Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 165 (2003) 237–247
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

# NMR assignment of protein side chains using residue-correlated labeling and NOE spectra

Geoffrey A. Mueller, Thomas W. Kirby, Eugene F. DeRose, and Robert E. London*

*National Institute of Environmental Health Sciences, Laboratory of Structural Biology, 111 Alexander Drive, P.O. Box 12233 MD MR-01, Research Triangle Park, NC 27709, USA*

Received 30 April 2003; revised 20 August 2003

Communicated by Christian Griesinger

**Abstract**

A new approach for the isotopic labeling of proteins is proposed that aims to facilitate side chain resonance assignments. Residue-correlated (RC) labeling is achieved by the expression of a protein on a medium containing a mixture of labeled, e.g., [U-$^{13}$C,$^{15}$N]amino acids, and NMR silent, [U-$^{2}$H]amino acids. *De novo* synthesis of amino acids was suppressed by feedback inhibition by the amino acids in the growth medium and by the addition of β-chloro-L-alanine, a transaminase inhibitor. Incorporation of these amino acids into synthesized proteins results in a relative diminution of inter-residue NOE interactions and a relative enhancement of intra-residue NOEs. Comparison of the resulting NOE spectra with those obtained from a uniformly labeled sample allows identification of intra-residue NOE peaks. Thus, this approach provides direct information for sidechain assignments in the NOE spectra, which are subsequently used for structural analysis. We have demonstrated the feasibility of this strategy for the 143 amino acid nuclease inhibitor NuiA, both at 35 °C, corresponding to a rotational correlation time of 9.5 ns, and at 5 °C, corresponding to a rotational correlation time of 22 ns.


## 1. Introduction

The determination of protein structure by NMR spectroscopy typically involves resonance assignment followed by assignment and interpretation of NOESY data. Assignment of the sidechain resonances is often a critical and rate limiting step in this process. The most useful approaches to sidechain assignments have involved TOCSY and COSY experiments, specifically the HCCH–TOCSY [1,2], the related H(CCO)NH– and (H)C(CO)NH–TOCSY [3,4], and the HCCH–COSY experiments [5,6]. The primary limitation of TOCSY experiments results from the limited magnetization transfer efficiencies that decline with increasing molecular size. The magnetization transfer efficiency has been calculated to be ~14% for a protein of 13 kDa to ~8% for a protein of 30 kDa and 4% for a 50 kDa protein [7]. The low transfer efficiency and rapid transverse relaxation of protonated carbons in larger proteins limits the signal-to-noise that can be attained in these experiments, and thus the utility of the TOCSY experiments in general [8]. In addition, the need to irradiate a large bandwidth in the TOCSY experiment leads to high power requirements with resultant sample heating. Spectral overlap can also be a problem in the HCCH–TOCSY experiment due to limited dispersion in the $^{13}$C dimension. Enhanced resolution can be obtained by utilizing $^{15}$N dispersion with the related H(CCO)NH– and (H)C(CO)NH–TOCSY experiments, but the transfer efficiency problem is exacerbated by the need to transfer polarization via the smaller $^{1}J_{NC}$ coupling interactions, and the carbonyl CSA relaxation limits the use of these experiments with larger proteins.

Many strategies to improve the TOCSY experiments have been proposed, in order to expand the range of applicability to larger proteins. Partial deuteration has been used to improve the relaxation characteristics of the protein sample and the H(CCO)NH– and (H)C(CO)NH–TOCSY experiments were appropriately optimized

* Corresponding author. Fax: 1-919-541-5707.
*E-mail address:* london@niehs.nih.gov (R.E. London).

1090-7807/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2003.08.006

for the sample [8]. However, the use of partial deuteration leads to multiple isotopomers with resulting spectral complications and ultimately the transfer efficiency will be limiting. Alternatively, the C(CC)(CO)NH–TOCSY experiment has been used to assign the $^{13}C$ sidechains in U-$^{2}H$,$^{13}C$,$^{15}N$-labeled molecules [9], followed by correlating the $^{1}H$ and $^{13}C$ shifts with HCCH–TOCSY data obtained on U-$^{13}C$,$^{15}N$-labeled samples. This approach also is limited by the fast transverse relaxation of protonated carbons in the doubly labeled proteins [8]. The HCCH–COSY is also unsuitable for larger proteins because the experiment is limited by the complexity and overlap of the phase sensitive data obtained.

Alternate strategies for sidechain assignment involve the use of short mixing times in the NOESY experiment to select the more proximate, intra-residue crosspeaks, and finally the selective labeling of individual residue types [10,11]. The use of short mixing times in NOESY experiments will select resonances based on inter-proton distances and this approach is subject to significant errors since the assumption that intra-residue $^{1}H$–$^{1}H$ distances are always less than inter-residue $^{1}H$–$^{1}H$ distances does not always hold. Fischer et al. [7] have proposed that the $^{13}C$–$^{13}C$ NOESY experiment can be a preferable alternative to TOCSY for larger proteins. This would compensate for transfer efficiency problems, but would still suffer from the poorer $^{13}C$ dispersion and spectral overlap.

Several groups have prepared proteins that contain specific protonated amino acids in a background of deuterated amino acids [10,12–18]. In these studies, the primary goals were spectral simplification or selective NOE information. However, assignment information can also be derived from NOE studies on such proteins, since the ratio of intra/intermolecular NOE values will be enhanced. Nevertheless, the use of these approaches to obtain sidechain assignments is not general, and can fail even if only a single residue type is protonated. This will occur if the inter-residue $^{1}H$–$^{1}H$ distances for particular pairs of protons on the protonated residues happen to be shorter than the intra-residue $^{1}H$–$^{1}H$ distances for that particular proton pair. Further, in order to provide a general assignment strategy, it would be necessary to prepare 20 samples, each containing a particular protonated amino acid type in a deuterated background.

In the present study, we propose an alternative approach for sidechain assignment which is based on the introduction of a residue-correlated (RC) labeling pattern achieved by expression on a medium containing a mixture of [U-$^{13}C$, $^{15}N$]amino acids and [U-$^{2}H$]amino acids. This approach seeks to position "NMR-active" residues that are U-$^{1}H$,$^{13}C$,$^{15}N$-labeled among "NMR-silent" residues that are U-$^{2}H$-labeled. In the following discussion, we refer to this labeling distribution as residue-correlated (RC) labeling. This type of labeling can be readily achieved by growth of the *Escherichia coli* or other organism expressing the protein on a medium containing a mixture of [U-$^{13}C$,$^{15}N$]amino acids and [U-$^{2}H$]amino acids in a ratio $\alpha$:$1-\alpha$, where $\alpha$ is the fractional labeling parameter. For such a labeled protein, the intensities of the intra-residue NOEs will be $\alpha I_0$, while the intensities of the inter-residue $^{1}H$–$^{1}H$ NOEs will equal $\alpha^2 I_0$, where $I_0$ would be the intensity of the NOE peaks in the U-[$^{13}C$,$^{15}N$]-labeled sample. The decrease in spin diffusion resulting from the presence of the $1-\alpha$ fraction of deuterated residues will reduce the intensity losses resulting from the dilution of protonated residues in a manner which will depend on the fractional labeling parameter $\alpha$. In this paper, we evaluate the success of this labeling approach with a protein that has been studied extensively in our laboratory, the nuclease A inhibitor (NuiA) from *Anabaena* sp.

## 2. Methods

NuiA protein was over-expressed in *E. coli* and purified by the published method [19,20]. Residue-correlated labeling was effected by growing the culture in a medium consisting of two parts completely deuterium-labeled medium (Martek 9-d > 98% from Spectra Stable Isotopes, Columbia, MD) and 1 part $^{13}C$-, $^{15}N$-labeled medium (Martek 9-CN). Although, it was anticipated that the presence of a full complement of amino acids in the growth medium would suppress *de novo* synthesis, we also added 100 μM β-chloro-L-alanine to inhibit transaminases [11]. In general, the presence of a 1:2 ratio of [U-$^{13}C$]D-glucose:[U-$^{2}H$]D-glucose present in both media, would be expected to result in a complex labeling pattern of newly synthesized amino acids [21]. Since the Martek 9-d media also contained $D_2O$, the final water composition of the growth medium was 67% deuterated. The composition of the water is presumably not relevant if *de novo* amino acid synthesis is fully suppressed, but the deuteration may prove useful if there is some *de novo* synthesis, as is inevitable for glutamate and aspartate using the approach described here. The NMR sample contained 3 mM protein in 90 mM deuterated Tris–HCl, pH 7, 10% $D_2O$.

NMR spectra were recorded on a Varian INOVA 600 MHz instrument at 5 or 35 °C as noted in the text. The spectra were processed with NMRPipe [22] and analyzed with NMRView [23] software on LINUX workstations running Red Hat Linux v. 7.3. The CN NOESY–HSQC [24] was acquired using the RC labeled protein and compared to that analyzed previously for structure determination. Analysis of the NOESY spectra obtained on the variously labeled samples was based on our previous assignments for NuiA [19], which utilized the program ARIA, version 1.0 [25]. The maximum of the peak was used as the intensity. HNCA spectra [26]

were also collected on both samples and the intensities of the intra- and inter-residue Cα peaks were analyzed. Intensities were calculated from ellipsoidal fitting of peak volumes. This differs from the approach used for the NOESY data because the HNCA has sufficient dispersion to allow accurate quantitation with this method.

## 3. Results and discussion

The philosophy behind the proposed residue-correlated labeling approach is most easily illustrated by considering the labeling patterns for a pair of interacting residues, such as the valine–phenylalanine pair shown in Fig. 1. Using a labeling pool of 33% [U-$^{13}$C,$^{15}$N]amino acids and 67% [U-$^{2}$H]amino acids, the expected labeling probabilities for the protein produced from growth on this media are indicated. Thus, 11% of the molecules have both residues labeled, 22% have protonated valine/deuterated phenylalanine, 22% have protonated phenylalanine/deuterated valine, and 44% have both residues deuterated. Considering first the valine residue, only the first two species will contribute to the observed valine NOE spectrum, and only the first will contribute inter-residue NOEs. Hence, the relative inter/intra-residue NOE intensities for valine will be statistically reduced by a factor of 2/3, so that a comparison of the NOE spectra obtained using uniform and residue-correlated labeling patterns should allow straightforward assignment of intra-residue NOEs. Analogous arguments can be made for the phenylalanine NOE patterns.

### 3.1. Characterization of the NuiA labeling

In order to evaluate the success of the labeling approach, we compared the data from the CN–NOESY for U-$^{13}$C,$^{15}$N-labeled sample with the data from the RC-$^{2}$H,$^{13}$C,$^{15}$N-labeled sample. We previously used the automated assignment program ARIA to analyze the data obtained in the CN–NOESY experiments on the U-$^{13}$C,$^{15}$N-labeled sample [25]. These assignments served as the basis for our previous structure determination and for analyzing the spectra recorded on the RC-[$^{2}$H,$^{13}$C,$^{15}$N]NuiA samples. We initially focused on the $^{15}$N-separated part of the CN–NOESY since it exhibits the best dispersion, providing the most straightforward analysis. ARIA was able to assign 1694 NOE crosspeaks in the $^{15}$N NOESY spectra. These could be subdivided into amide–amide crosspeaks (280), ambiguous crosspeaks with three or more possible assignments (323), and crosspeaks corresponding to intra- and inter-residue interactions with less ambiguity (1091). Although the residue-correlated labeling protocol involved using media containing ~66% $^{2}H_2O$, the protein was ultimately purified using natural abundance $H_2O$ buffers, so that the amide groups would be predominantly protonated. Further, since amide protons are generally labile, any deuteration of these positions will be nearly random and not subject to correlated labeling. The remaining 1091 NOEs are shown in Table 1 according their classification.

The NOEs in [U-$^{13}$C,$^{15}$N] and RC-[$^{2}$H,$^{13}$C,$^{15}$N] samples were compared to match them with the assignments

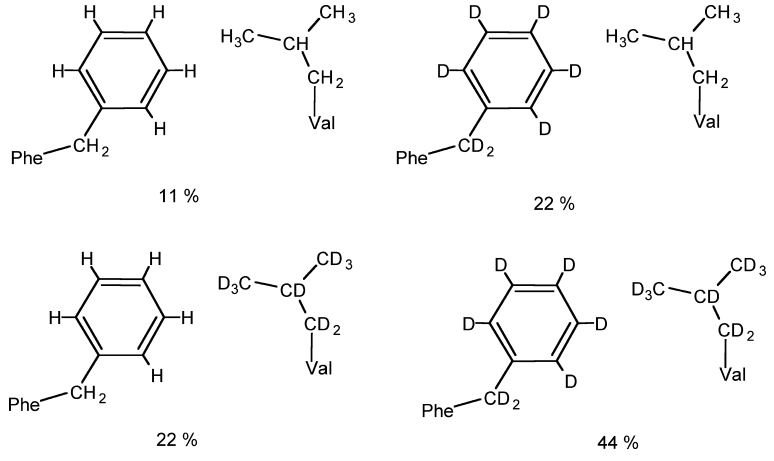


Fig. 1. Illustration of residue-correlated labeling for a pair of interacting valine and phenylalanine residues. The percentages shown correspond to growth on a medium containing 67% [U-$^{2}$H]amino acids and 33% [U-$^{13}$C,$^{15}$N]amino acids.

Table 1
Summary of NOE peaks in the $^{15}$N-edited NOESY identified by ARIA in NuiA

| NOE classification | NOEs identified in [U-$^{13}$C,$^{15}$N]NuiA | NOEs identified in RC-sample (%) |
|---|---|---|
| Intra-residue | 329 | 292 (89) |
| Inter-residue | 762 | 492 (65) |

in the fully labeled sample. By a match, we mean that the same peak can be identified in both spectra. In total, 72% of the NOEs in the fully labeled sample were identified in the RC-labeled sample, however there are significant differences between the two classifications as can be seen in Table 1. Eighty-nine percent of the intra-residue NOEs were found versus 65% of the inter-residue NOEs. This automated analysis is thus qualitatively consistent with expectations of the effects of residue-correlated labeling.

In the RC-labeled sample, some of the NOE peaks become sufficiently weak so that they are no longer identified by ARIA, while the remaining peaks show changes in relative intensity. As discussed above, it was anticipated that the inter-residue crosspeaks should be reduced by a multiplicative factor of $1-\alpha$ relative to the intra-residue crosspeaks. A comparison of the NOESY patterns obtained for the uniform and residue-correlated labeling patterns should facilitate the observation of these changes in intensity. In Fig. 2, we compare several NOE strips for three different residues: V122, L28, and G126. The first two strips (V122 and L28) are taken from the $^{15}$N-separated part of the CN–NOESY and G126 is taken from the $^{13}$C-separated part. The assignments of the intra-residue NOE crosspeaks are labeled in each strip. It is apparent that in each case the intensities of the inter-residue aliphatic peaks are significantly reduced in comparison to the intra-residue aliphatic peaks. It is particularly impressive that the long aliphatic sidechain of a leucine residue is assignable, when comparing the uniform and RC-labeled NOESY spectra (Fig. 2, residue L28). We note that due to the full protonation of amides after the sample is prepared in $H_2O$, the amide–amide and amide (donor)–aliphatic (acceptor) NOEs will be reduced by the same factor as the intra-residue NOEs, while the inter-residue aliphatic–aliphatic and aliphatic (donor)–amide (acceptor) NOEs will be reduced by $\alpha^2$ compared with the intra-residue reduction of $\alpha$ (see Table 2). These trends can be observed in the spectra shown in Fig. 2. Despite the 2/3 dilution of the [U-$^{1}$H,$^{13}$C,$^{15}$N]residues giving rise to the signals, the signal-to-noise of the spectra from the residue-correlated labeling sample remains high. As discussed above, this reflects a combination of factors, particularly the fact that we are observing intra-residue NOEs that generally correspond to shorter interproton distances, as well as the improved relaxation characteristics that result from the statistical deuteration of neighboring residues. As a rough estimate of this effect, Pachter *et al.* have calculated that for protonated amino acids in a deuterated protein, the intra-residue NOE intensities will be enhanced by ~50% relative to the values in the fully protonated protein [27]. The deuterated background probably also increases the extent of spin-diffusion within the protonated amino acids, explaining the greater intensity of the NOE crosspeaks from the more remote sidechain protons in the RC sample (see Fig. 2).

In order to quantify the relative enhancement of the intra- to inter-residue NOE intensities using RC labeling, we followed the approach described below. Each residue was treated individually and the ratio of the crosspeak intensities for the RC sample relative to the

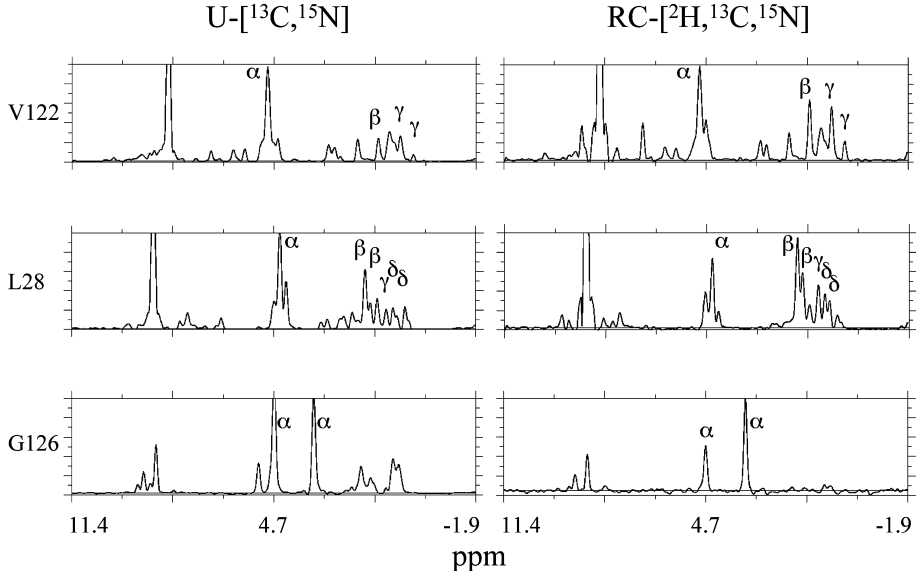


Fig. 2. Comparison of the F1 dimension slices from NOESY spectra taken on a uniformly labeled (U-[$^{13}$C,$^{15}$N]) versus a residue correlated (RC-[$^{2}$H,$^{13}$C,$^{15}$N]) samples. Residues Val122 and Leu28 are compared with 1D strips taken from the $^{15}$N-separated NOESY. The Gly126 spectra is taken at the C$\alpha$ chemical shift in the $^{13}$C-separated NOESY.

Table 2
Inter-residue NOE probabilities in samples with residue correlated labeling

| Interacting (donor) residue | Observed (acceptor) residue | Probabilitity (for $\alpha = 1/3$) | Inter-residue NOE? | NOE relative to uniform sample |
|---|---|---|---|---|
| *I: Amide–Amide* | | | | |
| 15NH | 15NH | 1/9 | Yes | |
| 14NH | 15NH | 2/9 | Yes | |
| 15NH | 14NH | 2/9 | No | |
| 14NH | 14NH | 4/9 | No | |
| | | | | 1/3 |
| *II: Aliphatic–aliphatic* | | | | |
| 13CH | 13CH | 1/9 | Yes | |
| 12CD | 13CH | 2/9 | No | |
| 13CH | 12CD | 2/9 | No | |
| 12CD | 12CD | 4/9 | No | |
| | | | | 1/9 |
| *III: Aliphatic–Amide* | | | | |
| 13CH | 15NH | 1/9 | Yes | |
| 12CD | 15NH | 2/9 | No | |
| 13CH | 14NH | 2/9 | No | |
| 12CD | 14NH | 4/9 | No | |
| | | | | 1/9 |
| *IV: Amide–Aliphatic* | | | | |
| 15NH | 13CH | 1/9 | Yes | |
| 14NH | 13CH | 2/9 | Yes | |
| 15NH | 12CD | 2/9 | No | |
| 14NH | 12CD | 4/9 | No | |
| | | | | 1/3 |

uniformly labeled sample was averaged for the intra- and inter-residue peaks separately. Mathematically this can be expressed for the intra-residue peaks

$$R_{\text{intra},i} = \text{mean}(I_{\text{intra,RC}}/I_{\text{intra,U}}) \quad (1)$$

and similarly for the inter-residue peaks

$$R_{\text{inter},i} = \text{mean}(I_{\text{inter,RC}}/I_{\text{inter,U}}), \quad (2)$$

where $R_i$ is the ratio for each residue type ($i$) and $I$ is the intensity of the NOE crosspeak for the uniformly (U) and residue-correlated (RC) labeled samples. If the labeling was successful, $I_{\text{inter,RC}}$ should be reduced to a greater extent ($\alpha^2$), relative to $I_{\text{inter,U}}$ while $I_{\text{intra,RC}}$ is reduced ($\alpha$), relative to $I_{\text{intra,U}}$. The relative enhancement of the intra- to inter-residue NOE using this labeling scheme is therefore the ratio of $R_{\text{intra}}/R_{\text{inter}}$. We expect $R_{\text{intra}}/R_{\text{inter}}$ to be greater than 1 if the labeling was successful because the $R_{\text{inter}}$ should be smaller than $R_{\text{intra}}$. Note that by taking in essence a ratio of ratios, any normalization factor between the two experiments is divided out. Overall, for each residue for which there were data to compare, the average $R_{\text{intra}}/R_{\text{inter}}$ was $1.8 \pm 1.0$. Thus we obtained a large effect, but also a large standard deviation. To better understand this large variation the ratios were grouped according to residue type as shown in Table 3.

In order to use this labeling strategy as an effective replacement for the TOCSY experiment, it is critical that the ratio be greater than 1 in all cases. This would allow an analyst to confidently differentiate intra- from inter-residue crosspeaks. Naively, the following residue types would be considered a success because the mean

Table 3
Ratio of intensities for Inter/Intra-residue NOE interactions

| Residue type | Mean ratio $R_{\text{intra}}/R_{\text{inter}}$ | SD | Missing[a] |
|---|---|---|---|
| Asp | 0.9 | 0.2 | 5 |
| Glu | 1.0 | 0.4 | 28 |
| Ser | 1.1 | 0.2 | 19 |
| Tyr | 1.1 | 0.2 | 4 |
| Trp | 1.1 | 0.3 | 11 |
| Asn | 1.2 | 0.6 | 44 |
| Gln | 1.3 | 1.1 | 62 |
| Phe | 1.4 | 0.7 | 18 |
| Gly | 1.8 | 0.8 | 18 |
| Ala | 1.9 | 0.6 | 16 |
| Thr | 2.1 | 0.4 | 20 |
| Met | 2.4 | NA | 1 |
| Leu | 2.5 | 0.6 | 17 |
| Lys | 2.5 | 0.8 | 8 |
| Ile | 2.6 | 1.1 | 9 |
| Val | 2.8 | 1.3 | 24 |
| Arg | 4.0 | 0.8 | 3 |

See notes in text.
[a] Missing indicates NOE peaks that were present in the uniformly labeled sample but are not present in RC labeled sample.

ratio ±1 SD is still greater than 1: Gly, Ala, Thr, Met, Leu, Lys, Ile, Val, and Arg (Table 3). This statistic underestimates the success of the labeling approach, since some of the resonances are sufficiently reduced in intensity in the RC NOE spectra to be below the detection threshold, and hence were not considered. This effect is apparent in the spectra of the residues of the RC sample shown in Fig. 2, and leads to a 'survivor bias' of the analysis. Mathematically, the ratio of the above ratios is given by:

$$\frac{R_{\text{intra}}}{R_{\text{inter}}} = \left(\frac{I_{\text{intra,RC}}}{I_{\text{intra,U}}}\right)\left(\frac{I_{\text{inter,U}}}{I_{\text{inter,RC}}}\right). \tag{3}$$

For inter-residue NOE peaks which drop below the detection threshold, the $I_{\text{inter,RC}}$ in the denominator approaches zero, which would lead to a larger ratio. However, since these cases are not counted in the comparison, the survivor results in a reduced value for $R_{\text{intra}}/R_{\text{inter}}$. As mentioned above, ARIA identified a greater percentage of intra- than inter-residue NOE crosspeaks. To assess the extent of such a significant 'survivor bias', the number of times a particular residue type had missing NOEs in the RC spectra is listed in the final column of Table 3. The mean number of peaks missed by residue type is $18 \pm 16$. The missing number of NOE peaks for residues Asn and Gln are more than one standard deviation from the mean, indicating that for these residues there is a 'survivor' bias and that side chain assignments can be made confidently for these residues as well.

The remaining residue types, Asp, Glu, Ser, Tyr, Phe, and Trp require further consideration. Since glucose was present in our growth medium, there could be some *de novo* synthesis of amino acids from Krebs cycle intermediates. Such amino acids would in general exhibit a complex labeling pattern. The direct amination of TCA cycle intermediates α-ketoglutarate and fumarate to produce glutamate and aspartate, respectively, represents a potential means for dilution of these amino acid pools which obviates the transaminase block [28]. The dilution effects on the aromatic amino acids appears to indicate that there were insufficient concentrations of at least one of these in the growth medium to effectively suppress *de novo* synthesis, or as suggested above, failure of β-chloro-L-alanine to suppress the particular transaminases involved in aromatic amino acid synthesis.

In order to further characterize the residue-correlated labeling, we compared HNCA data obtained on both the uniformly labeled and the residue-correlated labeled samples. This experiment typically identifies crosspeaks between the observed amide proton and both the intra-residue, $C\alpha_i$ as well as $C\alpha_{i-1}$ of the preceding residue. We compared the ratio $R_{\text{HNCA}}$ of $\text{HNCA}_i/\text{HNCA}_{i-1}$ intensities in the uniformly (U) and residue-correlated (RC) samples according to:

$$R_{\text{HNCA}} = \frac{(\text{HNCA}_i/\text{HNCA}_{i-1})^{\text{RC}}}{(\text{HNCA}_i/\text{HNCA}_{i-1})^{\text{U}}} = \left(\frac{\text{HNCA}_i^{\text{RC}}}{\text{HNCA}_i^{\text{U}}}\right)\left(\frac{\text{HNCA}_{i-1}^{\text{U}}}{\text{HNCA}_{i-1}^{\text{RC}}}\right). \tag{4}$$

This ratio corrects for any coupling constant variations, conformational effects, etc., and so will vary from 1.0 only as a result of differences in labeling pattern. If there is no significant *de novo* amino acid synthesis, $\text{HNCA}_i^{\text{RC}}$ will be 1/3 the intensity of the uniformly labeled sample due to the effect of dilution, while $\text{HNCA}_{i-1}^{\text{RC}}$ should be 1/9 of the value in the uniformly labeled sample, for a mean ratio $R_{\text{HNCA}} = 3.0$. The measured mean value for this $R_{\text{HNCA}}$ was $3.5 \pm 3.2$. Hence, the results are consistent with expectations, although the error is large.

In order to further analyze these data, we consider the effects of *de novo* amino acid synthesis on both $\text{HNCA}_{i-1}^{\text{RC}}$ and $\text{HNCA}_i^{\text{RC}}$. For the $\text{HNCA}_{i-1}^{\text{RC}}$ correlations, the labeling of the $C\alpha_{i-1}$ position will in general be 1/3 whether Cα is derived from an amino acid added to the medium or whether it is synthesized *de novo*. This results from the fact that the glucose present in the medium will have a mean labeling level that is the same as the amino acid pool, i.e., 1/3 [U-$^{13}$C]D-glucose plus 2/3 [U-$^{2}$H]D-glucose. Thus, the ratio $\text{HNCA}_{i-1}^{\text{RC}}/\text{HNCA}_{i-1}^{\text{U}}$ is predicted to equal $\alpha^2$ or 1/9 whether or not *de novo* synthesis occurs. Alternatively, observation of the $\text{HNCA}_i$ correlation requires that *both* the amide nitrogen and Cα carbon of a given residue be labeled. Amino acids synthesized *de novo* by either direct amination or transamination will only have both positions labeled (1/3)*(1/3) = 1/9 of the time. Hence, $\text{HNCA}_i^{\text{RC}}$ will in general be reduced by *de novo* synthesis. We have calculated the ratio ($\text{HNCA}_i^{\text{U}}/\text{HNCA}_i^{\text{RC}}$) for each amino acid type, and normalized the results by the average value for all of the residues (Table 4). Based on the above discussion, this ratio will be smaller if the amino acids supplied in the medium are incorporated into the NuiA, and larger if there is *de novo* synthesis, which would have the effect of reducing the denominator, as discussed above. As is apparent from Table 4, the results of this analysis are generally in agreement with the results of the analysis of the NOE data. The Glu, Asp, and Asn residues gave the highest ratios, consistent with significant *de novo* biosynthesis, and the aromatic residues also gave values above the average, while the other aliphatic residues generally exhibited ratios below the average. It may be that the inhibition of phenylpyruvic transaminase by β-chloro-L-alanine is incomplete, explaining the poorer results with aromatic residues.

However, although some of the residues show evidence of dilution due to *de novo* biosynthesis, the expected suppression of inter-residue NOEs was still observable. Fig. 3 shows F1 dimension slices from NOESY spec-

Table 4
Analysis of HNCA experiment on uniformly labeled and residue-correlated labeled samples

| Residue *i* | $\left(\frac{HNCA_i^U}{HNCA_i^{RC}}\right)_N^a$ | Variance |
|---|---|---|
| Arg | 0.36[b] | — |
| Ile | 0.53 | 0.04 |
| Lys | 0.78 | 0.01 |
| Trp | 0.78 | 0.02 |
| Val | 0.79 | 0.04 |
| Leu | 0.82 | 0.02 |
| Thr | 0.83 | 0.01 |
| Met | 0.84[b] | — |
| Ala | 0.85 | 0.05 |
| Gln | 0.89 | 0.08 |
| Gly | 1.08 | 0.13 |
| Tyr | 1.16 | 0.02 |
| Ser | 1.19 | 0.35 |
| Phe | 1.30 | 0.14 |
| Glu | 1.37 | 0.28 |
| Asn | 1.64 | 0.36 |
| Asp | 1.78 | 0.49 |

[a] Values of the ratio for each residue type have been normalized by dividing each ratio by the mean value computed for all of the residues.
[b] Data derived from a single residue.

tra taken on a uniformly labeled (U-[$^{13}$C,$^{15}$N]) versus a residue-correlated (RC-[$^{2}$H, $^{13}$C,$^{15}$N]) samples for Tyr35, Phe66, and Tyr113. The first two rows show 1D strips taken at the Cα chemical shift in the $^{13}$C-separated NOESY, while for Tyr 113, the 1D NOESY strips are taken at the Cβ1 chemical shift in the $^{13}$C-separated NOESY. Although the residue-correlated labeling approach was statistically less successful for the aromatic residues, it is apparent from these spectra that there is nevertheless a significant reduction in the relative intensities of the inter/intra-residue NOEs.

### 3.2. Assignments of sidechain carbon resonances

To demonstrate that this method is viable as a replacement for the TOCSY approach, we show in Fig. 4 how these spectra could be used to assign side chain resonances. The spectra shown in column 4a are NOESY strips originating from the labeled atom for Val49 from the $^{15}$N and $^{13}$C cubes of the CN–NOESY using the RC-labeled sample. The high similarity among these strips supports the conclusion that this methodology can be used to assign side chains. Our initial assumption is that sequential assignments would be made using standard backbone scalar coupling experiments such as HNCACB [29], and the Cα and Cβ frequencies would already be known. This information then can be used to assign the Hα and Hβ resonances. In Fig. 4b we simulate a search for a $^{13}$C-separated NOESY strip with a similar pattern of resonances to the $^{15}$N-separated strip. Fig. 4b shows the 3 strips with Cα frequencies close to that assigned for Val49 from the HNCACB. The third strip is a clear match for the Hα strip, while the other two belong to Ile16 and Ile71, and are clearly not good matches. A similar protocol can be used for searching for the Hβ. Assignment of the carbon resonances at the γ position (or further) requires an alternative strategy. For the example of Val49, reasonable choices can be made for the proton frequencies given the consistency in the first three strips of 4a. The other two resonances, which are enhanced relative to the uniformly labeled sample, presumably arise from the gamma protons of valine. If we search the $^{1}$H{$^{13}$C} HSQC for proton carbon pairs that are likely to be valine near the most upfield proton frequency in the Hβ strip, there are five possibilities and their $^{13}$C NOESY strips are shown in Fig. 4c. It is

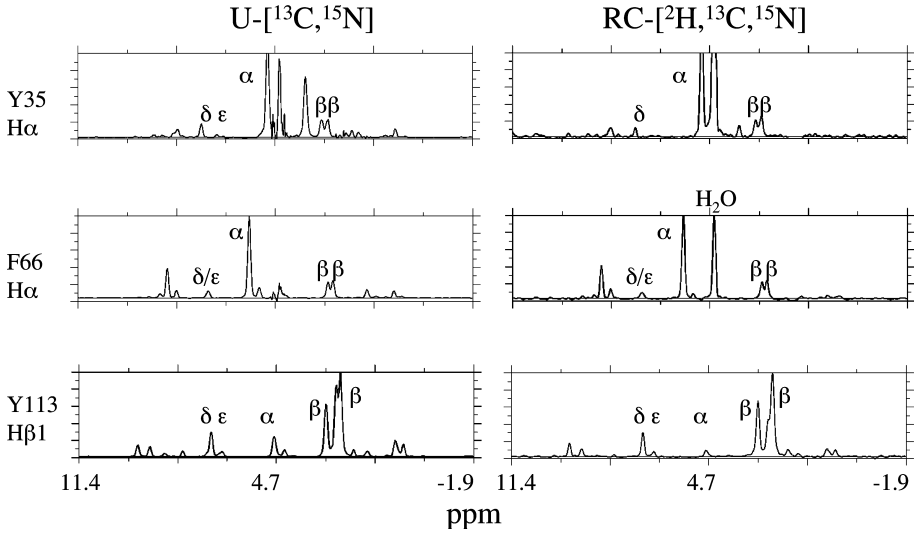


Fig. 3. Comparison of the F1 dimension slices from NOESY spectra taken on a uniformly labeled (U-[$^{13}$C,$^{15}$N]) versus a residue correlated (RC-[$^{2}$H, $^{13}$C,$^{15}$N]) samples. Residues Tyr35 and Phe66 are compared with 1D strips taken at the Cα chemical shift in the $^{13}$C-separated NOESY. The Tyr113 spectra are taken at the $C\beta_1$ chemical shift in the $^{13}$C-separated NOESY.

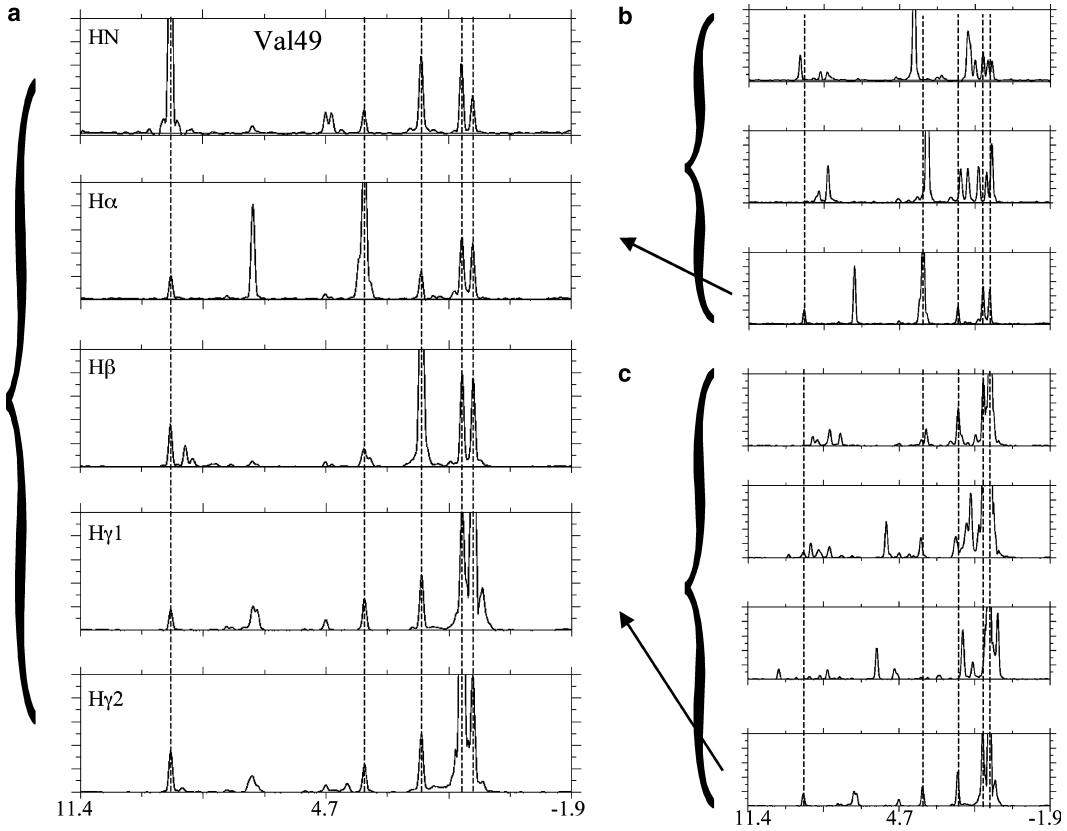


Fig. 4. An example of how to assign sidechains using NOE spectra from an RC-labeled sample. Part (a) shows all the F1 dimension slices for Val49 based on previous knowledge. Assuming the Cα and Cβ shift are already known from sequential assignments, (b) shows all the possible Cα spectra to match with the HN spectra in order to find the correct Hα chemical shift for Val49. The correct match is indicated with an arrow. After assigning the Hα and Hβ, the Hγ shifts can be tentatively assigned from the HN, Hα, and Hβ spectra in (a). Part (c) demonstrates how to find the Cγ chemical shift of Val49, by searching all methyl carbon strips at the assumed shift of $H\gamma_1$. One spectrum is a clear match (indicated with an arrow) with the HN, Hα, and Hβ spectra in (a). The $H\gamma_2$ chemical shift can be found in a similar fashion.

reasonably simple to match the pattern for the HN, Hα, and Hβ with the last strip shown in 4c. This demonstrates that it would be possible to assign long side chains using this methodology.

### 3.3. Application to larger proteins

In order to evaluate the feasibility of applying this assignment approach to a larger protein, we obtained TOCSY and NOESY data on uniformly labeled and residue-correlated NuiA, respectively, at 5 °C. The NuiA protein has a rotational correlation time of 9.5 ns at 35 °C [19]. Using Stokes Law with $T = 5$ °C and the viscosity of water at 5 °C, the estimated correlation time of NuiA is 22 ns at this temperature. This is comparable with the rotational correlation time of 23 ns for the 42 kDa complex of maltose binding protein with β-cyclodextrin at 37 °C [30]. In the comparison reported here, the NOESY spectra with the RC-labeled sample containing 33% [U-$^{13}$C,$^{15}$N]residues had superior signal to noise and the long aliphatic side chains were readily assignable (Figs. 2 and 5). In contrast, the H(CCO)NH–TOCSY spectra obtained on [U-$^{13}$C,$^{15}$N]NuiA (Fig. 5a,c,e) had insufficient signal/noise to allow many of the sidechain assignments to be made. More significantly, the NOESY spectra for Y85 (Fig. 5d) contain assignment information for the aromatic proton resonances, which are not obtained from the standard TOCSY experiments.

We have attempted to assign many more of the side chains with reasonable success. The major limitation arose for the residues that failed to show a clear correlated labeling pattern, as discussed above. However, although statistically less successful, the aromatic residues generally showed a significant relative enhancement of intra/inter-residue NOE intensities (e.g., Fig. 3), and the NOE resonances for Asp and Glu are generally very difficult to analyze due to poor shift dispersion. An additional limitation arose when the Hα frequency was close to water, making it difficult to find good spectra in the NOESY data due to the water suppression. Obtaining a $^{13}$C-separated NOESY in $D_2O$ would alleviate this problem. This is a problem that does not occur with the TOCSY assignment methodology. Potentially, one could collect the HNHA spectrum [31] in cases where it is hard to assign the Hα, and this offers the advantage of

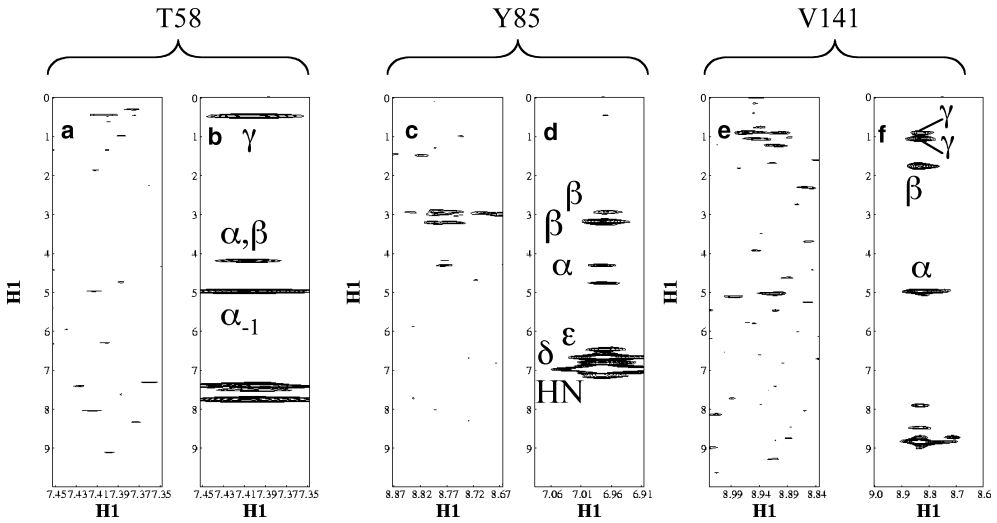


Fig. 5. Comparison of TOCSY data using uniformly labeled sample with NOESY data using residue-correlated labeling at 5 °C. The strips shown are taken from the F1 dimension of the H(CCO)NH–TOCSY (a,c,e) using a [U-$^{13}$C,$^{15}$N]sample and compared with corresponding strips from the F1 dimension of the NOESY (b,d,f) using the RC-labeled sample. Both data sets were acquired at 5 °C. The strips correspond to T58 (a,b), Y85 (c,d), V141 (e,f).

providing structural data about the phi angle. We note that this assignment strategy does not necessarily require an RC labeled sample. It may be possible to assign sidechains using the NOESY in a uniformly labeled sample, since the intra-residue NOE correlations are usually intense. However, as shown in Fig. 2, the RC labeling scheme greatly simplifies the approach by enhancing the intra-residue NOE correlations.

## 4. Conclusions

In the present study, we have demonstrated that residue-correlated labeling can be achieved in an *E. coli* expression system by growth on media that contain a mixture of [U-$^{13}$C,$^{15}$N] and [U-$^{2}$H]amino acids. The results for the protein NuiA (MW = 16 kDa) demonstrate that the approach worked extremely well for most residues, with poorer results for the aromatic residues, glutamate, aspartate, and serine. However, even for many of these amino acids, we were able to observe significant changes in the relative intra/inter-residue NOE peak intensities. Advantages of this approach include: (1) Assignment information is derived directly from NOE spectra rather than from other types of spectra, eliminating registration problems; (2) There are no deuterium isotope shift corrections, since the deuteration is on the unobserved amino acid residues (with the exception of some small long-range effects on backbone resonances); (3). Although the intensity of the intra-residue spectra is reduced by 1/$\alpha$ (1/3 for the example given), the NOE peaks under observation correspond to the stronger, intra-residue contributions, and hence they should be observable even at the effectively reduced concentrations. Further, the presence of a significantly deuterated background will reduce intensity losses due to spin diffusion. We have previously found that while the H(CCO)NH and (H)C(CCO)NH–TOCSY experiments [3,4] work extremely well for sidechain assignments in smaller proteins or in larger proteins that are sufficiently stable for studies at higher temperatures, we encountered significant problems for larger proteins that could not be studied at high temperatures. Alternatively, the more limited dispersion and greater degree of spectral overlap of the HCCH–TOCSY experiment, relative to the H(CCO)NH and (H)C(CCO)NH–TOCSY experiments, has also limited the use of this strategy for higher molecular weight proteins [2,7]. Wagner and co-workers have used fractional deuteration and optimized TOCSY sequences to accomplish complete side chain assignments[8]. Their approach also provided limited assignment information for the alpha protons due to the combined effects of deuteration and the limitations of the TOCSY approach discussed in the literature [7], and the partially deuterated protein resonances will in general exhibit multiple $^{13}$C and $^{1}$H resonances due to isotope shifts. However, the RC labeling approach has no such limitations. In contrast, the residue-correlated labeling approach for sidechain assignments should be generally feasible for all situations in which a NOESY experiment with reasonable sensitivity can be achieved. As noted above, the reduction in intensity due to label dilution will tend to be offset by the fact that this strategy targets intra-residue NOE interactions, which generally correspond to shorter distances, and hence have greater intensity. The assignment data thus obtained can be directly supplied to programs such as ARIA for structural analysis [25]. Other improvements

in the analysis may be possible with more sophisticated assignment schemes that involve pattern-matching algorithms such as neural networks to cluster similar spectra like those shown in Fig. 4a. Improvements in the labeling of the protein approach such as the use of transaminase deficient strains of *E. coli* [32] and the use of cell free expression systems which would completely eliminate *de novo* amino acid synthesis [33] are currently under evaluation.

Finally, we note that an interesting variant of the proposed approach would utilize a mixture of [U-$^{13}$C,$^{15}$N] and [U-$^{15}$N]amino acids. In the absence of deuteration, intra- and inter-residue NOEs can be separately measured using 4D isotope-edited [29,34–36] and isotope-filtered NOE experiments [37,38]. The 4D $^{15}$N,$^{13}$C-separated NOESY experiment [29,34] in particular, could be used to select primarily for intra-residue NOEs in such a sample. Inter-residue crosspeaks arising in this 4D NOESY experiment could be identified by their $^{1}$H chemical shifts measured from a 3D $^{13}$C filtered, $^{15}$N-edited NOESY experiment. Such an approach would greatly facilitate the analysis of NOESY data, both by facilitating sidechain assignments as described in the present study, and by allowing the use of the $^{13}$C-filtered NOESY experiment to exclusively assign inter-residue NOEs. While allowing separation of inter-residue NOE interactions, the lower sensitivity of such an approach would significantly limit its applicability at the present time. However, the recent introduction of rapid sampling methodology and cryogenic probe technology for obtaining multi-dimensional NMR spectra could ultimately render such an approach practicable [39].

## Acknowledgments

The authors would like to thank Dr. Yuan Chen (City of Hope) and Dr. Ron Venters (Duke University) for a critical reading of the manuscript.

## References

[1] A. Bax, G.M. Clore, A.M. Gronenborn, H-1-H-1 Correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spedctra of C-13 enriched proteins, Journal of Magnetic Resonance 88 (1990) 425–431.

[2] M. Ikura, S. Spera, G. Barbato, L.E. Kay, M. Krinks, A. Bax, Secondary structure and side-chain H-1 and C-13 resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy, Biochemistry 30 (1991) 9216–9228.

[3] S. Grzesiek, J. Anglister, A. Bax, Correlation of Backbone amide and aliphatic side-chain resonances in C-13/N-15 enriched proteins by isotropic mixing of C-13 magnetization, Journal of Magnetic Resonance, Series B 101 (1993) 114–119.

[4] T.M. Logan, E.T. Olejniczak, R.X. Xu, S.W. Fesik, A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments, Journal of Biomolecular NMR 3 (1993) 225–231.

[5] D.M. Baldisseri, J.G. Pelton, S.W. Sparks, D.A. Torchia, Complete H-1 and C-13 assignment of Lys and Leu sidechains of staphylococcal nuclease using HCCH-COSY and HCCH-TOCSY 3D NMR spectroscopy, FEBS Letters 281 (1991) 33–38.

[6] G.M. Clore, A. Bax, P.C. Driscoll, P.T. Wingfield, A.M. Gronenborn, Assignment of the side-chain H-1 and C-13 resonances of interleukin-1 beta using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy, Biochemistry 29 (1990) 8172–8184.

[7] M.W.F. Fischer, L. Zeng, E.R.P. Zuiderweg, Use of $^{13}$C-$^{13}$C NOE for the assignment of NMR lines of larger labeled proteins at large magnetic fields, Journal of the American Chemical Society 118 (1996) 12457–12458.

[8] Y. Lin, G. Wagner, Efficient side-chain and backbone assignment in large proteins: application to tGCN5, Journal of Biomolecular NMR 15 (1999) 227–239.

[9] B.T. Farmer 2nd, R.A. Venters, Assignment of side-chain C-13 resonances in perdeuterated proteins, Journal of the American Chemical Society 117 (1995) 4187–4188.

[10] C.H. Arrowsmith, R. Pachter, R.B. Altman, S.B. Iyer, O. Jardetzky, Sequence-specific 1H NMR assignments and secondary structure in solution of Escherichia coli trp repressor, Biochemistry 29 (1990) 6332–6341.

[11] K. Kato, C. Matsunaga, T. Igarashi, H. Kim, A. Odaka, I. Shimada, Y. Arata, Complete assignment of the methionyl carbonyl carbon resonances in switch variant anti-dansyl antibodies labeled with [1-$^{13}$C]methionine, Biochemistry 30 (1991) 270–278.

[12] H.L. Crespi, R.M. Rosenberg, J.J. Katz, Proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains, Science 161 (1968) 795–796.

[13] J.L. Markley, I. Putter, O. Jardetzky, High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease, Science 161 (1968) 1249–1251.

[14] P. Brodin, T. Drakenberg, E. Thulin, S. Forsén, T. Grundstrom, Selective proton labelling of amino acids in deuterated bovine calbindin D9K. A way to simplify 1H-NMR spectra, Protein Engineering 2 (1989) 353–357.

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

[16] C. Hilty, C. Fernandez, G. Wider, K. Wüthrich, Side chain NMR assignments in the membrane protein OmpX reconstituted in DHPC micelles, Journal of Biomolecular NMR 23 (2002) 289–301.

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

[18] L.Y. Lian, D.A. Middleton, Labelling approaches for protein structural studies by solution-state and solid-state NMR, Progress in Nuclear Magnetic Resonance Spectroscopy 39 (2001) 171–190.

[19] T.W. Kirby, G.A. Mueller, E.F. DeRose, M.S. Lebetkin, G. Meiss, A. Pingoud, R.E. London, The nuclease A inhibitor represents a new variation of the rare PR-1 fold, Journal of Molecular Biology 320 (2002) 771–782.

[20] G. Meiss, U. Bottcher, C. Korn, O. Gimadutdinow, A. Pingoud, Vectors for dual expression of target genes in bacterial and mammalian cells, Biotechniques 29 (2000) 476–478.

[21] D.M. LeMaster, J.E. Cronan Jr., Biosynthetic production of $^{13}$C-labeled amino acids with site-specific enrichment, Journal of Biological Chemistry 257 (1982) 1224–1230.

[22] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, Journal of Biomolecular NMR 6 (1995) 277–293.
[23] B.A. Johnson, R.A. Blevins, NMRVIEW: a computer program for the visualization and analysis of NMR data, Journal of Biomolecular NMR 4 (1994) 603–614.
[24] S.M. Pascal, R.D. Muhandiram, T. Yamazaki, J.D. Forman-Kay, L.E. Kay, Simultaneous acquisition of N-15- and C-13-edited NOE spectra of proteins dissolved in $H_2O$, Journal of Magnetic Resonance, Series B 103 (1994) 197–201.
[25] M. Nilges, Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities, Journal of Molecular Biology 245 (1995) 645–660.
[26] S. Grzesiek, A. Bax, Improved 3D triple resonance techniques applied to a 31 kDa protein, Journal of Magnetic Resonance 96 (1992) 432–440.
[27] R. Pachter, C.H. Arrowsmith, O. Jardetzky, The effect of selective deuteration on magnetization transfer in larger proteins, Journal of Biomolecular NMR 2 (1992) 183–194.
[28] G.N. Cohen, The Regulation of Cell Metabolism, Holt, Rinehart, & Winston, NY, 1968.
[29] D.R. Muhandiram, L.E. Kay, Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity, Journal of Magnetic Resonance, Series B 103 (1994) 203–216.
[30] D.W. Yang, L.E. Kay, Improved (HN)-H-1-detected triple resonance TROSY-based experiments, Journal of Biomolecular NMR 13 (1999) 3–10.
[31] H. Kuboniwa, S. Grzesiek, F. Delaglio, A. Bax, Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods, Journal of Biomolecular NMR 4 (1994) 871–878.
[32] D.M. LeMaster, F.M. Richards, NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation, Biochemistry 27 (1988) 142–150.
[33] T. Yabuki, T. Kigawa, N. Dohmae, K. Takio, T. Terada, Y. Ito, E.D. Laue, J.A. Cooper, M. Kainosho, S. Yokoyama, Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis, Journal of Biomolecular NMR 11 (1998) 295–306.
[34] L.E. Kay, G.M. Clore, A. Bax, A.M. Gronenborn, Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution, Science 249 (1990) 411–414.
[35] G.M. Clore, L.E. Kay, A. Bax, A.M. Gronenborn, Four-dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin 1 beta, Biochemistry 30 (1991) 12–18.
[36] G.W. Vuister, G.M. Clore, A.M. Gronenborn, R. Powers, D.S. Garrett, R. Tschudin, A. Bax, Increased resolution and improved spectral quality, in four-dimensional $^{13}C/^{13}C$-seperated HMQC-NOESY-HMQC spectra using pulsed field gradients, Journal of Magnetic Resonance, Series B 101 (1993) 210–213.
[37] C. Zwahlen, P. Legault, S.J.F. Vincent, G.J.R. Konrat, L.E. Kay, Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex, Journal of the American Chemical Society 119 (1997) 6711–6721.
[38] M. Ikura, A. Bax, 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 (1992) 2433–2440.
[39] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, Journal of the American Chemical Society 125 (2003) 1385–1393.