Journal of Biomolecular NMR **28:** 235–247, 2004.
KLUWER/ESCOM

Protein signal assignments using specific labeling and cell-free synthesis

Jianxia Shi[a], Jeffrey G. Pelton[a], Ho S. Cho[b] & David E. Wemmer[a,c,*]
[a]*Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720;* [b]*Xencor, 111 West Lemon Avenue, Monrovia, CA 91016;* [c]*Department of Chemistry, University of California, Berkeley, CA 94720, U.S.A.*

Received 16 June 2003; Accepted 4 September 2003

Key words: cell-free protein synthesis, NMR assignment, phosphoserine phosphatase (PSP), selective stable isotope labeling

Abstract

The goal of structural genomics initiatives is to determine complete sets of protein structures that represent recently sequenced genomes. The development of new high throughput methods is an essential aspect of this enterprise. Residue type and sequential assignments obtained from specifically labeled samples, when combined with 3D heteronuclear data, can significantly increase the efficiency and accuracy of the assignment process, the first step in structure determination by NMR. A protocol for the design of specifically labeled samples with high information content is presented along with a description of the experiments used to extract essential information using 2D versions of 3D heteronuclear experiments. *In vitro* protein synthesis methods were used to produce four specifically labeled samples of the 23.5 kDa protein phosphoserine phosphatase (PSP) from *Methanoccous jannaschii* (MJ1594). Each sample contained two $^{13}C/^{15}N$-labeled amino acids and one ^{15}N-labeled amino acid. The 135 type and 14 sequential assignments obtained from these samples were used in conjunction with 3D data obtained from uniformly $^{13}C/^{15}N$-labeled and $^{2}H/^{13}C/^{15}N$-labeled protein to manually assign the backbone $^{1}H^{N}$, ^{15}N, ^{13}CO, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ signals. Using an automated assignment algorithm, 30% more assignments were obtained when the type and sequential assignments were used in the calculations.

Abbreviations: HSQC – heteronuclear single quantum coherence spectroscopy; NMR – nuclear magnetic resonance; NOE – nuclear Overhauser effect; PSP – phosphoserine phosphatase; TPPI – time-proportional phase incrementation.

Introduction

Possibilities for the assignment and structure determination of proteins and other biomolecules by NMR have been greatly expanded by use of both specific and uniform isotopic labeling of molecules with ^{13}C, ^{15}N and ^{2}H. Current methodologies rely on uniform ^{13}C and ^{15}N enrichment (optionally with 2H) and the application of a battery of 3D and 4D triple resonance experiments (Clore and Gronenborn, 1994; Gardner and Kay, 1998). These methods work well for proteins of up to about 40 kDa in size as long as the HSQC spectrum is well resolved. Assignment ambiguities arise when there are multiple sets of overlapping NH signals. In these cases it is often not possible to correlate specific $^{13}C^{\alpha}$, $^{13}C^{\beta}$, or ^{13}CO signals with their corresponding NH resonances. Resolution of the ambiguities usually requires collection and analysis of additional data sets, which is both labor intensive and time consuming.

Before 3D and 4D triple-resonance techniques were developed, one method of obtaining assignment informaton on proteins was to record 2D NMR spectra on samples that contained one or at most a few isotopically labeled amino acids. Classic examples include work on staphaloccocal nuclease (Torchia et al.,

*To whom correspondence should be addressed: E-mail: dewemmer@lbl.gov

1988; Torchia et al., 1989) and bacteriophage T4 lysozyme (McIntosh et al., 1990). The disadvantages of these schemes are that multiple samples (typically 1 L growths from *E. coli*), each requiring tens of milligrams of expensive labeled amino acids, are needed (Torchia et al., 1989), and scrambling of the ^{15}N nuclei for a number of amino acids, particularly for Glu, Asp, and Ser, complicates the assignment process (McIntosh and Dahlquist, 1990). For these reasons selective labeling has seen limited applications.

Over the past several years, extensive efforts have been made toward optimizing the production of proteins by *in vitro* or cell-free methods that employ *E. coli* or other cell extracts (Guignard et al., 2002; Hirao et al., 2002; Jermutus et al., 1998; Kim et al., 1996; Kim and Swartz, 1999, 2000). The advantages of these methods for synthesizing specifically labeled proteins for NMR studies have been conclusively demonstrated by Yokoyama and co-workers (Kigawa et al., 1995, 1999; Yabuki et al., 1998). Scrambling of the isotopic labels is potentially very low in cell-free protein preparations as evidenced by HSQC spectra of the Ras protein specifically labeled with either ^{15}N-Asp or ^{15}N-Ser (Kigawa et al., 1995). In addition, the required quantities of expensive labeled amino acids are reduced by a factor of 50 to 100 because the *in vitro* reactions are conducted in much smaller volumes (10 ml) compared to traditional 1 L growths and all of the amino acids are used to produce the protein of interest. These advantages were combined with the carbonyl-^{13}C/^{15}N labeling strategy to identify a unique Pro-Thr pair and Thr NH assignment in the protein Ras (Yabuki et al., 1998). An additional advantage with the cell-free system is that the small reaction volume and absence of many cellular proteins make it easier to prepare and purify multiple samples, enabling high throughput protein production.

As part of our structural genomics effort, we are investigating methods of increasing the efficiency of both the protein assignment and structure determination process. Recently, we successfully incorporated single and multiple ^{15}N-labeled amino acids into the 211 residue protein phosphoserine phosphatase (PSP) [EC 3.1.3.3] (MW 23.5 kDa) from *Methanoccous jannaschii* using the RTS 500 *in vitro* protein production system produced by Roche Molecular Biochemicals (Cho et al., 2001). This prompted us to consider methods of exploiting the advantages of cell-free methods to produce specifically labeled samples that would maximize the information available for the protein assignment process. Although the ideas of using selective labeling for assignment are not new, this approach has not been systematically applied. Herein, a protocol for the design of specifically labeled samples produced by *in vitro* methods is described that yields multiple type and sequential assignments from each sample. These data, combined with several conventional triple-resonance experiments recorded on uniformly labeled protein, provided almost complete backbone assignments for PSP despite significant overlap of NH signals. The advantage of using data obtained from the *in vitro* samples with automated assignment algorithms, such as AutoAssign (Moseley et al., 2001; Zimmerman et al., 1997), is also demonstrated.

Materials and methods

Expression and purification of PSP produced by in vivo *methods*

The construction of the gene for PSP from *Methanococcus jannaschii* (gene MJ1594), overexpression in *E. coli*, and purification were conducted as described previously (Wang et al., 2001). Uniformly ^{15}N- and ^{13}C/^{15}N-labeled samples were prepared from B834 (DE3)/pSIS1244 *E. coli* cells grown in M9 minimal medium supplemented with ^{15}N-ammonium chloride or both ^{15}N-ammonium chloride and $^{13}C_6$-glucose (Isotec, Inc., Miamisburg, OH), respectively. A highly deuterated (80%) and ^{13}C/^{15}N-labeled sample of PSP was obtained by growing the same cells in 90% D_2O. Samples were dissolved in 300 µl (Shigemi NMR tubes) or 500 µl of a solution containing 20 mM $MgCl_2$, 0.1 mM EDTA, 0.1% sodium azide, 2 mM tri-(2-carboxyethyl) phosphine (TCEP), and 10% D_2O at pH 6.2. The concentrations of the samples ranged from 1 mM to 2.5 mM.

Expression and purification of PSP produced by in vitro *methods*

The gene encoding PSP (gene MJ1594) was excised from a pET-21a construct and ligated into the pIVEX 2.3-MCS vector (Roche Molecular Biochemicals, Indianapolis, IN) between the Nde I and Bam HI restriction sites. Except for amino acids, the reaction mixture was prepared by reconstituting the lyophilized RTS 500 *E. coli* HY lysate with the solutions provided. The manufacturer suggested adding each amino acid to a final concentration of 0.5 mM. Some amino acids occur at much higher frequencies than others within the PSP sequence, so we adopted a different strategy

for amino acid addition. On average, a given type of amino acid occurs about 10 times in the PSP sequence (211 amino acids/18 amino acid types). An amino acid that occurs at this average frequency (10) was added to a final concentration of 0.5 mM. The concentrations of the other amino acids were then calculated based on this 0.5 mM/10 residue ratio. For instance, Gly occurs 15 times and was added at a concentration of 0.75 mM, while Tyr occurs only 3 times and was added at a concentration of 0.15 mM. With the exception of His and Trp, which do not occur in the sequence and were not added to the reaction, no amino acid was added at a concentration of less than 0.1 mM.

PSP was synthesized using 15 μg of plasmid per 1 ml of reconstituted lysate. Some preparations of the lysate reconstitution buffer provided by Roche Molecular Biochemicals (Indianapolis, IN) included the stabilizer polyethylene glycol (PEG). There was no discernable difference in yield for reactions carried out with and without PEG. The standard lysate provided in the RTS 500 kit was used for all samples except KA and DSF. For KA the inhibitor amino-oxy-acetate was added at a concentration of 2 mM to reduce dilution of Ala (Lopukhov et al., 2002). For the DSF sample, a special lysate formulation provide by Roche Molecular Biochemicals, which contains an inhibitor effective in reducing the scrambling of Asp and Ser, was used. While preparing the DSF amino acid mixture, the ^{15}N-Ser stock solution was inadvertently contaminated with ^{15}N-Cys, and the signals for these two residues also appear in the DSF HSQC spectrum. After 24 hours of synthesis, the RTS reaction mixture was diluted to twice the starting volume with 25 mM Tris buffer (pH 8.4). The ETC sample was partially purified by incubation at 40 °C, centrifugation, and buffer exchange. This degree of purification proved sufficient for collection of high quality NMR data. All other samples were purified in a multi-step procedure. First, unwanted proteins were precipitated by incubation at 70 °C for 30 min and removed by centrifugation. Additional protein impurities were removed by passing the resulting solution over a 1 ml Hightrap Q column (Pharmacia, NJ) equilibrated with 25 mM Tris at pH 8.4. PSP does not stick to the column, and was eluted with the same buffer. In these cases, the PEG was removed by adjusting the pH of the protein solution to 3.0 and passing it over an SP Sepharose column (Pharmacia, NJ) equilibrated with 25 mM Tris-HCl (pH 3.0) and 50 mM NaCl. PSP was eluted with a buffer containing 25 mM Tris-HCl (pH 6.5) and 500 mM NaCl.

The ETC, DSF, KAC, KA, 100% TLY, and fractionally labeled TLY (100% Thr, 33% Leu, and 66% Tyr respectively) samples were produced from 1, 1, 3, 1, 3, and 1 ml reactions, respectively. In most cases, approximately 1.25 milligrams of purified protein were obtained from each milliliter of the RTS 500 *E. coli* lysate. Protein was dialyzed against the same buffer as used for protein produced by *in vivo* method, and concentrated. The final sample concentrations were 150 μM (ETC), 100 μM (DSF), 400 μM (KAC), 100 μM (KA), 400 μM (100% TLY) and 100 μM (fractionally labeled TLY).

NMR Spectroscopy

NMR spectra were recorded on four-channel DRX 500 and 600 NMR spectrometers equipped with ^{1}H/^{13}C/^{15}N gradient probes at 40 °C. Some spectra were also recorded on the DRX 500 equipped with a cryoprobe accessory. ^{15}N and ^{13}C decoupling were achieved by WALTZ16 (Shaka et al., 1983) modulation of a 1.6 kHz rf field and GARP (Shaka et al., 1985) modulation of a 3.8 kHz rf field, respectively. ^{1}H chemical shifts are referenced indirectly to DSS by setting the water resonance to 4.69 ppm at 40 °C. ^{15}N and ^{13}C chemical shifts were also referenced indirectly to sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as described (Wishart et al., 1995). NMR spectra were processed using the NMRPipe suite of programs (Delaglio et al., 1995) and analyzed using the NMRView package (Johnson and Blevins, 1994).

NMR spectroscopy – in vitro *samples*

2D ^{1}H-^{15}N HSQC spectra were recorded in 12 to 17 hours using the FHSQC pulse sequence (Mori et al., 1995). Typically, 128 complex points were acquired in the ^{15}N dimension of each experiment. 2D ^{1}H-^{15}N planes of HNCA spectra were recorded in 24 h on KAC (600 MHz) and TLY (500 MHz with cryoprobe accessory) using 38 complex points in the ^{15}N dimension. 2D ^{1}H-^{15}N planes of HNCO spectra were recorded in 12 hours at 600 MHz on the KAC sample using 32 complex points in the ^{15}N dimension. The 2D ^{1}H-^{13}C planes of the HNCACB and CBCA(CO)NH spectra were recorded in 1.5 days on the KAC sample at 500 MHz using a standard probe. A total of 52 complex points were collected in the ^{13}C dimension. Reduced dimensionality forms of the HNCACB and CBCA(CO)NH experiments were recorded on the KAC sample at 600 MHz. The pulse

sequences were derived from the 3D versions by combining the ^{15}N and ^{13}C time increments as described (Szyperski et al., 2002). The experiments were acquired with the ^{15}N carrier set to the edge of the spectrum (100 ppm). ^{15}N chemical shifts were derived by measuring the difference between the two resulting ^{13}C multiplets and adding the result to the ^{15}N carrier frequency.

NMR spectroscopy – in vivo *samples*

3D HN(CA)NH (Ikegami et al., 1997), HNCA (Grzesiek and Bax, 1992; Yamazaki et al., 1994), HNCO (Grzesiek and Bax, 1992; Yamazaki et al., 1994), HN(CO)CA (Clubb et al., 1992; Kay et al., 1994), HN(CA)CO (Clubb et al., 1992; Kay et al., 1994) and HNCACB (Muhandiram and Kay, 1994; Wittekind and Mueller, 1993) spectra were acquired on deuterated and ^{13}C/^{15}N-labeled PSP, while 3D HCANH (Larsson et al., 1999) and HBHA(CO)NH (Grzesiek and Bax, 1993) spectra were acquired on ^{13}C/^{15}N-labeled PSP at 600 MHz using minor modifications of pulse programs supplied by Bruker Bio-Spin (XWINNMR 3.0). These programs employ echo-antiecho gradient selection and sensitivity enhancement (Muhandiram and Kay, 1994) in the ^{15}N dimension and the States-TPPI method (Marion et al., 1989) of quadrature detection in the ^{13}C dimension. Typically, 25, 58, and 64 complex points were collected in the ^{15}N, ^{13}C, and ^{1}H dimensions, respectively with 32 scans per complex point and a recycle delay of two seconds for deuterated samples and one second for ^{13}C/^{15}N-labeled samples. The HN(CA)NH experiment was acquired twice, once by recording H_i^N-N_i-N_{i+1}/N_{i-1}/N_i correlations and once by recording H_i^N-N_i-H_{i+1}^N/ H_{i-1}^N/H_i^N correlations. The former version was especially useful because the narrower linewidths of the ^{15}N resonances (compared to $^{1}H^N$ signals) made it easier to resolve overlapping signals.

Automated assignments

Peak lists were derived from ^{1}H-^{15}N HSQC, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HCANH, and HBHA(CO)NH spectra. Each spectrum was zero-filled once in the indirect carbon or proton dimension and twice in the ^{15}N dimension (128 real points) prior to Fourier transformation. Peak shifts and intensities were determined using the program PIPP (Garrett et al., 1991). PIPP fits peaks with ellipses and uses interpolation to obtain peak centers. Inter-residue HN_i-H_{i-1}^{α} peak lists were obtained by peak picking signals at 3.0 ppm and above in HBHA(CO)NH spectra. The peak lists were aligned to one another using registration and shifting algorithms provided in the AutoPeak software package (version 1.3). Model data were derived by adding an uncertainty equal to one-half of the digital resolution to each of the manually determined chemical shifts. Uncertainties added to shift positions were as follows: HN in HSQC spectra (0.005), N in HSQC spectra (0.015), HN in 3D spectra (0.015), N in 3D spectra (0.15), CA (0.15), CB (0.25), HA (0.01), and CO (0.15). Amino acid type and sequential assignment information were derived from ETC, DSF, KAC, and TLY samples. In total, these data yielded 135 type assignments and 14 sequential assignment possibilities (Table 2). The assignments were derived from a combination of default and refined modes of execution within the program AutoAssign (version 1.11). Assignments were categorized as correct if the HN and ^{15}N chemical shifts were within 0.08 ppm and 0.5 ppm of the manually determined values. Tolerances used for calculations with both real and model peak lists were as follows: HN (0.035 ppm), N (0.35), CA (0.85), CB (0.95), HA (0.07), and CO (0.25).

Results and discussion

Sample design

Our strategy for the design of specifically labeled samples with high information content involves the selection of three residues, one each from groups 1 and 2 of Table 1 for ^{13}C/^{15}N-labeling, and a third residue from group 3 for ^{15}N-labeling. Selection of the first two amino acids from groups 1 and 2 ensures that these residues can be distinguished from one another based on $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts. Ideally, these residues will occur in pairs one or more times in the sequence, allowing for the extraction of sequential assignment information (see below). To simplify the assignment process, the amino acid chosen for ^{15}N-labeling should not immediately follow either of the ^{13}C/^{15}N-labeled residues. The amino acids listed in group 3 (Table 1) are good candidates for ^{15}N-labeling, since they generally occur at lower frequencies than those in groups 1 and 2 (McCaldon and Argos, 1988). Using this scheme, protein samples are identified using three letters, comprised of the one-letter code for the first two ^{13}C/^{15}N-labeled amino acids, followed by that for the ^{15}N-labeled residue.

Table 1. Sets of amino acids recommended for specific labeling (ppm)[a].

Group 1 residue	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	Group 2 residue	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	Group 3 residue	$^{13}C^{\alpha}$	$^{13}C^{\beta}$
Gln	56.6	29.1	Thr	62.2	69.6	Trp	57.7	30.2
Glu	57.4	30.1	Ser	58.6	63.8	His	56.4	30.0
Arg	57.0	30.7	Ile	61.6	38.6	Met	56.2	32.9
Lys	56.8	32.8	Val	62.5	32.7	Cys	57.4	34.1
Asn	53.4	38.7	Ala	53.2	18.9	Tyr	58.0	39.2
Asp	54.5	40.7				Phe	58.3	40.0
Leu	55.6	42.4				Gly	45.3	

[a]Average chemical shifts are taken from the BioMagResBank. (http://www.bmrb.wisc.edu).

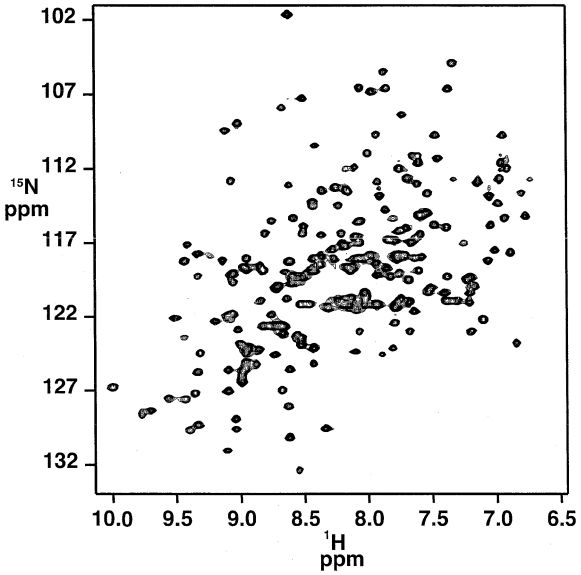

Figure 1. ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled PSP derived from *in vivo* methods recorded at 600 MHz and 40 °C.

The ^{1}H-^{15}N HSQC spectrum of uniformly ^{15}N-labeled PSP from *Methanococcus jannaschii* produced *in vivo* is shown in Figure 1. Overall, most signals are well dispersed, although there is significant overlap in some regions of the spectrum. Excluding sidechains, a total of 183 of 206 expected signals were readily identified.

Four samples containing two $^{13}C/^{15}N$-labeled amino acids and one ^{15}N-labeled amino acid were prepared (Table 2). Representative ^{1}H-^{15}N HSQC spectra for ETC, DSF, and KAC are shown in Figure 2. All of the expected signals were observed in each HSQC spectrum. Indeed, the spectrum of ETC (Figure 2a) is of high quality even though the sample was purified only by incubation at 40 °C for 6 h, centrifugation, and buffer exchange. As was shown for the cis-trans isomerase PpiB (Guignard et al., 2002), these data indicate that in favorable cases, spectra of proteins produced by *in vitro* synthesis can be obtained after minimal purification.

Attempts to specifically label Glu, Ser, and Asp *in vivo* usually result in at least some scrambling due to transamination (McIntosh and Dahlquist, 1990). Although there was no evidence for scrambling of Glu (ETC sample, Figure 2a), there was evidence for dilution or scrambling of Asp, Ser, and Ala. Early attempts to label PSP with ^{15}N-Asp and ^{15}N-Ser resulted in the appearance of weak signals due to Asn and Gly (data not shown). Similarly, signals resulting from Ala were reduced in intensity (dilution) by a factor of two in the KAC sample (Figure 2c). Using a lysate reconstitution buffer developed by Roche Molecular Biosciences, it was possible to label Asp and Ser (DSF sample; Figure 2b) without scrambling. However, the weaker than expected signals for Asp in the DSF sample (Figure 2b), indicate that this residue is still being diluted by unknown processes. A ready pathway for dilution of Ala involves pyruvate, a by-product of the energy source phosphoenolpyruvate. Pyruvate is converted into Ala by a pyridoxal-dependent transaminase which can be inhibited by amino-oxy-acetate (Lopukhov et al., 2002). Figure 2d shows the result of synthesizing PSP labeled with Lys and Ala, and with the addition of amino-oxy-acetate at a concentration of 2 mM. In the new preparation, the intensities of the Ala and Lys signals are essentially equal showing that dilution of the Ala signals has been inhibited.

Type assignments

One method for obtaining type assignments is to vary the degree of incorporation of two or three labeled

Table 2. PSP samples produced by *in vitro* synthesis and data used in Autoassign

Sample[a]	Amino acid frequency	Amino acid pairs identified	NMR experiments
ETC	27 Glu, 8 Thr, 2 Cys		HSQC
DSF	13 Asp, 5 Ser, 7 Phe		HSQC
KAC	33 Lys, 17 Ala, 2 Cys	3 Lys-Ala	HSQC, HNCA
		3 Ala-Lys	HNCO, CBCA(CO)NH
		5 Lys-Lys	HNCACB
TLY	8 Thr, 20 Leu, 3 Tyr	1 Thr-Leu	HSQC, HNCA
		1 Leu-Thr	
		1 Leu-Leu	

[a]Underlined residues were $^{13}C/^{15}N$–labeled. Other listed amino acids were ^{15}N-labeled.

amino acids. This technique, which is based on early specific labeling work of Ikura et al. (1990), is demonstrated with the sample TLY in Figures 3a and 3b. Two samples were produced. In the first sample (Figure 3a) only the labeled forms of Thr, Leu, and Tyr ($^{13}C/^{15}N$-labeled Thr, $^{13}C/^{15}N$-labeled Leu, and ^{15}N-labeled Tyr) were added to the reaction mixture. In the second sample (Figure 3b), the amino acids were added in their labeled and unlabeled forms in ratios of 1, 0.33, and 0.66, respectively. As shown in Figure 3c, the signals can be easily classified according to type based on the ratio of peak intensities in the two spectra. A similar plot is obtained using the ^{1}H-^{15}N HSQC spectrum of the uniformly-labeled protein (Figure 1) as the reference. Together the data in Figure 3 demonstrate that peak intensities reflect levels of incorporation in the *in vitro* system, permitting the rapid and accurate type assignments.

Identification of ^{15}N-labeled amino acids

The set of ^{15}N-labeled amino acids can be distinguished from the set of doubly-labeled amino acids by comparison of a ^{1}H-^{15}N HSQC spectrum with a 2D ^{1}H-^{15}N plane of an HNCA experiment. As shown in Figure 4 for TLY, the signals from the three ^{15}N-labeled Tyr residues appear in the HSQC spectrum but not in the HNCA (Figure 4a). Similarly for KAC, the two ^{15}N-labeled Cys residues appear only in the ^{1}H-^{15}N HSQC spectrum (Figure 4b). Complications can arise if any of the signals from the $^{13}C/^{15}N$-labeled residues are weak in the HNCA. This was the case for Thr 21 in TLY and for Ala26 and Ala169 in KAC. Each gave a weak signal in the HSQC spectrum and no signal (Thr 21) or very weak signal (Ala26, Ala69) in the HNCA. The reduced levels of scrambling observed using the *in vitro* system is especially valuable in these cases because a quick tally of the expected and observed signals will quickly reveal potential ambiguities.

Identification of $^{13}C/^{15}N$-labeled amino acids and sequential neighbors

The two doubly-labeled residues can be distinguished from one another through comparison of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts (Table 2) observed in a 2D ^{1}H-^{13}C plane of an HNCACB experiment. As shown in Figure 5 for KAC, many of the HN signals from the 33 Lys and 17 Ala residues can be distinguished based on differences in $^{13}C^{\beta}$ shifts (Lys $^{13}C^{\beta}$ $\sim$33 ppm; Ala $^{13}C^{\beta}\sim$19 ppm). Other HN signals were difficult to classify due to overlap in the 2D spectrum. This problem can be avoided by selecting combinations of $^{13}C/^{15}N$-labeled residues (groups 1 and 2 in Table 1) that minimizes the total number of signals.

Doubly-labeled amino acids that follow another doubly-labeled amino acid can be identified in 2D versions of HNCO (Figure 6) and CBCA(CO)NH spectra (Figure 7). For example, each of the 3 Lys-Ala, 3 Ala-Lys and 5 Lys-Lys pairs in the KAC sample was identified in the 2D ^{1}H-^{15}N plane of the HNCO experiment (Figure 6). In addition, the three sets of pairs (Lys-Ala, Ala-Lys, and Lys-Lys) were distinguished from one another using the 2D ^{1}H-^{13}C plane of the CBCA(CO)NH spectrum (Figure 7) along with type assignments obtained from the HNCACB spectrum (Figure 5) and intensity differences between Lys and Ala recorded on a sample prepared without a transaminase inhibitor (Figure 2c). The three Ala-Lys pairs (AK144, AK153, and AK187) were identified by the three characteristic Ala $^{13}C^{\beta}$ signals between 17 and

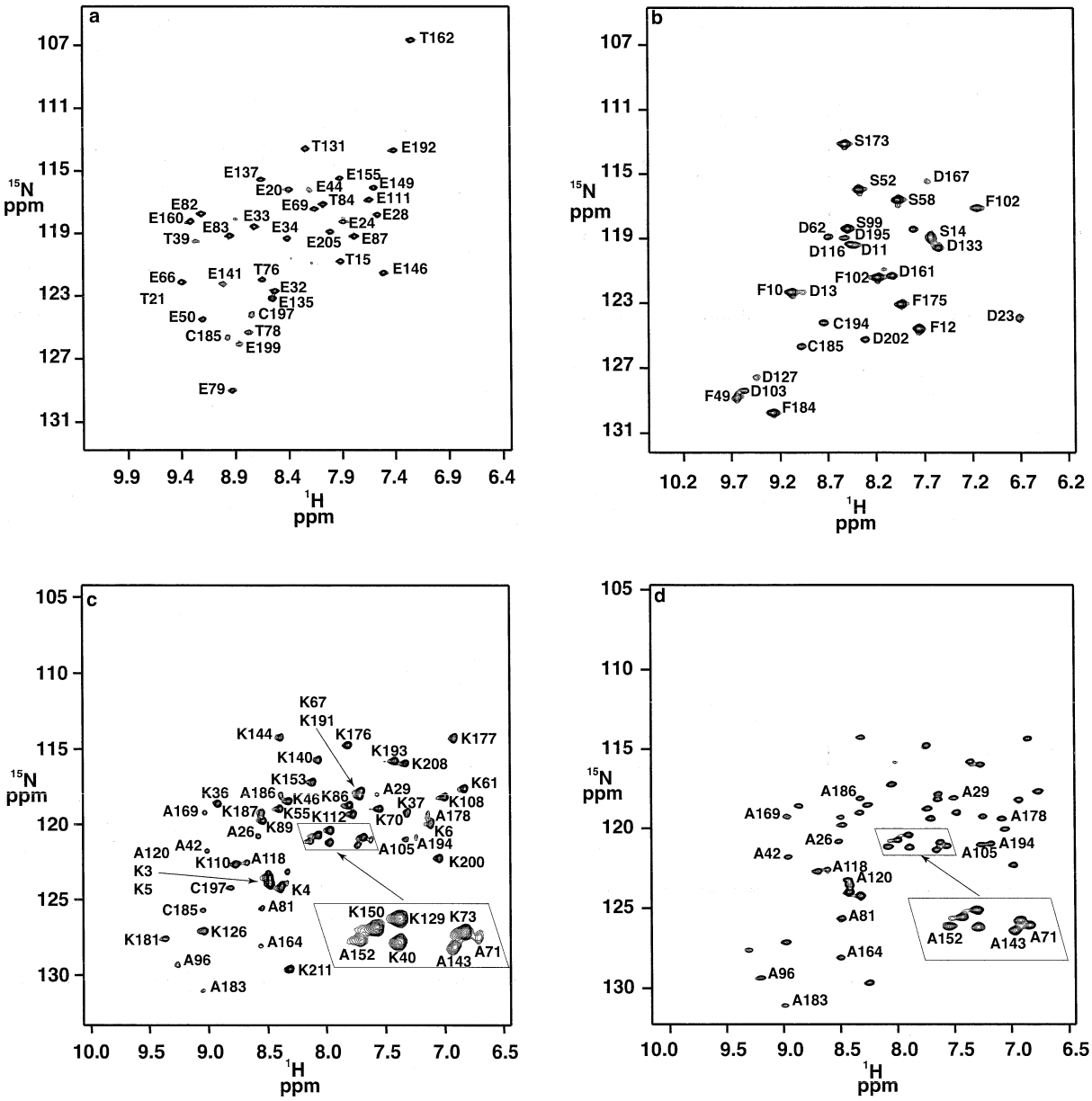

Figure 2. ^{1}H-^{15}N HSQC spectra of ETC (a), DSF (b), KAC (c) and KA (d) samples produced by *in vitro* methods. The spectra were recorded on a DRX 600 MHz spectrometer at 40 °C. During preparation of the DSF sample, the Ser stock solution was contaminated with ^{15}N-labeled Cys, and the signals for the two Cys residues in PSP appear in the spectrum. For ETC, DSF, and KAC, the first two residues were ^{13}C and ^{15}N-labeled, and the third residue was ^{15}N-labeled. For KA both residues were ^{15}N-labeled.

26 ppm (Figure 7) which result from C^{β}_{i-1}-N_i-H^N_i correlations. The Lys-Ala and Lys-Lys pairs were then distinguished from one another based on the previous peak classifications for the C-terminal residue of the pair (either Lys for Lys-Lys or Ala for Lys-Ala). Because of its high sensitivity, the CBCA(CO)NH experiment was also successfully recorded in reduced-dimensionality form (Szyperski et al., 2002) (not shown), providing the ^{15}N shifts of one of the three Ala (Ala-137) of the three Lys-Ala pairs and six Lys of the 3 Ala-Lys and 5 Lys-Lys pairs. Together with the sequence, these data reduce the number of possible sequential assignments for these signals to only 3 each for Lys-Ala and Ala-Lys, and 5 for Lys-Lys.

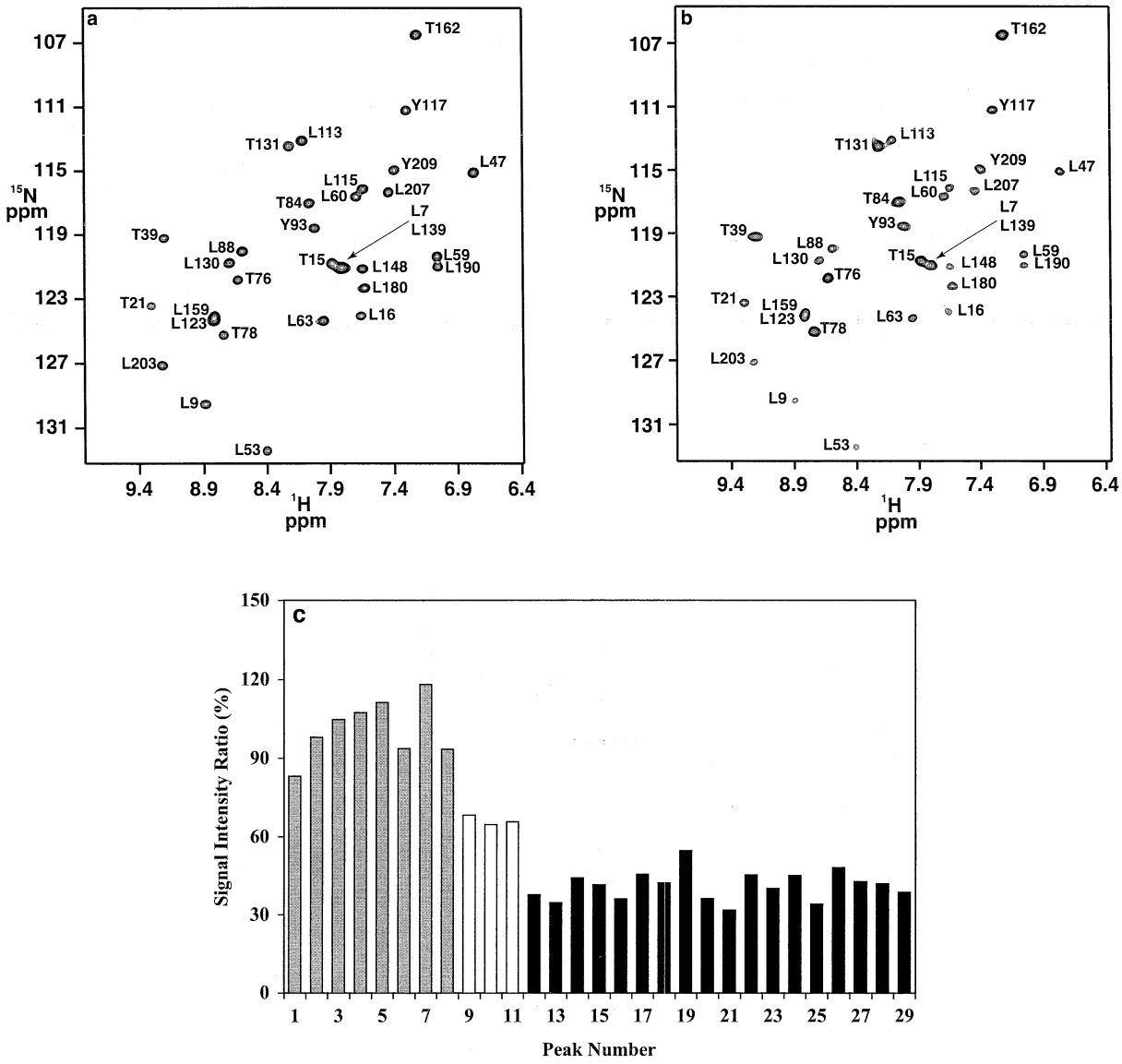

Figure 3. ^{1}H-^{15}N HSQC spectra of TLY with 100% isotope incorporation (a), with isotope-labeled Thr, Leu, and Tyr incorporated at 100%, 33%, and 66%, respectively (b), and a graphical summary of peak intensity ratios (c). Thr and Leu were labeled with ^{13}C and ^{15}N, and Tyr was labeled with ^{15}N. Average intensity ratios (in percent) were 101 (Thr), 37 (Leu), and 66 (Tyr). In (c), ratios for Thr, Leu, and Tyr are denoted by gray, black, and white bars, respectively. Two sets of peaks (Leu7 and Leu139; Leu123 and Leu159) are overlapped. In each case, one bar is used to represent the intensity ratio. The HSQC spectra were recorded in 0.5 days on a DRX 500 MHz spectrometer equipped with a cryoprobe.

Identification of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts was also useful in the analysis of 3D heteronuclear data recorded on the uniformly-labeled sample. For example, five amide signals near 7.1 ppm (^{1}H) and between 119 and 121 ppm (^{15}N) in the HSQC spectrum of uniformly ^{15}N-labeled PSP (Figure 1) were poorly resolved in 3D heteronuclear spectra due to limited resolution in the ^{15}N dimension. Two of these signals (due to Lys6 and Ala178) also appeared in the HSQC (Figure 2c) and HNCACB (Figure 5) spectra of the KAC sample, from which type assignments and both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts were obtained. The additional data from the KAC sample were sufficient to resolve the overlap in the 3D HNCA and 3D HNCACB spectra of the uniformly-labeled protein, making it possible to obtain assignments for all five signals.

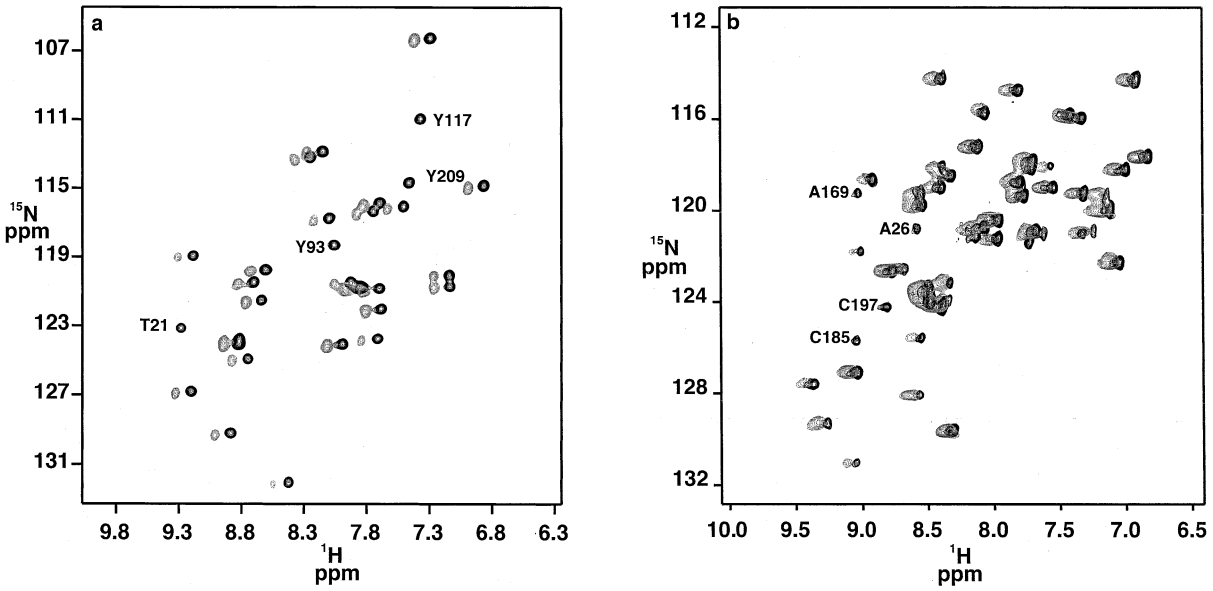

Figure 4. Overlays of a ^{1}H-^{15}N HSQC (black) and a ^{1}H-^{15}N plane (gray) of an HNCA spectrum of TLY (a) and KAC (b) at 600 MHz. The three ^{15}N-labeled Tyr residues in (a), and the two ^{15}N-labeled Cys residues in (b), which appear in the HSQC spectra but not in the HNCA spectra, are labeled. The spectra are shifted slightly relative to one another in the ^{1}H dimension for clarity. Weak peaks for Thr21 in (a) and Ala 26 and Ala 169 in (b) are also labeled. The signals for these residues are weak in the respective HNCA spectra, but are present at lower contours.

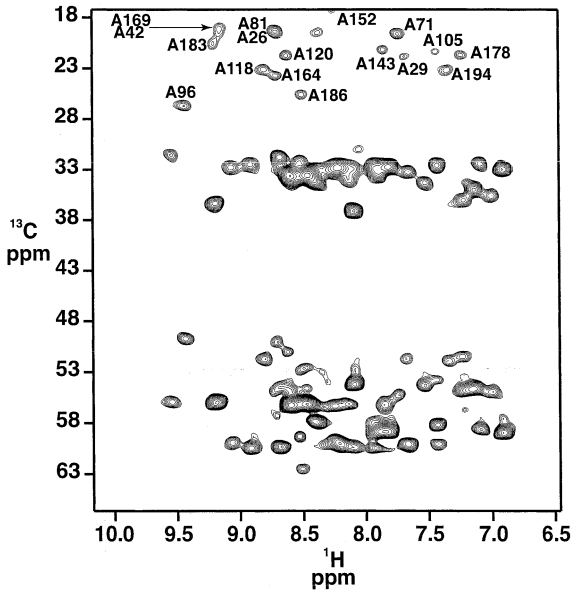

Figure 5. ^{1}H-^{13}C plane of an HNCACB spectrum of KAC recorded at 500 MHz. Lys and Ala $^{13}C^{\alpha}$ signals resonate between 48 and 63 ppm, while Lys $^{13}C^{\beta}$ signals resonate between 32 and 38 ppm and Ala $^{13}C^{\beta}$ signals (labeled) resonate between 18 and 28 ppm. Although not shown, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ signals are opposite in intensity.

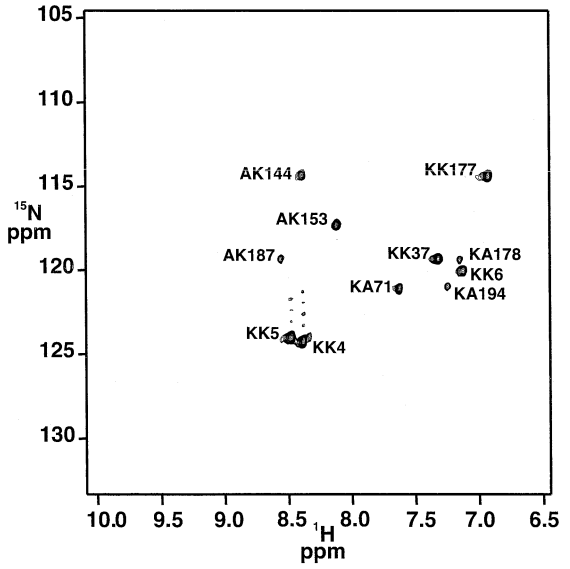

Figure 6. ^{1}H-^{15}N plane of an HNCO spectrum recorded on KAC at 600 MHz. NH signals appear for the ^{13}C/^{15}N-labeled amino acid that immediately follows another ^{13}C/^{15}N-labeled residue in the sequence. The sample was designed so that no ^{15}N-labeled residues follow a ^{13}C/^{15}N-labeled residue. Labels denote the pair of amino acids that give rise to each peak.

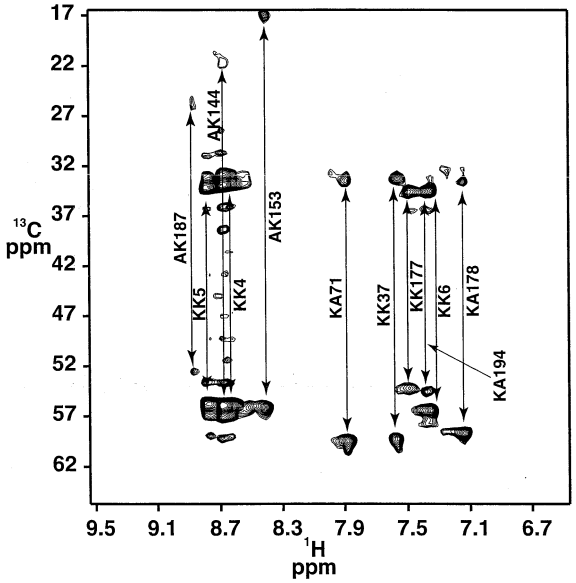

Figure 7. ^{1}H-^{13}C plane of a CBCA(CO)NH spectrum recorded on KAC at 500 MHz.

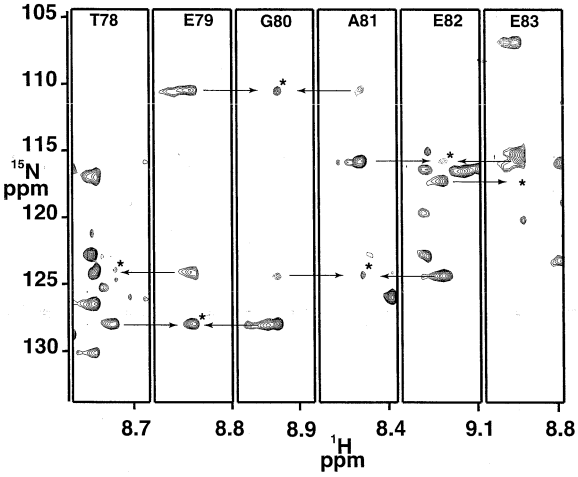

Figure 8. Representative strips from a 3D HN(CA)HN spectrum recorded in H_i^N-N_i-N_{i+1}/N_{i-1}/N_i mode on deuterated and uniformly ^{13}C/^{15}N-labeled PSP at 600 MHz. Asterisks denote $^{1}H^{N}$-^{15}N intra-residue correlations. Arrows show inter-residue connections.

Manual resonance assignments

Concomitant with *in vitro* protein production and experiments, 3D heteronuclear data were recorded on uniformly ^{15}N-, ^{13}C/^{15}N-, and ^{2}H/^{13}C/^{15}N-labeled PSP samples. The HN(CA)NH experiment was particularly valuable in the assignment process because the cross peaks link NH signals on adjacent residues in much the same way that d_{NN} NOEs do in helical regions of proteins (Ikegami et al., 1997). The experiment was performed twice, once by recording H_i^N-N_i-N_{i+1}/N_{i-1}/N_i (Figure 8) correlations and once by recording H_i^N-N_i-H_{i+1}^N/H_{i-1}^N/H_i^N correlations. In the first pass, the sequential assignments were obtained by combining these results with $^{13}C_i^{\alpha}$/$^{13}C_{i-1}^{\alpha}$ correlations provided by HNCA (Grzesiek and Bax, 1992; Yamazaki et al., 1994) and HN(CO)CA (Clubb et al., 1992; Kay et al., 1994) experiments and the 135 HN type assignments and 14 sequential assignment possibilities (Table 2) provided by experiments on the ETC, DSF, KAC, and TLY samples. Additional assignments were then made by considering $^{13}C_i^{\beta}$/$^{13}C_{i-1}^{\beta}$ linkages obtained from HNCACB (Muhandiram and Kay, 1994; Wittekind and Mueller, 1993) and CBCA(CO)NH (Grzesiek and Bax, 1993; Muhandiram and Kay, 1994) experiments and H_i^{α}/H_{i-1}^{α} linkages obtained from HCANH (Larsson et al., 1999) and HBHA(CO)NH (Grzesiek and Bax, 1993) experiments. In total, the backbone signals for all residues ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$ and ^{13}CO) except Met1 and Glu2 have been assigned along with 95% of the $^{13}C^{\beta}$ signals. The assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number BMRB-5799.

Automated assignments

The impact of the residue type and sequential assignment information on automated assignment algorithms was investigated using the program AutoAssign (Zimmerman et al., 1997). The results, which were obtained by combining 135 residue type assignments and 14 restraints on sequential assignments from the specifically-labeled samples (Table 2) with peak lists from nine 2D and 3D spectra (see Materials and methods), are summarized in Figure 9. As a reference, model peak lists were created from the manual assignments and also used as input to AutoAssign. To approximate real data, an uncertainty equal to one-half the digital resolution was added to each manually determined shift before being added to the various model peak lists. As can be seen in the figure, the total number of correct assignments increased by about 30% when the sequential and type assignments were combined with both the real and model peak lists. For the real data, the total number of incorrect assignments remained about the same (14%) with and without addition of the *in vitro* data. In contrast, the number of incorrect assignments decreased dramatically (15%) when the type and sequen-

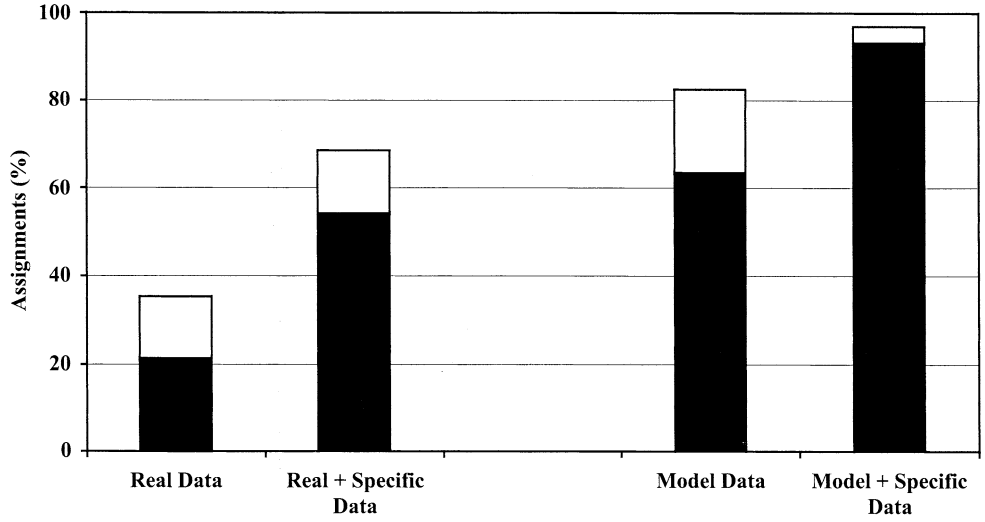

Figure 9. Percentage of ^{1}H-^{15}N assignments made with and without the 135 type and 14 sequential assignment possibilities obtained from the four *in vitro* samples (Table 2). Black and white bars denote the percentage of correct and incorrect assignments, respectively, compared to the manually determined shifts. The calculations were performed using real peak lists (left) and model peak lists generated from the manual determined assignments (right).

tial assignments were combined with the model peak lists.

Not surprisingly, AutoAssign was able to assign more residues correctly with model data (93%) than with real data (54%). At 23.5 kDa, PSP is a challenging case for automated assignment. Manual PSP assignments were obtained with additional data, particularly the HN(CA)NH spectrum, that was not easily adapted to AutoAssign. Overlap of ^{1}H-^{15}N signals also has some impact on the number of correct assignments. Small shifts of individual resonances from spectrum to spectrum, which occur in the real data, but not in the ideal data, probably also play a role. Limitations aside, these data provide a quantitative measure of the improvement possible when type and sequential assignments obtained with the triply-labeled samples are combined with 3D spectra to make protein assignments.

Conclusions

The triple labeling strategy presented here, in combination with cell-free protein synthesis, provides an efficient method for obtaining multiple type and sequential assignments that can be used to accelerate the protein assignment process. In this work a commercially available cell-free synthesis kit produced by Roche Molecular Biosciences was used to produce all of the protein samples. This kit is based on an *E. coli* extract similar to that reported by Kigawa et al. (1999). Based on the specific-labeling results of the Yokoyama group (Kigawa et al., 1995, 1999; Yabuki et al., 1998), and Guignard et al. (2002), we expect that the scrambling issues and labeling results presented here will be similar, although not necessarily identical, for proteins produced by other *E. coli*-based cell-free systems and that similar strategies can be used to suppress this. The advantages and disadvantages of using other extracts to produce samples for NMR, such as wheat germ (Madin et al., 2000; Sawasaki et al., 2002) remains to be explored in the context of this type of selective labeling.

Some scrambling and dilution processes do occur with the *E. coli* extract produced by Roche Molecular Biosciences, and were most apparent with Asp, Ser, and Ala. In contrast, there was no evidence for scrambling of Glu, even though it is known to do so *in vivo*. A modified lysate developed by Roche Molecular Biosciences effectively inhibited scrambling of Asp and Ser, and the transaminase inhibitor aminooxy-acetate inhibited the scrambling of Ala. Dilution of Asp in cell-free systems must still be addressed, but does not present a substantial concern for our overall approach.

The intensities of HN signals in HSQC spectra correlate with the fraction of labeled and unlabeled amino acid added to the reaction mixture, and can

be used to classify signals according to type. Comparison of HSQC spectra with 2D planes of HNCA and HNCACB spectra also allows for the classification of ^{15}N and $^{13}C/^{15}N$-labeled amino acids according to type, while comparison of HSQC spectra with 2D planes of HNCO and CBCA(CO)NH spectra allows for the identification of sequential assignments. Side-chain $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ signals obtained from HNCACB and CBCA(CO)NH spectra are valuable in sorting out crowded regions in 3D spectra of uniformly-labeled proteins. Additionally, the data obtained from this labeling strategy can be used to increase the number and accuracy of assignments obtained by automated routines such as AutoAssign.

In the future, continuing improvement in yields and ease of use should make cell-free protein synthesis an attractive option for producing both specifically- and uniformly-labeled proteins for NMR studies. The Japanese structural genomics effort is already using cell-free synthesis for uniform labeling of target proteins on a large scale (Yokoyama et al., 2000). While there has been remarkable progress in assignment of large proteins (Tugarinov et al., 2002), we believe that selective labeling of the type described here should be of value in many cases.

Acknowledgements

This work was supported by the NIH through GM62412 to the Berkeley Structural Genomic Center. We thank Roche Molecular Biochemicals, Indianapolis, IN for providing the lysate used for this work, Dr Seth Rubin and Modi Wetzler for a critical review of the manuscript, Steven Damo for setting up AutoAssign program, and Rosie Kim and Hisao Yokota for producing the initial PSP clone.

References

Cho H.S., Pelton, J.G., Wang, W., Yokota, H. and Wemmer, D.E. (2001) *Biochemica*, 27–29.

Clore, G.M. and Gronenborn, A.M. (1994) *Meth. Enzymol.*, **239**, 349–363.

Clubb, R.T., Thanabal, V. and Wagner, G. (1992) *J. Magn. Reson.*, **97**, 213–217.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Gardner, K.H., and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. and Dixon, N.E. (2002) *FEBS Lett.*, **524**, 159–162.

Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T. et al. (2002) *Nat. Biotechnol.*, **20**, 177–182.

Ikegami, T., Sato, S., Walchli, M., Kyogoku, Y. and Shirakawa, M. (1997) *J. Magn. Reson.*, **124**, 214–217.

Ikura, M., Krinks, M., Torchia, D.A. and Bax, A. (1990) *FEBS Lett.*, **266**, 155–158.

Jermutus, L., Ryabova, L.A. and Pluckthun, A. (1998) *Curr. Opin. Biotechnol.*, **9**, 534–548.

Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603–614.

Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) *J. Magn. Reson.*, **109A**, 129–133.

Kigawa, T., Muto, Y. and Yokoyama, S. (1995) *J. Biomol. NMR*, **6**, 129–134.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) *FEBS Lett.*, **442**, 15–19.

Kim, D.M. and Swartz, J.R. (1999) *Biotechnol. Bioeng.*, **66**, 180–188.

Kim, D.M. and Swartz, J.R. (2000) *Biotechnol. Prog.*, **16**, 385–390.

Kim, D.M., Kigawa, T., Choi, C.Y. and Yokoyama, S. (1996) *Eur. J. Biochem.*, **239**, 881–886.

Larsson, G., Wijmenga, S.S. and Schleucher, J. (1999) *J. Biomol. NMR*, **14**, 169–174.

Lopukhov, L.V., Ponomareva, A.A. and Yagodina, L.O. (2002) *Biotechniques*, **32**, 1248–1250.

Madin, K., Sawasaki, T., Ogasawara, T. and Endo, Y. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 559–564

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

McCaldon P, Argos P (1988) *Proteins*, **4**, 99–122.

McIntosh, L. P. and Dahlquist, F. W. (1990) *Q. Rev. Biophys.*, **23**, 1–38.

McIntosh, L.P., Wand, A.J., Lowry, D.F., Redfield, A.G. and Dahlquist, F.W. (1990) *Biochemistry*, **29**, 6341–6362.

Mori, S., Abeygunawardana, C., Johnson, M.O. and van Zijl, P.C. (1995) *J. Magn. Reson.*, **108B**, 94–98.

Moseley H.N., Monleon D and Montelione G.T. (2001) *Meth. Enzymol.*, **339**, 91–108.

Muhandiram, D.R. and Kay, L. E. (1994) *J. Magn. Reson.*, **103B**, 203–216.

Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 14652–14657

Shaka, A.J., Barber, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.

Shaka, A.J., Keeler, J., and Freeman, R. (1983) *J. Magn. Reson.*, **53**, 313–340.

Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N. and Montelione, G.T. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 8009–8014.

Torchia, D.A., Sparks, S.W. and Bax, A. (1988) *Biochemistry*, **27**, 5135–5141.

Torchia, D.A., Sparks, S.W. and Bax, A. (1989) *Biochemistry*, **28**, 5509–5524.

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L.E. *J. Am. Chem. Soc.*, **34**, 10025–10035

Wang, W., Kim, R., Jancarik, J., Yokota, H. and Kim, S.H. (2001) *Structure (Camb)*, **9**, 65–71.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.

Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **101B**, 201–205.

Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y., Laue, E.D., Cooper, J.A., Kainosho, M. and Yokoyama, S. (1998) *J. Biomol. NMR*, **11**, 295–306.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D.L., Dahlquist, F.W., Arrowsmith, C.H. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 6464–6465.

Yokoyama, S., Matsuo, Y., Hirota, H., Kigawa, T., Shirouzu, M., Kuroda, Yutaka, Kurumizaka, H., Kawaguchi, S., Ito, Y,. Shibata, T., Kainosho, M., Nishimura, Y., Inoue, Y., Kuramitsu, S. (2000) *Prog. Biophys. Mol. Biol.* **73**, 363–376.

Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., and Montelione, G.T. (1997) *J. Mol. Biol.*, **269**, 592–610.