids were used for the other residues. Phenylalanine and tyrosine residues are colored red and His-147 is colored blue (Takeda *et al.* 2009).

residues at 30 °C, except for a single, very weak peak assigned to the δ-^{13}C/^{1}H signal of Phe 110. At 40 °C, this weak δ-^{13}C/^{1}H cross-peak became sharper due to the increased ring flipping rate of Phe 110. On the contrary, 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 direction of the ζ ^{13}C-^{1}H bond coincides with the C^{β}-C^{γ} axis.

NOE data were collected for 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, while standard SAIL amino acids were used for the other residues (Takeda *et al.* 2009). 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 about 50. However, about half of the NOEs acquired for the δ-SAIL Phe/Tyr combination were intra-residual or sequential and thus contain limited information for the structure calculation. On the contrary, similar numbers of NOEs between sequentially distant residues were observed for all three combinations. EPPIb structures calculated with CYANA converged well for all three different combinations of SAIL Phe/Tyr. The precision of the structure was slightly lower for the δ-SAIL Phe/Tyr combination as compared with other structures, especially for two loop regions on the protein surface. This local structural difference is due to a complete lack of NOE restraints between the rings of Phe 55 and His 147, 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 Phe 27, Phe 110 and Phe 123, which is due to intermediate ring flipping rates, is clearly reflected in the ill-defined ring positions of these three Phe residues, although their backbone conformation was defined by other NOEs. The structure that was determined by the simultaneous use of all NOEs observed with the aforementioned three combinations of SAIL Phe/Tyr yielded a well-defined structure (Fig. 17) with an RMSD to the crystal structure of EPPIb (Edwards *et al.* 1997) of 1·43 Å for the

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

backbone atoms excluding unstructured loop regions. 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 structure determinations.

6.4 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 homodimer with a molecular weight of 28 kDa. When using uniformly labeled material, the structure determination of the SARS-CoV NP CTD in solution had been impeded by the poor quality of NMR spectra, especially for aromatic resonances. The SAIL protein yielded less crowded and better resolved spectra than uniform ^{13}C and ^{15}N labeling and enabled the homodimeric solution structure of this protein to be determined (Takeda *et al.* 2008a). The NMR structure is almost identical to the previously solved crystal structure, except for a disordered putative RNA-binding domain at the N-terminus.

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

The product of gene At3g16450.1 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 myrosinase-dependent defense against pathogens. Many myrosinase-binding proteins and myrosinase-binding protein-related proteins are composed of repeated homologous sequences with 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 3D structure of the At3g16450.1 protein from *Arabidopsis*, which consists of two tandem repeats, could be solved with SAIL (Takeda *et al.* 2008b). 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 low concentration (0·2 mM) of the protein product. Using NOESY data, the 3D 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.

6.6 Maltodextrin-binding protein

The solution structure of the 41-kDa maltodextrin-binding protein (MBP) was determined to prove that SAIL can solve the structures of proteins that are much larger than those accessible by NMR based on uniformly labeled samples (Kainosho *et al.* 2006). All peaks benefited from improved sensitivity and resolution and allowed the signals of SAIL-calmodulin and SAIL-MBP to be assigned readily by established methods (Cavanagh *et al.* 2006). Side-chain assignments were determined completely from the analysis of two data sets: HCCH-TOCSY data provided connectivities among all side-chain signals and HCCH-COSY data were used to identify the spin systems. The expected ^{1}H, ^{13}C and ^{15}N chemical shifts could be assigned 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 in the region of residues 229–241 that has been shown to

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

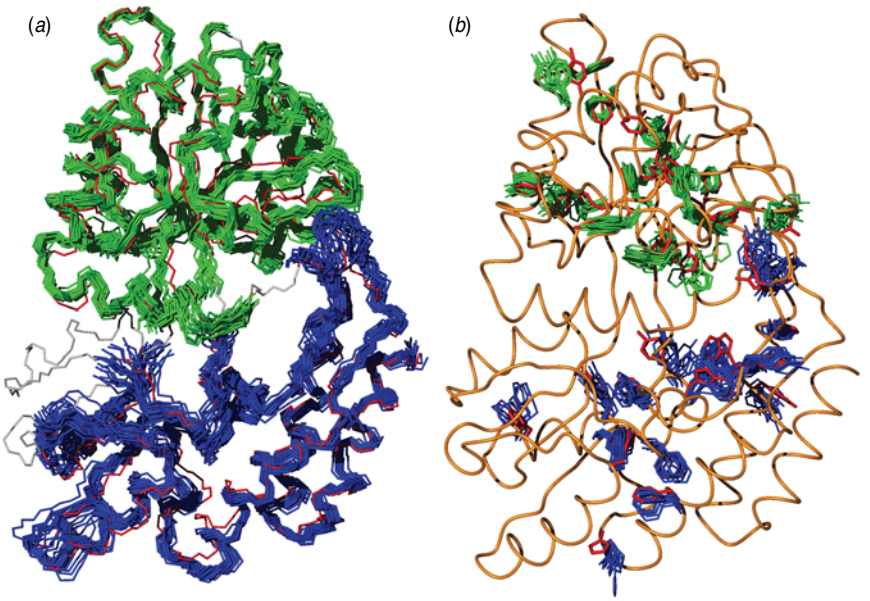


Fig. 18. Solution structure (Kainosho *et al.* 2006) and crystal structure (Sharff *et al.* 1993) of the maltose-binding protein (MBP). (*a*) Backbone of the N-terminal domain (SAIL-MBP in green, X-ray in red) and the C-terminal domain (SAIL-MBP in blue, X-ray in red). (*b*) Aromatic side-chains (SAIL-MBP in green and blue, X-ray in red), and backbone ribbon representation of the X-ray structure (gold). Superpositions of the MBP solution conformers onto the X-ray structures were performed separately for the two flexibly connected domains.

interact with the bound cyclodextrin (Sharff *et al.* 1993). Because of conformational heterogeneity, most of the expected resonances of residues 229–241 could not be assigned in earlier studies either (Mueller *et al.* 2000).

Distance restraints for the structure calculation were obtained from 3D ^{15}N- and ^{13}C-edited NOESY spectra (Marion *et al.* 1989) and from a 2D NOESY spectrum for the aromatic region. These spectra were simplified by a decreased number of signals as expected. Because of reduced spin-diffusion, maximum NOE intensities for SAIL proteins were typically reached at 1·5–3 times longer mixing times than with uniform labeling. A dense network of NOEs including 949 non-redundant, long-range distance restraints was established for SAIL-MBP. Of 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 (Güntert, 2003; Herrmann *et al.* 2002). The SAIL-MBP solution structures show good quality in terms of the agreement with the experimental data and other validation parameters.

The solution structure of the 41-kDa SAIL-MBP coincides closely with the crystal structure (Sharff *et al.* 1993) that was determined previously under slightly different conditions (Fig. 18). The 41-kDa SAIL-MBP solution structure is of similar precision and accuracy as those of smaller proteins, and the structural statistics are comparable to those commonly found in NMR structure determinations of smaller proteins. Previously (Mueller *et al.* 2000), a global fold of MBP was determined by NMR on the basis of NOEs between amide and methyl protons, residual dipolar

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

couplings for the polypeptide backbone and hydrogen bond restraints. That NMR study could determine well the global fold of the polypeptide backbone and the conformations of the methyl-containing side-chains of valine, leucine and isoleucine, but the approach used could not provide direct structural information on the other side-chains.

7. Conclusions and Outlook

With SAIL, the size limitation for high-quality NMR solution structures has been cleared up to at least 40 kDa, investigations of hitherto experimentally inaccessible aspects of protein structure, dynamics and interactions have become possible, and fully automated spectra analysis and structure determination of proteins has been greatly facilitated. A three to seven times better signal-to-noise ratio, reduced spectral crowding, and the simplification of coupling patterns are unique improvements that are achieved by using SAIL instead of uniform ^{13}C and ^{15}N labeling. The application of SAIL has hardly any drawbacks for NMR experiments and spectra analysis and would thus be the method of choice for stable isotope labeling in biomolecular NMR. The two factors that are currently limiting its ubiquitous application are the cost for supplying SAIL amino acids and the requirement to establish the efficient cell-free production of the proteins of interest. Both issues can be overcome, the first one by economies of scale in the commercial synthesis of the SAIL amino acids and the second by the more and more widespread proficiency of NMR research groups in cell-free protein expression techniques. On the other hand, studies that require only a small number of SAIL-labeled amino acid types can conveniently be carried out using cell-based expression by auxotrophic bacterial strains.

The development of the SAIL technology shows that sophisticated stable isotope labeling patterns that are designed for optimal results rather than synthetic convenience can be realized and used in practice. The capabilities in chemical and enzymatic synthesis that have been acquired for SAIL can now be turned to more ambitious goals, including, for instance, tailor-made optimal isotope labeling patterns for membrane proteins and solid-state NMR.

8. Acknowledgments

Financial support for the development of the SAIL technology by the Core Research for Evolutional Science and Technology (CREST) program of the Japan Science and Technology Corporation (JST), the Targeted Protein Research Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), Grants-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS) and by the Volkswagen Foundation is gratefully acknowledged.

9. References

Aberhart, D. J. & Russell, D. J. (1984). Steric course of ketopantoate hydroxymethyltransferase in *E. coli*. *Journal of the American Chemical Society* **106**, 4902–4906.

Aghazadeh, B., Zhu, K., Kubiseski, T. J., Liu, G. A., Pawson, T., Zheng, Y. & Rosen, M. K. (1998). Structure and mutagenesis of the Dbl homology domain. *Nature Structural Biology* **5**, 1098–1107.

Akke, M., Carr, P. A. & Palmer, A. G. (1994). Heteronuclear-correlation NMR spectroscopy with simultaneous isotope filtration, quadrature detection, and sensitivity enhancement using χ rotations. *Journal of Magnetic Resonance Series B* **104**, 298–302.

Arata, Y., Kato, K., Takahashi, H. & Shimada, I. (1994). Nuclear magnetic resonance study of antibodies – a

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

multinuclear approach. *Methods in Enzymology* **239**, 440–464.

Arnesano, F., Banci, L., Bertini, I., Felli, I. C., Luchinat, C. & Thompsett, A. R. (2003). A strategy for the NMR characterization of type II copper(II) proteins: the case of the copper trafficking protein CopC from *Pseudomonas syringae*. *Journal of the American Chemical Society* **125**, 7200–7208.

Arnold, L. D., Kalantar, T. H. & Vederas, J. C. (1985). Conversion of serine to stereochemically pure β-substituted α-amino acids via β-lactones. *Journal of the American Chemical Society* **107**, 7105–7109.

Arnold, L. D., May, R. G. & Vederas, J. C. (1988). Synthesis of optically pure α-amino acids via salts of α-amino-β-propiolactone. *Journal of the American Chemical Society* **110**, 2237–2241.

Arrowsmith, C. H. & Wu, Y. S. (1998). NMR of large (>25 kDa) proteins and protein complexes. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 277–286.

Atreya, H. S. & Chary, K. V. R. (2001). Selective 'unlabeling' of amino acids in fractionally ^{13}C labeled proteins: an approach for stereospecific NMR assignments of CH_3 groups in Val and Leu residues. *Journal of Biomolecular NMR* **19**, 267–272.

Axelsson, B. S., Otoole, K. J., Spencer, P. A. & Young, D. W. (1991). A versatile synthesis of stereospecifically labelled D-amino acids and related enzyme inhibitors. *Journal of the Chemical Society-Chemical Communications* 1085–1086.

Bartels, C., Billeter, M., Güntert, P. & Wüthrich, K. (1996). Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *Journal of Biomolecular NMR* **7**, 207–213.

Bartels, C., Güntert, P., Billeter, M. & Wüthrich, K. (1997). GARANT – a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *Journal of Computational Chemistry* **18**, 139–149.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *Journal of Magnetic Resonance* **88**, 425–431.

Bax, A. & Pochapsky, S. S. (1992). Optimized recording of heteronuclear multidimensional NMR spectra using pulsed field gradients. *Journal of Magnetic Resonance* **99**, 638–643.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242.

Bermel, W., Bertini, I., Duma, L., Felli, I. C., Emsley, L., Pierattelli, R. & Vasos, P. R. (2005). Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angewandte Chemie – International Edition* **44**, 3089–3092.

Bertini, I., Duma, L., Felli, I. C., Fey, M., Luchinat, C., Pierattelli, R. & Vasos, P. R. (2004). A heteronuclear direct-detection NMR spectroscopy experiment for protein-backbone assignment. *Angewandte Chemie – International Edition* **43**, 2257–2259.

Betton, J. M. (2003). Rapid translation system (RTS): a promising alternative for recombinant protein production. *Current Protein & Peptide Science* **4**, 73–80.

Beyer, J., Lang-Fugmann, S., Mühlbauer, A. & Steglich, W. (1998). A convenient synthesis of 4-hydroxy[1-^{13}C]benzoic acid and related ring-labelled phenolic compounds. *Synthesis-Stuttgart* 1047–1051.

Brenzel, S., Kurpiers, T. & Mootz, H. D. (2006). Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split Ssp DnaB intein and comparison to the split Sce VMA intein. *Biochemistry* **45**, 1571–1578.

Burk, M. J. (1991). C_2-symmetric bis(phospholanes) and their use in highly enantioselective hydrogenation reactions. *Journal of the American Chemical Society* **113**, 8518–8519.

Burk, M. J., Feaster, J. E., Nugent, W. A. & Harlow, R. L. (1993). Preparation and use of C_2-symmetric bis(phospholanes): production of α-amino acid derivatives via highly enantioselective hydrogenation reactions. *Journal of the American Chemical Society* **115**, 10125–10138.

Burk, M. J., Kalberg, C. S. & Pizzano, A. (1998). Rh-DuPHOS-catalyzed enantioselective hydrogenation of enol esters. Application to the synthesis of highly enantioenriched α-hydroxy esters and 1,2-diols. *Journal of the American Chemical Society* **120**, 4345–4353.

Burz, D. S., Dutta, K., Cowburn, D. & Shekhtman, A. (2006). Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nature Methods* **3**, 91–93.

Busche, A. E., Aranko, A. S., Talebzadeh-Farooji, M., Bernhard, F., Dötsch, V. & Iwai, H. (2009). Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. *Angewandte Chemie* (*English Edition*) **48**, 6128–6131.

Campbell, I. D., Dobson, C. M., Moore, G. R., Perkins, S. J. & Williams, R. J. (1976). Temperature dependent molecular motion of a tyrosine residue of ferrocytochrome C. *FEBS Letters* **70**, 96–100.

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Skelton, N. J. & Rance, M. (2006). *Protein NMR Spectroscopy. Principles and Practice*, 2nd edn. San Diego, CA: Academic Press.

Cavanagh, J., Palmer, A. G., Wright, P. E. & Rance, M. (1991). Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy. *Journal of Magnetic Resonance* **91**, 429–436.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

CHANDONIA, J. M. & BRENNER, S. E. (2006). The impact of structural genomics: expectations and outcomes. *Science* **311**, 347–351.

CHATTOPADHYAYA, R., MEADOR, W. E., MEANS, A. R. & QUIOCHO, F. A. (1992). Calmodulin structure refined at 1·7 Å resolution. *Journal of Molecular Biology* **228**, 1177–1192.

CHOU, J. J., LI, S. P., KLEE, C. B. & BAX, A. (2001). Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. *Nature Structural Biology* **8**, 990–997.

CHRUNYK, B. A., EVANS, J., LILLQUIST, J., YOUNG, P. & WETZEL, R. (1993). Inclusion body formation and protein stability in sequence variants of interleukin-1β. *Journal of Biological Chemistry* **268**, 18053–18061.

CLEMENS, M. J. & PRUJIN, G. J. (1999). *Protein Synthesis in Eukaryotic Cell-Free Systems*. New York: Oxford University Press.

CORNELL, W. D., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., SPELLMEYER, D. C., FOX, T., CALDWELL, J. W. & KOLLMAN, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179–5197.

COUGHLIN, P. E., ANDERSON, F. E., OLIVER, E. J., BROWN, J. M., HOMANS, S. W., POLLAK, S. & LUSTBADER, J. W. (1999). Improved resolution and sensitivity of triple-resonance NMR methods for the structural analysis of proteins by use of a backbone-labeling strategy. *Journal of the American Chemical Society* **121**, 11871–11874.

COWBURN, D., SHEKHTMAN, A., XU, R., OTTESEN, J. J. & MUIR, T. W. (2004). Segmental isotopic labeling for structural biological applications of NMR. *Methods in Molecular Biology* **278**, 47–56.

CRESPI, H. L., ROSENBERG, R. M. & KATZ, J. J. (1968). Proton magnetic resonance of proteins fully deuterated except for ^{1}H-leucine side chains. *Science* **161**, 795–796.

DAVANLOO, P., ROSENBERG, A. H., DUNN, J. J. & STUDIER, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences USA* **81**, 2035–2039.

DING, K. & GRONENBORN, A. M. (2004). Sensitivity-enhanced IPAP experiments for measuring one-bond $^{13}C'$-$^{13}C^{\alpha}$ and $^{13}C^{\alpha}$-$^1H^{\alpha}$ residual dipolar couplings in proteins. *Journal of Magnetic Resonance* **167**, 253–258.

DUAN, X. Q., GIMBLE, F. S. & QUIOCHO, F. A. (1997). Crystal structure of PI-Scel, a homing endonuclease with protein splicing activity. *Cell* **89**, 555–564.

EDISON, A. S., ABILDGAARD, F., WESTLER, W. M., MOOBERRY, E. S. & MARKLEY, J. L. (1994). Practical introduction to theory and implementation of multinuclear, multidimensional nuclear magnetic resonance experiments. *Methods in Enzymology* **239**, 3–79.

EDWARDS, K. J., OLLIS, D. L. & DIXON, N. E. (1997). Crystal structure of cytoplasmic *Escherichia coli* peptidyl-prolyl isomerase: Evidence for decreased mobility of loops upon complexation. *Journal of Molecular Biology* **271**, 258–265.

ERLENMEYER, E. (1893). Über die Condensation der Hippursäure mit Phtalsäureanhydrid und mit Benzaldehyd. *Justus Liebigs Annalen der Chemie* **275**, 1–20.

ERNST, R. R., BODENHAUSEN, G. & WOKAUN, A. (1987). *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Clarendon Press.

FERNÁNDEZ, C., HILTY, C., WIDER, G., GÜNTERT, P. & WÜTHRICH, K. (2004). NMR structure of the integral membrane protein OmpX. *Journal of Molecular Biology* **336**, 1211–1221.

FESIK, S. W. & ZUIDERWEG, E. R. P. (1988). Heteronuclear 3-dimensional NMR spectroscopy – a strategy for the simplification of homonuclear two-dimensional NMR spectra. *Journal of Magnetic Resonance* **78**, 588–593.

FIAUX, J., BERTELSEN, E. B., HORWICH, A. L. & WUTHRICH, K. (2002). NMR analysis of a 900K GroEL-GroES complex. *Nature* **418**, 207–211.

FOLMER, R. H. A., HILBERS, C. W., KONINGS, R. N. H. & HALLENGA, K. (1995). A ^{13}C double-filtered NOESY with strongly reduced artefacts and improved sensitivity. *Journal of Biomolecular NMR* **5**, 427–432.

GANI, D. & YOUNG, D. W. (1983). Synthesis of L-serine stereospecifically labelled at C-3 with deuterium. *Journal of the Chemical Society – Perkin Transactions* **1** 2393–2398.

GARDNER, K. H. & KAY, L. E. (1997). Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ1 Methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society* **119**, 7599–7600.

GARDNER, K. H. & KAY, L. E. (1998). The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* **27**, 357–406.

GARDNER, K. H., ZHANG, X. C., GEHRING, K. & KAY, L. E. (1998). Solution NMR studies of a 42 kDa *Escherichia coli* maltose binding protein β-cyclodextrin complex: chemical shift assignments and analysis. *Journal of the American Chemical Society* **120**, 11738–11748.

GARRETT, D. S., SEOK, Y. J., LIAO, D. I., PETERKOFSKY, A., GRONENBORN, A. M. & CLORE, G. M. (1997). Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry* **36**, 2517–2530.

GOFF, S. A. & GOLDBERG, A. L. (1987). An increased content of protease La, the lon gene product, increases protein degradation and blocks growth in Escherichia coli. *Journal of Biological Chemistry* **262**, 4508–4515.

GOTO, N. K., GARDNER, K. H., MUELLER, G. A., WILLIS, R. C. & KAY, L. E. (1999). A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins. *Journal of Biomolecular NMR* **13**, 369–374.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

Goto, N. K. & Kay, L. E. (2000). New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Current Opinion in Structural Biology* **10**, 585–592.

Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1987a). Novel 3-dimensional NMR techniques for studies of peptides and biological macromolecules. *Journal of the American Chemical Society* **109**, 7227–7228.

Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1987b). A practical approach to 3-dimensional NMR spectroscopy. *Journal of Magnetic Resonance* **73**, 574–579.

Gronwald, W. & Kalbitzer, H. R. (2004). Automated structure determination of proteins by NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **44**, 33–96.

Grzesiek, S., Anglister, J., Ren, H. & Bax, A. (1993). ^{13}C line narrowing by ^{2}H decoupling in $^{2}H/^{13}C/^{15}N$-enriched proteins. Application to triple-resonance 4D J-connectivity of sequential amides. *Journal of the American Chemical Society* **115**, 4369–4370.

Grzesiek, S. & Bax, A. (1992a). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114**, 6291–6293.

Grzesiek, S. & Bax, A. (1992b). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance* **99**, 201–207.

Grzesiek, S., Cordier, F., Jaravine, V. & Barfield, M. (2004). Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy* **45**, 275–300.

Güntert, P. (2003). Automated NMR protein structure calculation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **43**, 105–125.

Güntert, P. (2009). Automated structure determination from NMR spectra. *European Biophysics Journal* **38**, 129–143.

Güntert, P., Berndt, K. D. & Wüthrich, K. (1993). The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. *Journal of Biomolecular NMR* **3**, 601–606.

Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. *Journal of the American Chemical Society* **111**, 3997–4004.

Güntert, P., Braun, W. & Wüthrich, K. (1991a). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *Journal of Molecular Biology* **217**, 517–530.

Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology* **273**, 283–298.

Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1991b). Structure determination of the Antp(C39S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *Journal of Molecular Biology* **217**, 531–540.

Güntert, P. & Wüthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *Journal of Biomolecular NMR* **1**, 447–456.

Guignard, L., Ozawa, K., Pursglove, S. E., Otting, G. & Dixon, N. E. (2002). NMR analysis of in vitro-synthesized proteins without purification: a high-throughput approach. *FEBS Letters* **524**, 159–162.

Henrich, B., Lubitz, W. & Plapp, R. (1982). Lysis of *Escherichia coli* by induction of cloned phi X174 genes. *Molecular and General Genetics* **185**, 493–497.

Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology* **319**, 209–227.

Hirata, R., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990). Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H^{+}-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **265**, 6726–6733.

Huang, Y. P. J., Moseley, H. N. B., Baran, M. C., Arrowsmith, C., Powers, R., Tejero, R., Szyperski, T. & Montelione, G. T. (2005). An integrated platform for automated analysis of protein NMR structures. *Methods in Enzymology* **394**, 111–141.

Ikeya, T., Takeda, M., Yoshida, H., Terauchi, T., Jee, J., Kainosho, M. & Güntert, P. (2009). Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system. *Journal of Biomolecular NMR* **44**, 261–272.

Ikeya, T., Terauchi, T., Güntert, P. & Kainosho, M. (2006). Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA. *Magnetic Resonance in Chemistry* **44**, S152–S157.

Ikura, M. & Bax, A. (1992). Isotope-filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *Journal of the American Chemical Society* **114**, 2433–2440.

Ikura, M., Kay, L. E. & Bax, A. (1990). A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

IKURA, M., SPERA, S., BARBATO, G., KAY, L. E., KRINKS, M. & BAX, A. (1991). Secondary structure and side-chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR-spectroscopy. *Biochemistry* **30**, 9216–9228.

ISHIMA, R., LOUIS, J. M. & TORCHIA, D. A. (2001). Optimized labeling of $^{13}CHD_2$ methyl isotopomers in perdeuterated proteins: potential advantages for ^{13}C relaxation studies of methyl dynamics of larger proteins. *Journal of Biomolecular NMR* **21**, 167–171.

IWAI, H. & PLÜCKTHUN, A. (1999). Circular β-lactamase: stability enhancement by cyclizing the backbone. *FEBS Letters* **459**, 166–172.

IWAI, H., ZÜGER, S., JIN, J. & TAM, P. H. (2006). Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Letters* **580**, 1853–1858.

JARDETZKY, O. & ROBERTS, G. C. K. (1981). *NMR in Molecular Biology*. New York: Academic Press.

JEE, J. & GÜNTERT, P. (2003). Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *Journal of Structural and Functional Genomics* **4**, 179–189.

JOHNSON, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods in Molecular Biology* **278**, 313–352.

KAINOSHO, M. (1997). Isotope labelling of macromolecules for structural determinations. *Nature Structural Biology* **4**, 858–861.

KAINOSHO, M. & AJISAKA, K. (1975). Conformational analysis of amino acids and peptides using specific isotope substitution. II. Conformation of serine, tyrosine, phenylalanine, aspartic acid, asparagine, and asparatic acid β-methyl ester in various ionization states. *Journal of the American Chemical Society* **97**, 5630–5631.

KAINOSHO, M., TORIZAWA, T., IWASHITA, Y., TERAUCHI, T., ONO, A. M. & GÜNTERT, P. (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57.

KAINOSHO, M. & TSUJI, T. (1982). Assignment of the three methionyl carbonyl carbon resonances in *Streptomyces* subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. *Biochemistry* **21**, 6273–6279.

KALBITZER, H. R., LEBERMAN, R. & WITTINGHOFER, A. (1985). ^{1}H-NMR spectroscopy on elongation factor Tu from *Escherichia coli* – resolution enhancement by perdeuteration. *FEBS Letters* **180**, 40–42.

KANE, P. M., YAMASHIRO, C. T., WOLCZYK, D. F., NEFF, N., GOEBL, M. & STEVENS, T. H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H^{+}-adenosine triphosphatase *Science* **250**, 651–657.

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). *Journal of Biomolecular NMR* **18**, 75–76.

KAY, L. E. (2005). NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance* **173**, 193–207.

KAY, L. E., IKURA, M., TSCHUDIN, R. & BAX, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance* **89**, 496–514.

KAY, L. E., KEIFER, P. & SAARINEN, T. (1992a). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* **114**, 10663–10665.

KAY, L. E., NICHOLSON, L. K., DELAGLIO, F., BAX, A. & TORCHIA, D. A. (1992b). Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T_1 and T_2 values in proteins. *Journal of Magnetic Resonance* **97**, 359–375.

KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., WYCKOFF, H. & PHILLIPS, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666.

KIGAWA, T., MUTO, Y. & YOKOYAMA, S. (1995). Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *Journal of Biomolecular NMR* **6**, 129–134.

KIGAWA, T., YABUKI, T., YOSHIDA, Y., TSUTSUI, M., ITO, Y., SHIBATA, T. & YOKOYAMA, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Letters* **442**, 15–19.

KIM, D. M., KIGAWA, T., CHOI, C. Y. & YOKOYAMA, S. (1996). A highly efficient cell-free protein synthesis system from *Escherichia coli*. *European Journal of Biochemistry* **239**, 881–886.

KIM, D. M. & SWARTZ, J. R. (2000). Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnology Progress* **16**, 385–390.

KLABUNDE, T., SHARMA, S., TELENTI, A., JACOBS, W. R. & SACCHETTINI, J. C. (1998). Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Structural Biology* **5**, 31–36.

KLAMMT, C., LÖHR, F., SCHÄFER, B., HAASE, W., DÖTSCH, V., RÜTERJANS, H., GLAUBITZ, C. & BERNHARD, F. (2004). High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry* **271**, 568–580.

KOBAYASHI, M., YAGI, H., YAMAZAKI, T., YOSHIDA, M. & AKUTSU, H. (2008). Dynamic inter-subunit interactions in thermophilic F-1-ATPase subcomplexes studied by cross-correlated relaxation-enhanced polarization transfer NMR. *Journal of Biomolecular NMR* **40**, 165–174.

KORADI, R., BILLETER, M., ENGELI, M., GÜNTERT, P. & WÜTHRICH, K. (1998). Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY. *Journal of Magnetic Resonance* **135**, 288–297.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

KORADI, R., BILLETER, M. & GÜNTERT, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. *Computer Physics Communications* **124**, 139–147.

KRAMER, G., KUDLICKI, W. & HARDESTY, B. (1999). *Cell-Free Coupled Transcription-Translation Systems from Escherichia coli*. New York: Oxford University Press.

LANG, M., LANG-FUGMANN, S. & STEGLICH, W. (2002). *Organic Syntheses* **78**, 113–122.

LEE, W., REVINGTON, M. J., ARROWSMITH, C. & KAY, L. E. (1994). A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Letters* **350**, 87–90.

LEMASTER, D. M., LAIUPPA, J. C. & KUSHLAN, D. M. (1994). Differential deuterium isotope shifts and one bond ^{1}H-^{13}C scalar couplings in the conformational analysis of protein glycine residues. *Journal of Biomolecular NMR* **4**, 863–870.

LEMASTER, D. M. & RICHARDS, F. M. (1988). NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. *Biochemistry* **27**, 142–150.

LI, G. G., PATEL, D. & HRUBY, V. J. (1993). An efficient procedure for the demethylation of aryl-methyl ethers in optically pure unusual amino acids. *Tetrahedron Letters* **34**, 5393–5396.

LIAN, L. Y. & MIDDLETON, D. A. (2001). Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39**, 171–190.

LÓPEZ-MÉNDEZ, B. & GÜNTERT, P. (2006). Automated protein structure determination from NMR spectra. *Journal of the American Chemical Society* **128**, 13112–13122.

LUGINBÜHL, P., GÜNTERT, P., BILLETER, M. & WÜTHRICH, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *Journal of Biomolecular NMR* **8**, 136–146.

MACHONKIN, T. E., WESTLER, W. M. & MARKLEY, J. L. (2002). ^{13}C{^{13}C} 2D NMR: a novel strategy for the study of paramagnetic proteins with slow electronic relaxation rates. *Journal of the American Chemical Society* **124**, 3204–3205.

MADIN, K., SAWASAKI, T., OGASAWARA, T. & ENDO, Y. (2000). A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proceedings of the National Academy of Sciences USA* **97**, 559–564.

MALMODIN, D., PAPAVOINE, C. H. M. & BILLETER, M. (2003). Fully automated sequence-specific resonance assignments of heteronuclear protein spectra. *Journal of Biomolecular NMR* **27**, 69–79.

MARION, D., KAY, L. E., SPARKS, S. W., TORCHIA, D. A. & BAX, A. (1989). Three-dimensional heteronuclear NMR of ^{15}N labeled proteins. *Journal of the American Chemical Society* **111**, 1515–1517.

MARKLEY, J. L., PUTTER, I. & JARDETZKY, O. (1968). High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* **161**, 1249–1251.

MATSUO, H., KUPCE, E., LI, H. J. & WAGNER, G. (1996a). Increased sensitivity in HNCA and HN(CO)CA experiments by selective C^{β} decoupling. *Journal of Magnetic Resonance Series B* **113**, 91–96.

MATSUO, H., KUPCE, E. & WAGNER, G. (1996b). Resolution and sensitivity gain in HCCH-TOCSY experiments by homonuclear C^{β} decoupling. *Journal of Magnetic Resonance Series B* **113**, 190–194.

MAURIZI, M. R. (1987). Degradation in vitro of bacteriophage lambda N protein by Lon protease from *Escherichia coli*. *Journal of Biological Chemistry* **262**, 2696–2703.

MCINTOSH, L. P. & DAHLQUIST, F. W. (1990). Biosynthetic incorporation of N-15 and C-13 for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Quarterly Reviews of Biophysics* **23**, 1–38.

MEDEK, A., OLEJNICZAK, E. T., MEADOWS, R. P. & FESIK, S. W. (2000). An approach for high-throughput structure determination of proteins by NMR spectroscopy. *Journal of Biomolecular NMR* **18**, 229–238.

MELACINI, G. (2000). Separation of intra- and intermolecular NOEs through simultaneous editing and *J*-compensated filtering: a 4D quadrature-free constant-time *J*-resolved approach. *Journal of the American Chemical Society* **122**, 9735–9738.

METZLER, W. J., WITTEKIND, M., GOLDFARB, V., MUELLER, L. & FARMER II, B. T. (1996). Incorporation of ^{1}H/^{13}C/^{15}N-{Ile, Leu, Val} into a perdeuterated, ^{15}N-labeled protein: potential in structure determination of large proteins by NMR. *Journal of the American Chemical Society* **118**, 6800–6801.

MONTELIONE, G. T., ZHENG, D. Y., HUANG, Y. P. J., GUNSALUS, K. C. & SZYPERSKI, T. (2000). Protein NMR spectroscopy in structural genomics. *Nature Structural Biology* **7**, 982–985.

MUELLER, G. A., CHOY, W. Y., YANG, D., FORMAN-KAY, J. D., VENTERS, R. A. & KAY, L. E. (2000). Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *Journal of Molecular Biology* **300**, 197–212.

MUONA, M., ARANKO, A. S. & IWAI, H. (2008). Segmental isotopic labelling of a multidomain protein by protein ligation by protein prans-splicing. *ChemBioChem: A European Journal of Chemical Biology* **9**, 2958–2961.

MURRAY III, A. & WILLIAMS, D. L. (1958). *Organic Syntheses with Isotopes. Part II: Organic Compounds Labeled with Isotopes of the Halogens, Hydrogen, Nitrogen, Oxygen, Phosphorus and Sulfur*. New York: Interscience.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

NERI, D., SZYPERSKI, T., OTTING, G., SENN, H. & WÜTHRICH, K. (1989). Sterospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510–7516.

NILGES, M. (1995). Calculation of protein structures with ambiguous distance restraints – Automated assignment of ambiguous NOE crosspeaks and disulfide connectivities. *Journal of Molecular Biology* **245**, 645–660.

NISHIYAMA, K., OBA, M., UENO, R., MORITA, A., NAKAMURA, Y. & KAINOSHO, M. (1994). Synthesis of phenylalanines regiospecifically labelled with deuterium in the aromatic ring. *Journal of Labelled Compounds & Radiopharmaceuticals* **34**, 831–837.

O'DONOGHUE, S. I. & NILGES, M. (1999). Calculation of symmetric oligomer structures from NMR data. In *Structure Computation and Dynamics in Protein NMR*, vol. 17 (eds. N. R. Krishna & L. J. Berliner), pp. 131–161. New York: Kluwer Academic/Plenum Publishers.

OBA, M., IWASAKI, A., HITOKAWA, H., IKEGAME, T., BANBA, H., URA, K., TAKAMURA, T. & NISHIYAMA, K. (2006). Preparation of L-serine and L-cystine stereospecifically labeled with deuterium at the β-position. *Tetrahedron-Asymmetry* **17**, 1890–1894.

OBA, M., KOBAYASHI, M., OIKAWA, F., NISHIYAMA, K. & KAINOSHO, M. (2001). Synthesis of $^{13}C/D$ doubly labeled L-leucines: probes for conformational analysis of the leucine side-chain. *Journal of Organic Chemistry* **66**, 5919–5922.

OBA, M., TERAUCHI, T., MIYAKAWA, A., KAMO, H. & NISHIYAMA, K. (1998). Stereoselective deuterium-labelling of diastereotopic methyl and methylene protons of L-leucine. *Tetrahedron Letters* **39**, 1595–1598.

OBA, M., TERAUCHI, T., MIYAKAWA, A. & NISHIYAMA, K. (1999). Asymmetric synthesis of L-proline regio- and stereoselectively labelled wiith deuterium. *Tetrahedron-Asymmetry* **10**, 937–945.

OGURA, K., TERASAWA, H. & INAGAKI, F. (1996). An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *Journal of Biomolecular NMR* **8**, 492–498.

OHKI, S. Y., ETO, M., KARIYA, E., HAYANO, T., HAYASHI, Y., YAZAWA, M., BRAUTIGAN, D. & KAINOSHO, M. (2001). Solution NMR structure of the myosin phosphatase inhibitor protein CPI-17 shows phosphorylation-induced conformational changes responsible for activation. *Journal of Molecular Biology* **314**, 839–849.

OHKI, S. Y. & KAINOSHO, M. (2008). Stable isotope labeling methods for protein NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy* **53**, 208–226.

OJIMA, I., KATO, K., NAKAHASHI, K., FUCHIKAMI, T. & FUJITA, M. (1989). New and effective routes to fluoro analogues of aliphatic and aromatic amino acids. *Journal of Organic Chemistry* **54**, 4511–4522.

OKUMA, K., ONO, A. M., TSUCHIYA, S., OBA, M., NISHIYAMA, K., KAINOSHO, M. & TERAUCHI, T. (2009). Assymetric synthesis of (2*S*,3*R*)- and (2*S*,3*S*)-[2-^{13}C;3-^{2}H] glutamic acid. *Tetrahedron Letters* **50**, 1482–1484.

OLEJNICZAK, E. T., XU, R. X. & FESIK, S. W. (1992). A 4D-HCCH-TOCSY experiment for assigning the side-chain ^{1}H-resonance and ^{13}C-resonance of proteins. *Journal of Biomolecular NMR* **2**, 655–659.

OLLERENSHAW, J. E., TUGARINOV, V., SKRYNNIKOV, N. R. & KAY, L. E. (2005). Comparison of $^{13}CH_3$, $^{13}CH_2D$, and $^{13}CHD_2$ methyl labeling strategies in proteins. *Journal of Biomolecular NMR* **33**, 25–41.

OSCHKINAT, H., CIESLAR, C., HOLAK, T. A., CLORE, G. M. & GRONENBORN, A. M. (1989). Practical and theoretical aspects of 3-dimensional homonuclear Hartmann-Hahn-nuclear Overhauser enhancement spectroscopy of proteins. *Journal of Magnetic Resonance* **83**, 450–472.

OSTLER, G., SOTERIOU, A., MOODY, C. M., KHAN, J. A., BIRDSALL, B., CARR, M. D., YOUNG, D. W. & FEENEY, J. (1993). Stereospecific assignments of the leucine methyl resonances in the ^{1}H NMR spectrum of *Lactobacillus casei* dihydrofolate reductase. *FEBS Letters* **318**, 177–180.

OTOMO, T., ITO, N., KYOGOKU, Y. & YAMAZAKI, T. (1999a). NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* **38**, 16040–16044.

OTOMO, T., TERUYA, K., UEGAKI, K., YAMAZAKI, T. & KYOGOKU, Y. (1999b). Improved segmental isotope labeling of proteins and application to a larger protein. *Journal of Biomolecular NMR* **14**, 105–114.

OTTING, G. & WÜTHRICH, K. (1990). Heteronuclear filters in two-dimensional [^{1}H,^{1}H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Quarterly Reviews of Biophysics* **23**, 39–96.

PALMER, A. G., CAVANAGH, J., BYRD, R. A. & RANCE, M. (1992). Sensitivity improvement in 3-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance* **96**, 416–424.

PALMER, A. G., CAVANAGH, J., WRIGHT, P. E. & RANCE, M. (1991a). Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance* **93**, 151–170.

PALMER, A. G., CAVANAGH, J., WRIGHT, P. E. & RANCE, M. (1991b). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance* **93**, 151–170.

PERVUSHIN, K. & ELETSKY, A. (2003). A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment. *Journal of Biomolecular NMR* **25**, 147–152.

PERVUSHIN, K., RIEK, R., WIDER, G. & WÜTHRICH, K. (1997). Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

biological macromolecules in solution. *Proceedings of the National Academy of Sciences USA* **94**, 12366–12371.

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. *Journal of the American Chemical Society* **120**, 6394–6400.

Pratt, J. M. (1984). *Transcription and Translation: A Practical Approach*. New York: IRL Press.

Prestegard, J. H., Bougault, C. M. & Kishore, A. I. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chemical Reviews* **104**, 3519–3540.

Rajesh, S., Nietlispach, D., Nakayama, H., Takio, K., Laue, E. D., Shibata, T. & Ito, Y. (2003). A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp. *Journal of Biomolecular NMR* **27**, 81–86.

Ramer, S. E., Moore, R. N. & Vederas, J. C. (1986). Mechanism of formation of serine β-lactones by Mitsunobu cyclization: synthesis and use of L-serine stereospecifically labelled with deuteriun at C-3. *Canadian Journal of Chemistry* **64**, 706–713.

Riegel, E. & Zwilgmeyer, F. (1943). *Organic Syntheses* **II**, 126.

Riek, R., Wider, G., Pervushin, K. & Wüthrich, K. (1999). Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proceedings of the National Academy of Sciences USA* **96**, 4918–4923.

Romanelli, A., Shekhtman, A., Cowburn, D. & Muir, T. W. (2004). Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proceedings of the National Academy of Sciences USA* **101**, 6397–6402.

Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. & Kay, L. E. (1996). Selective methyl group protonation of perdeuterated proteins. *Journal of Molecular Biology* **263**, 627–636.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13585–13590.

Santoro, J. & King, G. C. (1992). A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. *Journal of Magnetic Resonance* **97**, 202–207.

Seeholzer, S. H., Cohn, M., Putkey, J. A., Means, A. R. & Crespi, H. L. (1986). NMR studies of a complex of deuterated calmodulin with melittin. *Proceedings of the National Academy of Sciences USA* **83**, 3634–3638.

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. & Kay, L. E. (1996). Assignment of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and HN resonances in an ^{15}N, ^{13}C, ^{2}H labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and ^{2}H-decoupling. *Journal of the American Chemical Society* **118**, 6570–6579.

Sharff, A. J., Rodseth, L. E. & Quiocho, F. A. (1993). Refined 1·8-Å structure reveals the mode of binding of β-cyclodextrin to the maltodextrin binding protein. *Biochemistry* **32**, 10553–10559.

Shi, J., Pelton, J. G., Cho, H. S. & Wemmer, D. E. (2004). Protein signal assignments using specific labeling and cell-free synthesis. *Journal of Biomolecular NMR* **28**, 235–247.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751–755.

Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.

Skrisovska, L. & Allain, F. H. T. (2008). Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: application to the RRMs of Npl3p and hnRNP L. *Journal of Molecular Biology* **375**, 151–164.

Smith, B. O., Ito, Y., Raine, A., Teichmann, S., BenTovim, L., Nietlispach, D., Broadhurst, R. W., Terada, T., Kelly, M., Oschkinat, H., Shibata, T., Yokoyama, S. & Laue, E. D. (1996). An approach to global fold determination using limited NMR data from larger proteins selectively protonated at specific residue types. *Journal of Biomolecular NMR* **8**, 360–368.

Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. & Alakhov, Y. B. (1988). A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**, 1162–1164.

Sprangers, R. & Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622.

Staunton, D., Owen, J. & Campbell, I. D. (2003). NMR and structural genomics. *Accounts of Chemical Research* **36**, 207–214.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. (2000). A novel NMR method for determining the interfaces of large protein-protein complexes. *Nature Structural Biology* **7**, 220–223.

Takeda, M., Chang, C. K., Ikeya, T., Güntert, P., Chang, Y. H., Hsu, Y. L., Huang, T. H. & Kainosho, M. (2008a). Solution structure of the C-terminal dimerization domain of SARS coronavirus nucleocapsid protein determined by the SAIL-NMR method. Implications for RNA packaging as revealed by nucleic acid interactions. *Journal of Molecular Biology* **380**, 608–622.

Takeda, M., Ikeya, T., Güntert, P. & Kainosho, M. (2007). Automated structure determination of proteins

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

with the SAIL-FLYA NMR method. *Nature Protocols* **2**, 2896–2902.

Takeda, M., Ono, A. M., Terauchi, T. & Kainosho, M. (2009). Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. *Journal of Biomolecular NMR* **44**, 261–265.

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. (2008b). Structure of the putative 32 kDa myrosinase-binding protein from *Arabidopsis* (At3g16450.1) determined by SAIL-NMR. *FEBS Journal* **275**, 5873–5884.

Tate, S., Ushioda, T., Utsunomiya-Tate, N., Shibuya, K., Ohyama, Y., Nakano, Y., Kaji, H., Inagaki, F., Samejima, T. & Kainosho, M. (1995). Solution structure of a human cystatin A variant, cystatin A^{2-98} M65L, by NMR spectroscopy. A possible role of the interactions between the N- and C-termini to maintain the inhibitory active form of cystatin A. *Biochemistry* **34**, 14637–14648.

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. *Organic Letters* **10**, 2785–2787.

Torchia, D. A., Sparks, S. W. & Bax, A. (1988). Delineation of α-helical domains in deuteriated Staphylococcal nuclease by 2D NOE NMR spectroscopy. *Journal of the American Chemical Society* **110**, 2320–2321.

Torizawa, T., Ono, A. M., Terauchi, T. & Kainosho, M. (2005). NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. *Journal of the American Chemical Society* **127**, 12620–12626.

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

Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F. & Dötsch, V. (2005). Efficient strategy for the rapid backbone assignment of membrane proteins. *Journal of the American Chemical Society* **127**, 13504–13505.

Tugarinov, V., Choy, W. Y., Orekhov, V. Y. & Kay, L. E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences USA* **102**, 622–627.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Wenger, R. K., Yao, H. Y. & Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research* **36**, D402–D408.

Viswanatha, V. & Hruby, V. J. (1980). Conversion of L-tyrosine to L-phenylalanine. Preparation of L-[3′,5′-$^{13}C_2$]phenylalanine. *Journal of Organic Chemistry* **45**, 2010–2012.

Vitali, F., Henning, A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M. & Allain, F. H. T. (2006). Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO Journal* **25**, 150–162.

Vuister, G. W., Kim, S. J., Wu, C. & Bax, A. (1994). 2D and 3D NMR study of phenylalanine residues in proteins by reverse isotopic labeling. *Journal of the American Chemical Society* **116**, 9206–9210.

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. *Biophysics of Structure and Mechanism* **2**, 139–158.

Wang, H., Janowick, D. A., Schkeryantz, J. M., Liu, X. H. & Fesik, S. W. (1999). A method for assigning phenylalanines in proteins. *Journal of the American Chemical Society* **121**, 1611–1612.

Williams, N. K., Liepinsh, E., Watt, S. J., Prosselkov, P., Matthews, J. M., Attard, P., Beck, J. L., Dixon, N. E. & Otting, G. (2005). Stabilization of native protein fold by intein-mediated covalent cyclization. *Journal of Molecular Biology* **346**, 1095–1108.

Williams, N. K., Prosselkov, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D. R., Curmi, P. M. G., Otting, G. & Dixon, N. E. (2002). In vivo protein cyclization promoted by a circularly permuted Synechocystis sp PCC6803 DnaB mini-intein. *Journal of Biological Chemistry* **277**, 7790–7798.

Williams, R. M. (1989). Synthesis of Optically Active α-Amino Acids. Chapter 2: Homologation of the β-carbon, in *Organic Chemistry Series*, Ed. Baldwin, J. E. Pergamon Press: Oxford **7**, 134–166.

Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *Journal of Magnetic Resonance Series B* **101**, 201–205.

Wu, P. S. C., Ozawa, K., Lim, S. P., Vasudevan, S. G., Dixon, N. E. & Otting, G. (2007). Cell-free transcription/translation from PCR-amplified DNA for high-throughput NMR studies. *Angewandte Chemie-International Edition* **46**, 3356–3358.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. New York: Wiley.

Wüthrich, K. (1998). The second decade – into the third millennium. *Nature Structural Biology* **5**, 492–495.

Xu, R., Ayers, B., Cowburn, D. & Muir, T. W. (1999). Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proceedings of the National Academy of Sciences USA* **96**, 388–393.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press

YAGI, H., TSUJIMOTO, T., YAMAZAKI, T., YOSHIDA, M. & AKUTSU, H. (2004). Conformational change of H^+-ATPase β monomer revealed on segmental isotope labeling NMR spectroscopy. *Journal of the American Chemical Society* **126**, 16632–16638.

YAMAZAKI, T., FORMAN-KAY, J. D. & KAY, L. E. (1993). Two-dimensional NMR experiments for correlating $^{13}C\beta$ and $^{1}H\delta/\varepsilon$ chemical shifts of aromatic residues in ^{13}C-labeled proteins via scalar couplings. *Journal of the American Chemical Society* **115**, 11054–11055.

YAMAZAKI, T., LEE, W., ARROWSMITH, C. H., MUHANDIRAM, D. R. & KAY, L. E. (1994). A suite of triple-resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high-sensitivity. *Journal of the American Chemical Society* **116**, 11655–11666.

YAMAZAKI, T., OTOMO, T., ODA, N., KYOGOKU, Y., UEGAKI, K., ITO, N., ISHINO, Y. & NAKAMURA, H. (1998). Segmental isotope labeling for protein NMR using peptide splicing. *Journal of the American Chemical Society* **120**, 5591–5592.

YAMAZAKI, T., TOCHIO, H., FURUI, J., AIMOTO, S. & KYOGOKU, Y. (1997). Assignment of backbone resonances for larger proteins using the ^{13}C-^{1}H coherence of a $^{1}H_{\alpha}$-, ^{2}H, ^{13}C, and ^{15}N-labeled sample. *Journal of the American Chemical Society* **119**, 872–880.

YEE, A., CHANG, X. Q., PINEDA-LUCENA, A., WU, B., SEMESI, A., LE, B., RAMELOT, T., LEE, G. M., BHATTACHARYYA, S., GUTIERREZ, P., DENISOV, A., LEE, C. H., CORT, J. R., KOZLOV, G., LIAO, J., FINAK, G., CHEN, L., WISHART, D., LEE, W., MCINTOSH, L. P., GEHRING, K., KENNEDY, M. A., EDWARDS, A. M. & ARROWSMITH, C. H. (2002). An NMR approach to structural proteomics. *Proceedings of the National Academy of Sciences USA* **99**, 1825–1830.

YEE, A., PARDEE, K., CHRISTENDAT, D., SAVCHENKO, A., EDWARDS, A. M. & ARROWSMITH, C. H. (2003). Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. *Accounts of Chemical Research* **36**, 183–189.

YOKOYAMA, S., HIROTA, H., KIGAWA, T., YABUKI, T., SHIROUZU, M., TERADA, T., ITO, Y., MATSUO, Y., KURODA, Y., NISHIMURA, Y., KYOGOKU, Y., MIKI, K., MASUI, R. & KURAMITSU, S. (2000). Structural genomics projects in Japan. *Nature Structural Biology* **7**, 943–945.

YOSHIDA, H., FURUYA, N., LIN, Y. J., GÜNTERT, P., KOMANO, T. & KAINOSHO, M. (2008). Structural basis of the role of the NikA ribbon-helix-helix domain in initiating bacterial conjugation. *Journal of Molecular Biology* **384**, 690–701.

ZANGGER, K., OBERER, M., KELLER, W. & STERK, H. (2003). X-filtering for a range of coupling constants: application to the detection of intermolecular NOEs. *Journal of Magnetic Resonance* **160**, 97–106.

ZHAO, W. T., ZHANG, Y. H., CUI, C. X., LI, Q. Q. & WANG, J. J. (2008). An efficient on-column expressed protein ligation strategy: application to segmental triple labeling of human apolipoprotein E3. *Protein Science* **17**, 736–747.

ZUBAY, G. (1973). *In vitro* synthesis of protein in microbial systems. *Annual Review of Genetics* **7**, 267–287.

ZÜGER, S. & IWAI, H. (2005). Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nature Biotechnology* **23**, 736–740.

ZWAHLEN, C., LEGAULT, P., VINCENT, S. J. F., GREENBLATT, J., KONRAT, R. & KAY, L. E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/*boxB* RNA complex. *Journal of the American Chemical Society* **119**, 6711–6721.

https://doi.org/10.1017/S0033583510000016 Published online by Cambridge University Press