Chapter 1
Isotope Labeling Methods for Large Systems

Patrik Lundström, Alexandra Ahlner, and Annica Theresia Blissing

Abstract A major drawback of nuclear magnetic resonance (NMR) spectroscopy compared to other methods is that the technique has been limited to relatively small molecules. However, in the last two decades the size limit has been pushed upwards considerably and it is now possible to use NMR spectroscopy for structure calculations of proteins of molecular weights approaching 100 kDa and to probe dynamics for supramolecular complexes of molecular weights in excess of 500 kDa. Instrumental for this progress has been development in instrumentation and pulse sequence design but also improved isotopic labeling schemes that lead to increased sensitivity as well as improved spectral resolution and simplification. These are described and discussed in this chapter, focusing on labeling schemes for amide proton and methyl proton detected experiments. We also discuss labeling methods for other potentially useful positions in proteins.

1.1 Introduction

In the first few years following the first successful nuclear magnetic resonance (NMR) spectroscopy experiments [1, 2], the technique was primarily the physicist's tool. After all, it was a nuclear phenomenon that apparently was of little relevance for chemistry. This dramatically changed when chemical shifts of spectral lines due to chemical environment were demonstrated [3, 4]. Now, molecules could be identified based on their NMR spectra and this is the major reason why NMR has been the most important spectroscopic method ever since. Around the same time, two almost as important nuclear magnetic phenomena were discovered. One was the scalar coupling between nuclear spins mediated by electrons in the covalent bonds separating the nuclei. In contrast to the direct dipolar coupling, this coupling leads to splitting of spectral lines, and equivalently provides a way to transfer magnetization, in isotropic samples. This is important as it provides a way of correlating two nuclei. The other important discovery was the Overhauser and nuclear Overhauser effect (NOE) that correctly predicted that nuclear polarization depends on the spin state of nearby (in space) unpaired electrons or nuclei [5–7] and enables measurements of internuclear distances. The chemical shift, the scalar coupling and the NOE in principle provide the tools needed for structure calculations of complex molecules such as proteins. The scalar coupling is used to establish correlations between nuclei and thus to assign the

P. Lundström (✉) • A. Ahlner • A.T. Blissing
Division of Molecular Biotechnology, Department of Physics, Chemistry and Biology, Linköping University, SE-58183 Linköping, Sweden
e-mail: patlu@ifm.liu.se

H.S. Atreya (ed.), *Isotope labeling in Biomolecular NMR*, Advances in Experimental Medicine and Biology 992,
DOI 10.1007/978-94-007-4954-2_1,

resonances and the NOE is used to measure their separation. Using protocols such as distance geometry and simulated annealing, the three-dimensional structure can then be calculated.

Because of signal overlap, one dimensional NMR experiments are not feasible for structure calculations or other high resolution applications even for small proteins. Luckily, methods for recording two dimensional homonuclear experiments to establish correlations between nuclei were developed in the late 1970s and in 1982 Kurt Wüthrich and coworkers had completed the resonance assignments of the 6.5 kDa protein basic pancreatic trypsin inhibitor using ^{1}H homonuclear experiments [8]. In 1985 the same group were able to calculate the solution structure of the protein bull seminal protease inhibitor (6 kDa) primarily from distance restraints derived from NOESY experiments [9] and in the years that followed, the solution structures of several other proteins of similar size were calculated. However, the process of obtaining resonance assignments was labor intensive and the signal overlap was too severe for the method to be practical for proteins larger than approximately 10 kDa. A partial remedy that did not require isotopic labeling was to generalize the experiments to three dimensions and thus to correlate three different protons using two different mixing sequences, such as one NOE and one Hartmann-Hahn transfer period [10]. While these experiments were useful to increase spectral resolution and to establish many correlations between nuclei in a single experiment they suffered from the shortcoming of small scalar couplings between protons separated by three covalent bonds. Many potential experiments would thus need prohibitively long transfer times with concomitant reduction in sensitivity. In contrast, many heteronuclear scalar couplings are significantly stronger and it was recognized that if a heteronucleus such as ^{15}N or ^{13}C is used in the third dimension, superior sensitivity could be achieved [11–13]. Today, the most important of these heteronuclear experiments are NOESY-HSQC and TOCSY-HSQC and by using these types of experiments proteins up to 20 kDa can often be assigned and their structures calculated. Using this approach, the assignment process is performed by first identifying spin systems using the TOCSY-HSQC experiments and connecting them sequentially by the aid of the NOESY-HSQC experiments.

When triple-resonance experiments were developed [14], a new, more effective way of obtaining resonance assignments was possible by correlating the amide proton and nitrogen of one residue with one or two carbon nuclei of the same residue and of the preceding residue. Side-chain assignments could then be completed with ease using TOCSY experiments and chemical shift information of one or more carbon nuclei obtained by the triple-resonance experiments. The structures themselves were still mainly calculated from distance restraints obtained from NOESY-HSQC experiments. With these methods and with the addition of residual dipolar coupling (RDC) restraints [15], structure calculations by NMR could be completed in less time and with higher precision than before. A necessary price that had to be paid for using these conceptually simpler and less cumbersome experiments was that the proteins must be simultaneously labeled with ^{13}C and ^{15}N.

Another major breakthrough came with the advent of transverse relaxation optimized spectroscopy (TROSY) pulse sequences [16]. For these to work well, most non-labile protons must be replaced by deuterons. Using these methods it is possible to increase the size limit significantly and in favorable cases it is possible to perform backbone resonance assignments and structure calculations of proteins approaching 100 kDa [17]. A significant drawback of these methods is that most protons that are used as distance restraints, including in the protein core, have been removed and a NOE driven structure calculation then has to rely only on distances between amide protons leading to very few restraints for each residue. A way of improving the situation is to add side-chain protons at strategic places in an otherwise deuterated background [18]. For large proteins this would be at side-chain methyl groups, because of their favorable relaxation properties and because of their numerous contacts. They can be assigned using experiments that correlate the methyl groups with the protein backbone [19] and add crucial information about the protein core.

The progress in NMR spectroscopy applied to large systems has been extremely rapid in recent years and the applications have been impressive. One reason for this is the development of high-field instruments and cryogenically cooled probes with superior sensitivity compared to standard instruments

two decades ago. Another important aspect is the development of new pulse sequences that allow magnetization transfer in a spin state selective way so that relaxation losses are minimized. However, neither sensitive instruments nor clever pulse sequences suffice for recording spectra of high sensitivity for large systems. An equally important requirement is labeling schemes designed to enhance resolution and to reduce relaxation rates and spectral crowding. These labeling schemes are the focus of this chapter. The definition I will use for a large system is that the usual combination of pulse sequences and isotopic labeling schemes, i.e. non-TROSY pulse sequences and fully protonated uniformly ^{15}N and/or ^{13}C labeled samples, will fail. It is not possible to provide a certain number for what this means in terms of molecular weight but a biomolecule or a complex of biomolecules larger than 30 kDa fulfills this criterion for most applications.

1.2 Spin Relaxation and TROSY

There are two main challenges with NMR spectroscopy applied to large systems. One is signal overlap due to spectral crowding. While this is a concern, many applications have focused on oligomeric complexes so that although the complex tumbles as a large unit, the number of signals is manageable. Also for a monomeric protein as large as the 82 kDa *E. coli* malate synthase G, Tugarinov et al. recorded a beautiful ^{15}N-^{1}H correlation spectrum that, while crowded, had most peaks resolved [17]. Furthermore, if necessary it is possible to selectively label a subset of the amino acid residues as will be discussed below. The more serious problem is line-broadening and concomitant reduced sensitivity for large systems due to rapid transverse relaxation. This, in practice, sets the limit for how large systems that can be studied. A useful strategy for the study of large systems is thus to try to increase the tumbling rate by increasing the temperature. For proteins this can of course not be done indefinitely since stability is decreased at high temperature. Proteins from thermophilic organisms are useful in this regard. However, it is usually not possible to improve spectral quality sufficiently for high molecular weight systems by only increasing the temperature. The method that has opened the door to NMR studies of larger proteins is line narrowing by TROSY [20].

To see how line narrowing is achieved in TROSY type experiments, we consider a spin-pair *IS* that is scalar coupled with coupling constant J_{IS} and evaluate the time evolution of the two components of *I* transverse magnetization. Without loss of generality we will disregard from chemical shift evolution. If the two components are relaxed by the dipole-dipole interaction with spin *S* and by *I* chemical shift anisotropy they evolve as [21]

$$\frac{d}{dt}\begin{pmatrix} I^+S^\alpha(t) \\ I^+S^\beta(t) \end{pmatrix} = \begin{pmatrix} i\pi J_{IS} + \bar{R}_2 + \eta_{xy} & \left(R_{2I}^{DD} - R_{2IS}^{DD}\right)/2 \\ \left(R_{2I}^{DD} - R_{2IS}^{DD}\right)/2 & -i\pi J_{IS} + \bar{R}_2 - \eta_{xy} \end{pmatrix} \begin{pmatrix} I^+S^\alpha(t) \\ I^+S^\beta(t) \end{pmatrix} \tag{1.1}$$

where

$$\bar{R}_2 = \left(R_{2I}^{DD} + R_{2IS}^{DD}\right)/2 + R_{2I}^{CSA} \tag{1.2}$$

$$R_{2I}^{DD} = \frac{d^2}{8}\left[4J(0) + 3J\left(\omega_I\right) + 6J(\omega_S) + 6J(\omega_I + \omega_S) + J(\omega_I - \omega_S)\right] \tag{1.3}$$

$$R_{2IS}^{DD} = \frac{d^2}{8}\left[4J(0) + 3J\left(\omega_I\right) + 6J(\omega_I + \omega_S) + J(\omega_I - \omega_S)\right] \tag{1.4}$$

$$R_{2I}^{CSA} = \frac{c^2}{6}\left[4J(0)+3J\left(\omega_I\right)\right] \tag{1.5}$$

$$\eta_{xy} = \frac{\sqrt{3}}{6}cdP_2\left(\cos\theta\right)\left[4J(0)+3J\left(\omega_I\right)\right] \tag{1.6}$$

where $d = \mu_0\hbar\gamma_I\gamma_S\left\langle r_{IS}\right\rangle^{-3}/4\pi; c = \gamma_I B_0\Delta\sigma/\sqrt{3}; \mu_0$ is the permeability of vacuum; $\hbar$ is the reduced Planck constant; γ_I and γ_S are the magnetogyric ratios; r_{IS} is the internuclear distance; B_0 is the static magnetic field strength, $\Delta\sigma$ is the anisotropy of the (axially symmetric) chemical shift tensor and $P_2\left(\cos\theta\right)$ is the second order Legendre polynomial of the cosine of the angle between the principal frames of the dipolar and chemical shift anisotropy interactions. $J\left(\omega\right)$ is the spectral density that is usually modeled by the model-free formalism [22–24]

$$J\left(\omega\right) = \frac{2}{5}\left[\frac{S^2\tau_c}{1+\left(\omega\tau_c\right)^2} + \frac{\left(1-S_f^2\right)\tau_f'}{1+\left(\omega\tau_f'\right)^2} + \frac{\left(S_f^2-S^2\right)\tau_s'}{1+\left(\omega\tau_s'\right)^2}\right] \tag{1.7}$$

where $S^2 = S_f^2S_s^2; S_f^2$ and S_s^2 are the generalized order parameters for fast and slow internal motions, respectively; $\tau_f' = \tau_f\tau_c/(\tau_f+\tau_c), \tau_s' = \tau_s\tau_c/(\tau_s+\tau_c); \tau_f$ and τ_s are the correlation times for the fast and slow internal motions respectively and τ_c is the correlation time for molecular tumbling.

If $2\pi J_{IS}{}^2 \gg \left(R_{2I}^{DD} - R_{2IS}^{DD}\right)^2$ the off-diagonal elements are unimportant and cross-relaxation between the two components can be neglected. Equation 1.1 shows that the relaxation rate of one component is reduced while the relaxation rate of the other is enhanced by the cross-correlation relaxation rate between the dipole-dipole and chemical shift anisotropy tensors (η_{xy}) and consequently one component will be broad while the other will be narrow. TROSY experiments select the narrow component in NMR spectra [20] and are designed not to mix slowly and rapidly relaxing components. The auto-correlated and cross-correlated contributions to relaxation have different field dependence which means that complete cancellation occurs at certain field strength. This argument is however based on the assumption that the spin-pair is isolated from dipole-dipole interactions with remote spins. In proteins this is not the case and especially for protonated samples there is additional broadening. The line-narrowing effect by the TROSY principle for an isolated ^{15}N-^{1}H spin-pair is shown in Fig. 1.1.

For IS_2 and IS_3 spin systems there are additional cross-correlations between different dipolar interactions. The TROSY principle can be applied in these cases as well and line narrowing can be achieved by selecting the appropriate lines in such multiplets. TROSY experiments have been described for methyl [25] and also methylene [26] groups. The theoretical description in these cases is rather involved and the interested reader is referred to the original works.

The original TROSY experiment for recording ^{15}N-^{1}HN correlation maps is shown in Fig. 1.2a [20]. In this experiment the slowly relaxing component is selected in both dimensions and it is noteworthy that polarization originating from ^{15}N contributes to the signal. Gradient-selected sensitivity enhanced versions [28, 29] as well as versions that actively suppress the rapidly relaxing components without increasing the phase cycle have subsequently been described [30]. TROSY detection for ^{15}N and ^{1}HN is also readily incorporated into triple-resonance pulse sequences [31, 32]. This has been instrumental for resonance assignments of large systems.

In the case of applications to methyl groups the heteronuclear multiple-quantum coherence (HMQC) experiment is used to record TROSY correlation maps. Slowly relaxing single quantum ^{1}H coherence, produced by the only 90° ^{1}H pulse, is converted to slowly relaxing heteronuclear double and zero quantum coherence by a 90° ^{13}C pulse. It is then converted back to slowly relaxing single quantum ^{1}H coherence by a second 90° ^{13}C pulse. Because ^{13}C pulses are used exclusively to change the

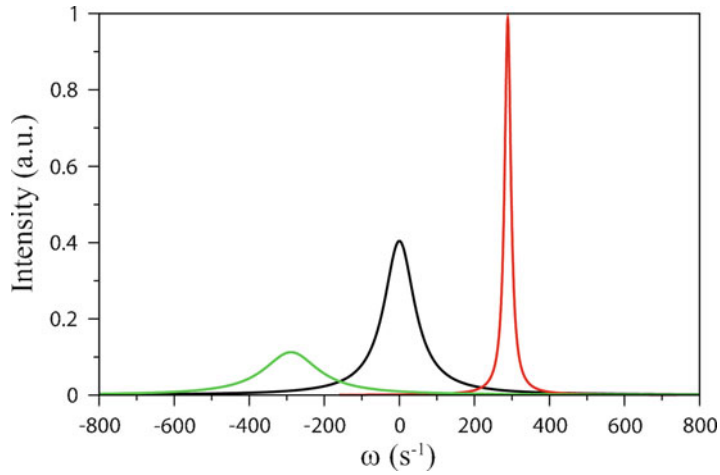

Fig. 1.1 Illustration of the TROSY effect on the line shape. The figure was prepared by solving Eq. 1.1 for an isolated ^{15}N-^{1}H spin-pair, tumbling with a correlation time of 30 ns at a static magnetic field of 18.8 T. The scalar coupling constant is 92 Hz. The *red* and *green lines* correspond to the narrow and broad components of the doublet, respectively, whereas the *black line* represents the line shape in a decoupled spectrum

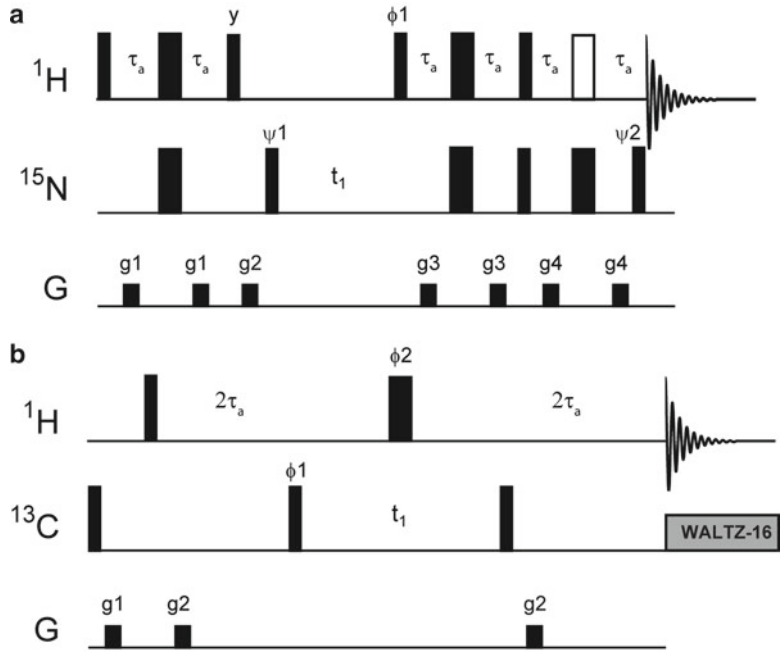

Fig. 1.2 (**a**) The original ^{15}N-^{1}HN TROSY pulse sequence [20]. *Narrow* and *wide rectangles* depict 90° and 180° pulses, respectively. The *open rectangle* represents a WATERGATE element. The delay τ_a is equal to $1/(4J_{NH})$. The phase cycling is ψ1 = y,–y,–x,x; ψ2 = 4(x),4(–x); ϕ1 = 4(y),4(–y) and receiver = y,–y,–x,x. (**b**) The basic HMQC pulse sequence [27]. In this case the delay τ_a is equal to $1/(4J_{CH})$. The phase cycling is ϕ1 = x,–x; ϕ2 = x,x,y,y,–x,–x,–y,–y; receiver = x,–x,–x,x. Quadrature detection is achieved by States-TPPI phase incrementation in both cases

coherence level after the initial ^{1}H pulse, rapidly and slowly relaxing coherences are never mixed [25]. The HMQC pulse sequence is shown in Fig. 1.2b. A way of achieving enhanced resolution with only modest (10%) loss in sensitivity is to use heteronuclear zero-quantum (HZQC) instead of HMQC experiments [33].

1.3 Labeling Protocol for Large Proteins

Isotopically enriched proteins are usually expressed in M9 minimal medium [34], referred to as M9 medium in the following. M9 medium is primarily composed of 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 0.5 g/L NaCl. These salts are dissolved in D_2O and the medium is supplemented with 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 mg/L biotin, 10 mg/L thiamine and antibiotics. Stock solutions for these components must be dissolved in D_2O. 0.5–1 g/L $^{15}NH_4Cl$ is also added to the medium. Depending on whether simultaneous labeling with ^{13}C is required or not the carbon source is either 2–3 g/L [$^{13}C_6$, $^{2}H_7$]-glucose or 2–3 g/L [$^{12}C_6$, $^{2}H_7$]-glucose. The use of protonated glucose in combination with 100% D_2O as solvent is strongly discouraged since many hydrogen positions are derived from both these sources. In addition to lowering the sensitivity because of introduction of a high magnetogyric ratio nucleus it leads to a mixture of different isotope shifts and further broadened lines. It is imperative to not autoclave deuterated growth media since this leads to partial protonation due to $H_2O \leftrightarrow D_2O$ exchange. Instead, sterile filtration should be performed. Below is an expression protocol that has frequently been used to produce highly deuterated samples.

1. Transfer one or more freshly transformed *E. coli* colonies of BL21(DE3) strain to 30 mL LB (in H_2O) supplemented with the appropriate antibiotic(s) and grow cells at 37 °C in a shaking incubator until $OD_{600} = 1.0$ is reached.
2. Spin down the cells at 1,200 g, 15 min. at room temperature (25 °C).
3. Resuspend a fraction of the cells in 10% of the isotopically labeled M9 medium to achieve OD_{600} of 0.1–0.2. Grow the cells at 37 °C until $OD_{600} = 1.0$. Pour the starter culture directly into the remaining 90% of the isotopically enriched M9 medium.
4. Grow the cells at 37 °C until $OD_{600} = 0.6–1.0$.
5. Induce over expression with 0.5–1 mM IPTG. Perform over expression either at 37 °C for 2–5 h or at room temperature or 16 °C overnight. The final OD_{600} will depend on the growth medium and on the duration of over expression. This step should be modified if selectively labeled precursors or amino acids are used. These compounds are added to the growth medium 1 h before over expression is induced.

A purification protocol has to be developed for each protein. Denaturation and refolding is usually necessary if complete exchange of deuterons to protons, or vice versa, at amide positions is required. Typically, the protein is unfolded in 6 M GdnCl or 8 M urea after cell lysis. Refolding is done by exchange to a buffer that favors the folded state and can be performed by several different methods including dialysis, on column or by rapid dilution [35].

1.4 Labeling Methods for Amide Proton Detected Experiments

Generally, structural investigations of large proteins require perdeuterated samples although methods for dealing with quite large systems using proteins only $^{13}C/^{15}N$ labeled have been described. For instance, Xu et al. have developed a strategy for assigning spectra of large uniformly ^{13}C, ^{15}N labeled proteins [36]. By using TROSY-HNCA, ^{13}C,^{15}N-NOESY, ^{13}C,^{13}C-NOESY and MQ-HCCH-TOCSY

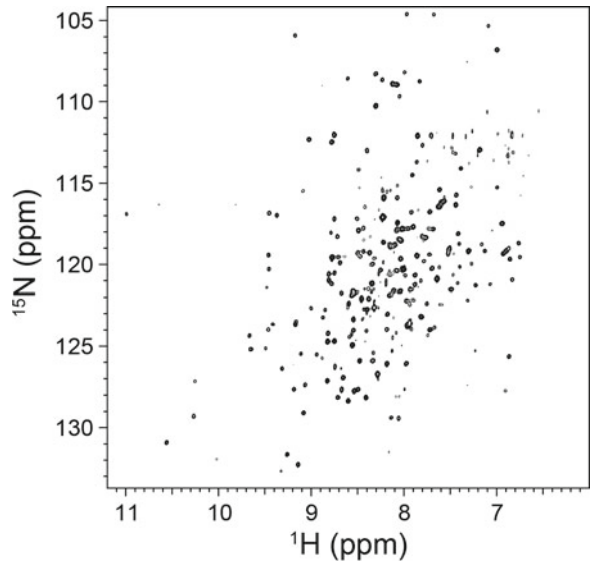

Fig. 1.3 ^{15}N-^{1}HN correlation map of a perdeuterated homodimer of the kinase domain from EphB2, 2×289 residues, 65 kDa, recorded with a TROSY pulse sequence at 25 °C and a static magnetic field of 18.8 T

spectra, clusters with the same HN shifts were identified, and individual spins systems were determined, creating dipeptide fragments that were pieced together and assigned to the sequence. They were able to assign the 42 kDa maltose binding protein and the 65 kDa protein hemoglobin [36].

The reason why perdeuteration is beneficial in TROSY experiments is because the amide proton is not only relaxed by the dipolar interaction with its attached ^{15}N spin and its chemical shift anisotropy but also by dipolar interactions with (especially) other protons that are close in space. Since this means that the total dipolar interaction increases while the chemical shift anisotropy interaction stays the same, the TROSY effect is diminished. It also means that $2\pi J_{IS}{}^{2} \gg \left(R_{2I}^{DD} - R_{2IS}^{DD}\right)^{2}$ no longer holds which leads to cross-relaxation between the two components and a further reduction of the TROSY effect. For large proteins it is thus essential to reduce the pool of protons. This can be done by perdeuteration, replacing aliphatic and aromatic protons with deuterons. By reducing the number of possible interactions with other protons, the TROSY effect is retained.

As is evident from Fig. 1.3, which shows a TROSY ^{15}N-^{1}HN correlation map of a perdeuterated ^{15}N/^{13}C labeled sample of dimeric kinase domain of the protein EphB2 (2×289 residue, 65 kDa), spectra of excellent quality are possible even for quite large systems. The spectrum was recorded at 25 °C at a static magnetic field strength 18.8 T and the protein was purified from inclusion bodies by solubilization in 6 M GdnHCl and subsequently refolded by dialysis. TROSY pulse sequences in combination with perdeuteration are quite sensitive even for a monomeric protein as large as the *E. coli* malate synthase G that has a molecular weight of 82 kDa (723 residues). The protein was assigned to 95% at backbone positions and 97% at Cα, Cβ, CO positions [17]. In experiments involving transverse magnetization of a nucleus that is primarily relaxed by chemical shift anisotropy neither the TROSY effect nor perdeuteration is of much help. Furthermore, since the chemical shift anisotropy mediated relaxation scales with the square of the external magnetic field going from 600 to 800 MHz in fact reduces sensitivity [17]. Another cause of sensitivity loss in many of these experiments is due to imperfections in the large number of pulses in these sequences. It is thus imperative to keep the number of applied pulses to a minimum and to use optimal pulses. In this regard recent developments in pulse sequences promise much for the future. For instance, a combination of TROSY and multiple-quantum evolution elements on average leads to gains in sensitivity of a factor 1.8 when compared

to the conventional TROSY-HN(CO)CA experiment for the membrane protein-detergent complex KcsA with a rotational correlation time of around 60 ns [37]. The reason for the gain in sensitivity is a combination of these relaxation-optimized elements and that ten less pulses on ^{13}C are required. Another complication for highly deuterated samples of high molecular weight proteins relates to slow proton longitudinal relaxation and consequently the need for long inter scan delays in NMR experiments. Acceleration of longitudinal proton relaxation between scans can be achieved by the band-selective excitation short-transient (BEST) technique where band-selective pulse centered on the amide region are substituted for hard pulses on proton [38]. This means that experiments can be repeated very rapidly, resulting in higher sensitivity per unit time.

A drawback with perdeuteration is that experiments that require interactions with and between aliphatic and aromatic protons are not possible. The most important example is the NOESY experiment that is used to obtain distance restraints for structure calculations. For very large systems, like malate synthase G described above, the most viable strategy is to introduce protons at strategic positions, where relaxation properties are favorable, such as methyl groups. Selective labeling schemes for methyl groups are described later in this text. Indeed, when the global fold of this protein was solved by NMR spectroscopy, the relatively few distance restraints obtained from ^{1}HN and $^{1}H^{methyl}$ had to be complemented by other measured parameters such as residual dipolar couplings and secondary chemical shifts [39].

Amide proton detected experiments for even larger systems, comprising homo-oligomeric supramolecular complexes, have also been described. These complexes benefit from having the spectral appearance of monomers and from the fact that the intensity of the NMR signal is multiplied with the number of monomers per oligomer. Wüthrich and coworkers have recorded ^{15}N-^{1}H correlation maps of the heptameric co-chaperonin GroES (72 kDa) bound to either SR1 (400 kDa) or GroEL (800 kDa). While GroES was perdeuterated it did not matter whether the other proteins were perdeuterated or not. Because of significant relaxation losses during transfer periods TROSY experiments perform significantly worse than ones incorporating cross-relaxation induced polarization transfer (CRIPT) for systems of this size. In CRIPT spectra of GroES bound to GroEL, the slowly relaxing component for most of the expected 94 resonances were observed [40].

To decrease spectral crowding for large proteins and for assignment purposes, selective ^{15}N labeling of only one or several residue types can be employed. For small to medium-sized proteins this can be achieved by growth and expression in rich medium supplemented with ^{15}N labeled amino acids for the selected residues and high levels of all other amino acids in unlabeled form [41]. For large proteins a similar strategy is growth in M9 medium in 100% D_2O supplemented with 1 g/L glucose, 1 g/L natural abundance NH_4Cl, 1–3 g/L deuterated algal lysate and 4 mg/L deuterated and ^{15}N labeled forms of each of the selected amino acids [42]. Since the algal lysate contains amino acids it is necessary to carefully optimize the amount of labeled amino acids to avoid cross-labeling due to transamination. Completely clean labeling of the desired residue types was not possible unless bacterial strains deficient in transaminases were used [42].

1.5 Labeling Methods for Aromatic Side-Chain Positions

TROSY line-narrowing can also be realized for other positions than backbone amides. Two examples of this are $^{1}H^{15}N\varepsilon 2$ moieties of the tryptophan side-chain and $^{1}H^{13}C$ groups of the side-chains of Phe and Tyr.

1.5.1 Labeling of Tryptophan Side-Chains

Tryptophan side-chains frequently provide important interactions in the early stages of protein folding [43] and participate in interactions with ligands or other proteins [44]. Also, tryptophan residues have

been shown to be sensitive to solvation and temperature [45]. These characteristics combined make this amino acid particularly useful as a biophysical probe for NMR spectroscopy.

Since tryptophan residues report on several biophysical events, there are many ways to study these occurrences. Löhr et al. have written a three-dimensional experiment to correlate tryptophan side-chain with $^{13}C\beta$ [46]. Monitoring perturbations of the chemical shifts upon interaction with various ligands provide information on the nature of the association, without the need for complete structure determination. Similarly, cross-saturation experiments can be employed to characterize large protein interaction complexes [47].

A simple and cost-effective way of selectively labeling tryptophan side-chains is to add perdeuterated [2,4-$^{13}C_2$]-indole, which is a precursor of tryptophan, to the culture medium prior to induction of protein expression. This method results in excellent incorporation with virtually no scrambling of label [45].

1.5.2 Labeling of Aromatic Side-Chain Positions

Because of the large chemical shift anisotropy for the carbon positions of aromatic side-chains, substantial line narrowing by the TROSY principle can also be achieved in this case. TROSY experiments applied to aromatic side-chains have been described [16]. In this case, maximum line-narrowing is achieved at static magnetic fields between 11.7 and 18.8 T. For uniformly ^{13}C labeled proteins these experiments require constant-time evolution in the indirect dimension and since it is not feasible to selectively protonate a protein at certain aromatic side-chain positions by using glucose or related molecules as carbon sources, these experiments are not widely used although one application is HCCH-TOCSY experiments with TROSY detection in both carbon dimensions [48]. It should be noted however that specifically labeled aromatic amino acids can be added to a deuterated background. The amino acids should optimally be $^{13}C^1H$ labeled at the positions that are detected and $^{12}C^2H$ elsewhere. The in vitro stereo-array isotope labeling (SAIL) protocol produces aromatic amino acid side-chains with alternating $^{13}C^1H$ and $^{12}C^2H$ isotopomers and could thus be useful in this regard [49].

1.6 Labeling Methods for Methyl Groups

Methyl groups provide the most sensitive probes in NMR experiments because of their rapid rotation around a three-fold axis. Additional line-narrowing required for applications to large systems is achieved by methyl TROSY experiments. For applications involving high molecular weight systems, non-methyl positions must be deuterated and scalar couplings to adjacent carbons must not be present to avoid using constant-time evolutions periods. The use of methyl groups as probes for supra-molecular structure and dynamics has been reviewed extensively by Ruschak and Kay [50]. In this section, we will sometimes use the notation H for ^1H and D for ^2H for increased clarity.

A clean way of selectively labeling certain methyl side-chains is to add the commercially available compounds α-ketobutyrate (precursor of Ile) and/or α-ketoisovalerate (precursor of Leu, Val) that are specifically labeled with ^{13}C at the methyl groups to the growth medium [18]. This is usually referred to as ILV labeling and is shown in Fig. 1.4. Although these compounds can potentially be degraded into precursors for other amino acids the authors noticed essentially no labeling at other positions. A very useful feature of this method is that labeling of the methyl groups can be customized. For instance, in addition to having all methyl groups $^{13}CH_3$ it is also possible to have them $^{13}CH_2D$ or $^{13}CHD_2$. One can also label different methyl groups differently. By using α-ketoisovalerate labeled with $^{13}CH_3$ at one methyl group and with $^{12}CD_3$ at the other only the *proR* or *proS* methyl group of Leu and Val is detected. For applications involving high molecular weight systems, the non-methyl position

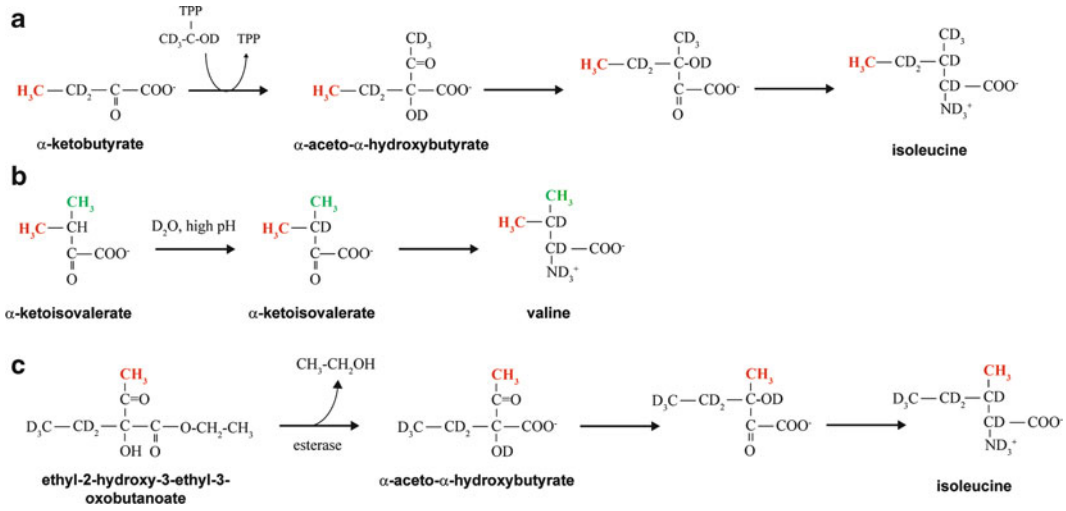

Fig. 1.4 Synthesis of the amino acids Ile and Val from the precursors α-ketobutyrate and α-ketoisovalerate, respectively for high molecular weight NMR applications. The methyl group originally at α-ketobutyrate is highlighted in *red* and the two methyl groups originally at α-ketoisovalerate are shown in *red* and *green*. The labeling pattern of the highlighted methyl groups can be customized independently. (**a**) Synthesis of Ile, selectively labeled at the δ1 position. (**b**) Synthesis of Val from the precursor α-ketoisovalerate. The initial step is deuteration of position 3 that can be performed in-house to cut costs. The methyl groups of Leu will be labeled in the same manner if α-ketoisovalerate is added to the growth medium. (**c**) Selective labeling of Ileγ2, achieved by adding α-aceto-α-hydroxybutyrate to the growth medium. This precursor is purchased in its ethyl ester form and deesterification is performed by treatment with esterase as illustrated by the first step

of α-ketoisovalerate can be deuterated in-house by incubation of 25 mM α-ketoisovalerate in D_2O at pH 12.5 for 2–3 h at 45 °C [18]. The completeness of the reaction can be followed by NMR.

To produce these samples, a similar protocol for bacterial growth and over expression as given above is used. The methyl groups of Ileδ1, Leu and Val are labeled to 90% if the precursors are supplied in concentrations of 50 mg/L for α-ketobutyrate and 100 mg/L for α-ketoisovalerate 1 h prior to induction of protein expression [18]. Using this approach it has for instance been possible to measure dynamics for systems as large as the proteasome 20S core particle of 670 kDa [51] and to reveal the structural basis for signal-sequence recognition by the translocase motor SecA [52].

To compare $^{13}CHD_2$ and $^{13}CH_3$ labeled methyl groups in NMR applications involving supra-molecules, Religa and Kay compared the relative sensitivity of experiments involving these isotopomers. The sensitivity of Ile, Leu, Val $^{13}CH_3$-labeled samples was found to be between 1.5 and 2 fold higher than the corresponding data sets obtained with $^{13}CHD_2$ probes. For supra-molecules, labeling with $^{13}CH_3$ isotopomers is thus the method of choice if maximum sensitivity is desired. However, applications that require $^{13}CHD_2$ moieties can still be performed with high sensitivity [53].

Recently, Ruschak et al. suggested a method to instead label Ileγ2 to enable measurements at this position in large proteins [54]. The protocol is based on adding the precursor α-aceto-α-hydroxybutyrate that is $^{13}C^1H_3$ labeled at the relevant methyl group and $^{12}C/^2H$ labeled elsewhere. For reasons of stability the compound is purchased in its ethyl ester form and de-esterified by incubation with esterase. An amount corresponding to 100 mg/L of the acid form is added to the growth medium. Contrary to what is observed for the ILV labeling scheme, scrambling leads to the presence of weak correlations of *proR* Valγ and Leuδ [54]. However, these do not complicate the interpretation of the spectra considerably.

Methyl groups in Met residues are also most useful probes. For applications involving large proteins it is, of course, essential that all positions except the Met methyl groups are deuterated. This is conveniently

achieved by supplying this residue, selectively ^{13}C, ^{1}H labeled at the methyl group, to a deuterated growth medium at a concentration of 100 mg/L 1 h prior to induction [55].

1.7 Cell-Free Protein Synthesis

Some proteins cannot be expressed successfully in *E coli*. Proteins that are involved in apoptosis, are rapidly degraded, have low solubility or are toxic can instead be expressed in a cell-free system. Cell-free expression is an *in vitro* method to produce proteins by isolating the protein translational machinery of eukaryotic or prokaryotic cells, allowing protein synthesis to take place *in vitro* and has since long been used for NMR applications [56]. Since the protein production takes place in an artificial environment, specific labels can be incorporated by supplying amino acids with the desired labeling pattern suitable for NMR studies. It is also possible to produce perdeuterated protein samples [57] and introduce non-natural amino acids [58].

Cell-free protein synthesis can be achieved either by batch-mode or continuous-exchange. The batch-mode protocol includes mixing the reactants (cell extract, template DNA, RNA polymerase, NTPs) in a buffered solution and incubating. However, build-up of by-products and depletion of NTPs are limiting factors. The continuous-exchange protocol involves incubating the reaction in a dialysis membrane, allowing exchange of solutes and thereby removing by-products and supplying NTPs to the synthesis reaction across the membrane. The continuous-flow cell-free system is a variation of this concept. Here, NTPs are continuously added to the reaction mixture while by-products are simultaneously removed through a membrane that retains the translational components.

1.8 Concluding Remarks

We have herein described isotopic labeling schemes and experimental methods for characterization of high molecular weight systems by NMR spectroscopy. Current methodology allows resonance assignments and structural characterization of single chain proteins approaching 100 kDa [17, 39] as well as relating structure and dynamics to function for supra-molecular systems as large as the 20S core particle of the ribosome [51, 55, 59]. Key to most labeling techniques is substitution of deuterons for protons at most positions. Protons are only retained at strategic positions, such as at backbone amide positions or at certain methyl groups, to gain as much sensitivity and resolution enhancement as possible in TROSY type experiments. The continuous development of new labeling schemes in combination with improved pulse sequences, more powerful magnets and more sensitive probes ensures that even larger and more complex systems can be studied with NMR spectroscopy in the future.

References

1. Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption of nuclear magnetic moments in a solid. Phys Rev 69:37–38
2. Bloch F, Hansen WW, Packard ME (1946) Nuclear induction. Phys Rev 69:680
3. Proctor WG, Yu FC (1950) The dependence of a nuclear magnetic resonance frequency upon chemical compounds. Phys Rev 77:717
4. Dickinson WC (1950) Dependence of the ^{19}F nuclear resonance position on chemical compound. Phys Rev 77: 736–737
5. Overhauser AW (1953) Polarization of nuclei in materials. Phys Rev 92:411–415
6. Carver TR, Slichter CP (1953) Polarization of nuclear spins in metals. Phys Rev 92:212–213

7. Solomon I (1955) Relaxation processes in a system of two spins. Phys Rev 99:559–566
8. Wagner G, Wüthrich K (1982) Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra: basic pancreatic trypsin inhibitor. J Mol Biol 155:347–366
9. Williamson MP, Havel TF, Wüthrich K (1985) Solution conformation of proteinase inhibitor IIa from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. J Mol Biol 182:295–315
10. Oschkinat H, Griesinger C, Kraulis PJ, Sorensen OW, Ernst RR, Gronenborn AM, Clore GM (1988) 3-dimensional NMR spectroscopy of a protein in solution. Nature 332:374–376
11. Fesik SW, Zuiderweg ERP (1988) Heteronuclear 3-dimensional NMR spectroscopy – a strategy for the simplification of homonuclear two-dimensional NMR spectra. J Magn Reson 78:588–593
12. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of 3-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn multiple quantum coherence and nuclear Overhauser multiple quantum coherence spectroscopy – application to interleukin-1 beta. Biochemistry 28:6150–6156
13. Ikura M, Kay LE, Tschudin R, Bax A (1990) 3-dimensional NOESY-HMQC spectroscopy of a ^{13}C labeled protein. J Magn Reson 86:204–209
14. Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514
15. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114
16. Pervushin K, Riek R, Wider G, Wüthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labeled proteins. J Am Chem Soc 120:6394–6400
17. Tugarinov V, Muhandiram R, Ayed A, Kay LE (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. J Am Chem Soc 124:10025–10035
18. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (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. J Biomol NMR 13:369–374
19. Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc 125:13868–13878
20. 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 biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371
21. Cavanagh J, Fairbrother WJ, Palmer AG 3rd, Rance M, Skelton NJ (2007) Protein NMR spectroscopy: principles and practice. Elsevier Academic Press, Burlington
22. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules.1. Theory and range of validity. J Am Chem Soc 104:4546–4559
23. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 104:4559–4570
24. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM (1990) Deviations from the simple two-parameter model-free approach to the interpretation of ^{15}N nuclear magnetic relaxation of proteins. J Am Chem Soc 112:4989–4991
25. Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. Magn Reson Chem 41:843–852
26. Miclet E, Williams DC Jr, Clore GM, Bryce DL, Boisbouvier J, Bax A (2004) Relaxation-optimized NMR spectroscopy of methylene groups in proteins and nucleic acids. J Am Chem Soc 126:10560–10570
27. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced H-1-C-13 NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125:10420–10428
28. Czisch M, Boelens R (1998) Sensitivity enhancement in the TROSY experiment. J Magn Reson 134:158–160
29. Weigelt J (1998) Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. J Am Chem Soc 120:10778–10779
30. Nietlispach D (2005) Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. J Biomol NMR 31:161–166
31. 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. Proc Natl Acad Sci USA 95:13585–13590
32. Yang DW, Kay LE (1999) Improved ^{1}HN-detected triple resonance TROSY-based experiments. J Biomol NMR 13:3–10
33. Tugarinov V, Sprangers R, Kay LE (2004) Line narrowing in methyl-TROSY using zero-quantum ^{1}H-^{13}C NMR spectroscopy. J Am Chem Soc 126:4921–4925
34. Maniatis T, Sambrook J, Fritsch EF (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 68–69

35. Middelberg APJ (2002) Preparative protein refolding. Trends Biotechnol 20:437–443
36. Xu YQ, Zheng Y, Fan JS, Yang DW (2006) A new strategy for structure determination of large proteins in solution without deuteration. Nat Methods 3:931–937
37. Bayrhuber M, Riek R (2011) Very simple combination of TROSY, CRINEPT and multiple quantum coherence for signal enhancement in an HN(CO)CA experiment for large proteins. J Magn Reson 209:310–314
38. Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043
39. Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci USA 102:622–627
40. Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900 K GroEL GroES complex. Nature 418:207–211
41. McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q Rev Biophys 23:1–38
42. Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2004) Uniform and residue-specific 15N-labeling of proteins on a highly deuterated background. J Biomol NMR 29:289–297
43. Baldwin RL (2002) Making a network of hydrophobic clusters. Science 295:1657–1658
44. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9
45. Rodriguez-Mias RA, Pellecchia M (2003) Use of selective Trp side chain labeling to characterize protein-protein and protein-ligand interactions by NMR spectroscopy. J Am Chem Soc 125:2892–2893
46. Löhr F, Katsemi V, Betz M, Hartleib J, Rüterjans H (2002) Sequence-specific assignment of histidine and tryptophan ring ^{1}H, ^{13}C and ^{15}N resonances in ^{13}C/^{15}N- and ^{2}H/^{13}C/^{15}N-labelled proteins. J Biomol NMR 22:153–164
47. Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada I (2000) A novel NMR method for determining the interfaces of large protein-protein complexes. Nat Struct Biol 7:220–223
48. Meissner A, Sorensen OW (1999) Optimization of three-dimensional TROSY-type HCCH NMR correlation of aromatic ^{1}H-^{13}C groups in proteins. J Magn Reson 139:447–450
49. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Guntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57
50. Ruschak AM, Kay LE (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. J Biomol NMR 46:75–87
51. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622
52. Gelis I, Bonvin A, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131: 756–769
53. Religa TL, Kay LE (2010) Optimal methyl labeling for studies of supra-molecular systems. J Biomol NMR 47: 163–169
54. Ruschak AM, Velyvis A, Kay LE (2010) A simple strategy for ^{13}C,^{1}H labeling at the Ile-gamma 2 methyl position in highly deuterated proteins. J Biomol NMR 48:129–135
55. Religa TL, Sprangers R, Kay LE (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science 328:98–102
56. Kigawa T, Muto Y, Yokoyama S (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. J Biomol NMR 6:129–134
57. Etezady-Esfarjani T, Hiller S, Villalba C, Wüthrich K (2007) Cell-free protein synthesis of perdeuterated proteins for NMR studies. J Biomol NMR 39:229–238
58. Goerke AR, Swartz JR (2009) High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. Biotechnol Bioeng 102:400–416
59. Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE (2010) The proteasome antechamber maintains substrates in an unfolded state. Nature 467:868–871