Journal of Biomolecular NMR, **24:** 231–235, 2002.
KLUWER/ESCOM

Biosynthetically directed fractional ^{13}C labeling facilitates identification of Phe and Tyr aromatic signals in proteins

Jaison Jacob[a], John M. Louis[b], Issa Nesheiwat[b] & Dennis A. Torchia[a,*]
[a]*Molecular Structural Biology Unit, National Institute of Dental and Craniofacial Research, and* [b]*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.*

Received 12 July 2002; Accepted 17 September 2002

Key words: amino acid, aromatic, NMR, non-random ^{13}C labeling, proteins

Abstract

Analysis of 2D [^{13}C,^{1}H]-HSQC spectra of biosynthetic fractionally ^{13}C labeled proteins is a reliable, straightforward means to obtain stereospecific assignments of Val and Leu methyl sites in proteins. Herein we show that the same fractionally labeled protein sample facilitates observation and identification of Phe and Tyr aromatic signals. This is the case, in part, because the fractional ^{13}C labeling yields aromatic rings in which some of the ^{13}C-^{13}C J-couplings, present in uniformly labeled samples, are absent. Also, the number of homonuclear J-coupling partners differs for the δ-, ε- and ζ-carbons. This enabled us to vary their signal intensities in distinctly different ways by appropriately setting the ^{13}C constant-time period in 2D [^{13}C,^{1}H]-HSQC spectra. We illustrate the application of this approach to an 18 kDa protein, c-VIAF, a modulator of apoptosis. In addition, we show that cancellation of the aromatic ^{13}C CSA and ^{13}C-^{1}H dipolar interactions can be fruitfully utilized in the case of the fractionally labeled sample to obtain high resolution ^{13}C constant-time spectra with good sensitivity.

Introduction

It is widely appreciated that strong ^{13}C-^{13}C J-coupling can impede resonance assignment of aromatic rings, particularly those of phenylalanine (Vuister et al., 1994; Wang et al., 1999), but occasionally those of tyrosine and tryptophan as well. Useful approaches have been developed to overcome this problem in the case of Phe residues, but require the preparation of protein samples beyond those normally needed to obtain signal assignments (Vuister et al., 1994; Wang et al., 1999). Herein we show that the biosynthetic fractionally ^{13}C labeled protein sample, normally prepared to obtain stereospecfic methyl assignments (Neri et al., 1989; Senn et al., 1989) can simplify the analysis of constant time, CT, 2D [^{13}C,^{1}H]-HSQC aromatic spectra and facilitate identification and assignment of Phe and Tyr aromatic signals.

Biosynthetic fractional ^{13}C labeling is achieved by expressing proteins in a minimal medium containing a 1:9 mixture of [$^{13}C_6$]glucose and natural abundance glucose as the sole carbon source (Neri et al., 1989; Senn et al., 1989). Fractional ^{13}C labeling was introduced (Neri et al., 1989; Senn et al., 1989) to derive stereospecific signal assignments of Val and Leu methyl groups in proteins in a reliable manner from a simple analysis of 2D [^{13}C,^{1}H]-HSQC spectra. In addition, it has been shown that analysis of carbon multiplet intensities of free amino acids derived from hydrolyzed fractionally ^{13}C labeled proteins is a powerful means to quantitatively investigate metabolic pathways in bacteria (Szyperski, 1995; Szyperski et al., 1992).

*To whom correspondence should be addressed. E-mail: dtorchia@dir.nidcr.nih.gov

Materials and methods

Sample preparation

The C-terminal domain of human VIAF protein (B.W.M. Richter and C.S. Duckett, manuscript in preparation), spanning the region from K87 to D238, was expressed using the pET11a vector and *E. coli* BL21 (DE3) (Novagen, Inc., WI). The nucleotide sequence of the cloned DNA was confirmed by sequencing. Cells were grown at 37 °C either in Luria-Bertani medium or in a modified minimal medium with $^{15}NH_4Cl$ as the sole nitrogen source. In addition, the growth media contained either [$^{13}C_6$]glucose or 10% [$^{13}C_6$]glucose/90% [$^{12}C_6$]glucose as the sole carbon source. Cells derived from 1 l of culture were suspended in 20 volumes of buffer A (50 mM Tris-HCl pH 8.2, 10 mM EDTA, 10 mM DTT) and lysed by sonication at 4 °C in the presence of 100 μg/ml lysozyme. The insoluble recombinant protein was washed by resuspension in buffer B (50 mM Tris-HCl pH 8.2, 10 mM EDTA, 10 mM DTT, 1 M urea and 0.5% Triton X-100). In both cases, the insoluble fraction was pelleted by centrifugation at 20 000 *g* for 30 min at 4 °C. The final pellet was solubilized in 50 mM Tris-HCl, pH 8.0, 7.5 M guanidine-HCl, 5 mM EDTA and 10 mM DTT to yield a concentration of approximately 20 mg ml^{-1}. Thirty mg of protein was applied to a Superdex-75 column (HiLoad 2.6 cm × 60 cm, Amersham Biosciences, NJ) equilibrated in buffer C (50 mM Tris-HCl, pH 8, 4 M guanidine-HCl, 5 mM EDTA, 2 mM DTT) and at a flow-rate of 3 ml min^{-1} at ambient temperature. Peak fractions were pooled and 20 mg (0.2 mg ml^{-1} in buffer C) of protein was folded at room temperature by dialysis against 4 l of 50 mM Tris-HCl, pH 8, 2 M guanidine-HCl and 2 mM DTT for 12–14 h and twice against 4 l of 20 mM sodium phosphate, pH 7.4, 2 mM DTT for 4–6 h. The protein was concentrated to ∼ 10 mg ml^{-1} and applied to Superdex-75 column (HiLoad 2.6 cm × 60 cm) in 20 mM sodium phosphate pH 6.6, 0.02 mM EDTA, 0.01% NaN_3 and 2 mM DTT. Peak fractions were combined, concentrated to ∼ 1 mM protein and stored at 4 °C.

Data acquisition and analysis

NMR spectra were obtained at 35 °C and at a 0.8 mM sample concentration using either a Bruker DMX500 or a Bruker DMX750 spectrometer. A CT [^{13}C,^{1}H]-HSQC spectrum of a uniformly $^{13}C/^{15}N$ enriched sample was acquired at 500.14 MHz using an 18 ms CT ^{13}C evolution period with a ^{1}H π pulse used to refocus ^{1}H-^{13}C J-coupling. CT [^{13}C,^{1}H]-HSQC spectra of a 10% ^{13}C enriched sample were acquired at 499.53 MHz and at 749.0 MHz using a 36 ms CT period without the ^{1}H π pulse. A cold probe, which offered an ca. three-fold increase in sensitivity over the conventional probe, was used to record spectra of the 10% ^{13}C enriched sample at 499.53 MHz. All experiments at 749 MHz were carried out using a conventional probe because a cold probe was not available. In the absence of the ^{1}H π pulse, ^{1}H-^{13}C J-coupling and ^{13}C chemical shift both evolve during the CT period. The upfield doublet (anti-TROSY) component relaxes significantly more rapidly than the downfield component, due to cross correlation between the ^{13}C-^{1}H dipolar and ^{13}C CSA interactions (Goldman, 1984; Pervushin et al., 2000). Because the upfield component decays nearly completely during the 36 ms CT period, there was no need to acquire and process the data in a manner that suppressed this component. All data were processed and analyzed using nmrPipe (Delaglio et al., 1995).

A total of 400 complex t2 (^{1}H) data points were acquired on both DMX750 and DMX500 spectrometers, with sweep widths of 7507.7 Hz and 4496.4 Hz, respectively. The t2 data were apodized using a Lorentz-to-Gauss window function, zero-filled and Fourier transformed. In the ^{13}C time domain (t1), the number of data points was typically set to achieve the maximum possible resolution allowed by the CT period (see figure captions for details). The t1 data were extended by linear prediction, apodized using a cosine squared bell window function, zero-filled and Fourier transformed.

Results and discussion

Our interest in using a fractionally ^{13}C labeled protein sample to help assign Phe/Tyr aromatic signals was occasioned by difficulties in interpreting the aromatic region of a CT [^{13}C,^{1}H]-HSQC spectrum (Vuister and Bax, 1992) of uniformly 98% ^{13}C labeled c-VIAF, Figure 1A. Biochemical evidence indicates that VIAF is a novel protein that modulates apoptosis (B.W.M. Richter and C.S. Duckett, manuscript in preparation). c-VIAF is the C-terminal domain of the full length 238 residue protein, and contains four Tyr and six Phe residues. A portion of the aromatic region of an 18 ms CT ^{1}H-^{13}C HSQC spectrum of a uniformly 98% ^{13}C labeled sample, recorded at 500.14 MHz,

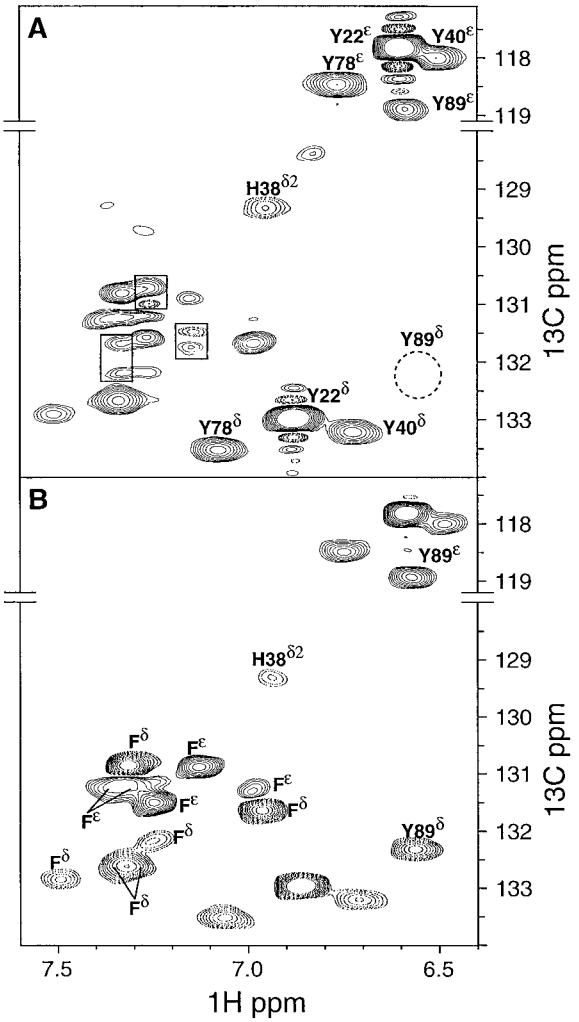

Figure 1. Comparison of Phe and Tyr aromatic regions of 18 ms CT [^{13}C-^{1}H] HSQC spectra of c-VIAF; (A) uniformly 98% ^{13}C labeled and (B) 10% ^{13}C fractionally labeled. The former spectrum was acquired at 500.14 MHz using a standard triple resonance probe with 128 complex t1 points, an F1 spectral width of 44 ppm and 8 scans per free induction decay, while the latter spectrum was recorded at 499.53 MHz using a high-sensitivity cold probe with 59 complex t1 points, an F1 spectral width of 26 ppm and 192 scans per free induction decay. In (A) the dashed circle indicates the location expected of the missing Y89 $C^{\delta}H$ correlation, and the boxes enclose correlations in the Phe region of the spectrum that are strong J-coupling artifacts. In (B) both Phe and Tyr the δ- and ε-correlations are distinguished by the opposite signs of their intensities. The δ-correlations are negative, as indicated by dashed contours, while the ε-correlations are positive. The Y89 $C^{\delta}H$ correlation and Phe $C^{\delta}H$ and $C^{\varepsilon}H$ correlations are observed with chemical shifts that are consistent with cross peaks seen in ^{13}C separated NOESY and ^{1}H TOCSY spectra.

is shown in Figure 1A. Although four Tyr C^{ε}-H correlations are seen, only three Tyr C^{δ}-H correlations are evident. Tyr89 C^{β}-H/C^{δ}-H cross-peaks observed in a 3D ^{13}C-separated NOESY spectrum predicted a Y89 C^{δ}-H correlation in the location indicated by the dashed circle in Figure 1A. This prediction was consistent with the absence of a Y89 H^{δ}-H^{ε} correlation in the 2D ^{1}H TOCSY spectrum (of a sample in natural abundance ^{13}C, not shown) because the Tyr89 H^{δ} and H^{e} protons would have nearly identical chemical shifts at 6.57 ppm. The Tyr89 C^{δ}-H correlation is missing in Figure 1(A), presumably because the Tyr89 C^{γ} and C^{δ} chemical shifts also coincide. Such a coincidence of chemical shifts results in strong C^{γ}-C^{δ}-coupling which spoils refocusing of carbon J-coupling during the 18 ms ^{13}C CT period of the pulse sequence. In general, strong coupling is expected to occur more frequently in the case of Phe residues, because the Phe δ-, ε- and ζ-carbons have similar chemical shifts (Wishart et al., 1995), and is the apparent source of the artifacts enclosed within boxes in Figure 1A.

In contrast with the results just discussed, all four Tyr C^{δ}-H correlations are observed in the 18 ms CT 2D [^{13}C,^{1}H]-HSQC spectrum of the 10% ^{13}C fractionally labeled sample, Figure 1B. As expected, the Tyr89 C^{δ}-H correlation has a proton shift of 6.56 ppm, nearly identical to that of the Tyr89 H^{ε}. It should also be noted that in Figure 1B the intensities of the Tyr C^{δ} and $C^{\varepsilon}H$ correlations have opposite signs. These observations are in agreement with the analysis of ^{13}C multiplet component intensities of tyrosine derived from fractionally ^{13}C labeled and hydrolyzed P22 c2 repressor (Szyperski, 1995). Szyperski (1995) reported that 80–90% of the tyrosine δ-carbons were directly bonded to a single ^{13}C, whereas the 80–90% of the ε-carbons were bonded to either zero or two ^{13}C spins. The Tyr89 C^{δ} correlation is observed in Figure 1B because a significant fraction of the Tyr89 δ-carbons are J-coupled solely to the Tyr89 ε-carbons in the fractionally labeled sample.

The ^{13}C-^{13}C multiplet signal intensities of the sharp, intense Tyr22 δ- and ε-carbon signals were clearly observed (data not shown) in the ^{13}C coupled 2D [^{13}C,^{1}H]-HSQC spectrum of c-VIAF, and are in qualitative agreement with the corresponding signal intensities of the carbon multiplets measured for the free amino acid (Szyperski, 1995). In addition, the opposite signs of Tyr δ- and ε-carbon signal intensities seen in Figure 1B was observed in a spectrum (data not shown) of 10% fractionally ^{13}C labeled ribosomal protein S4 Δ41. These observations suggest

that the ^{13}C labeling patterns will be similar in proteins that are derived from aerobic *E. coli* growths. But, as previously noted (Szyperski, 1995), the use of auxotrophic strains or substantially different growth conditions could significantly modify the observed ^{13}C labeling patterns.

Because Phe and Tyr residues are synthesized by a common pathway, the signs of Phe δ- and ε-carbon signal intensities observed in Figure 1B are the same as observed for Tyr. The active metabolic pathways identified by Szypersky's (1995) analysis indicate that the fraction of Phe ζ-carbons bonded to a single ^{13}C is comparable to the sum of the fractions bonded to zero or two ^{13}C spins. Because signals corresponding to these respective labeling patterns have opposite signs (in an 18 ms CT experiment) signal cancellation significantly reduces the intensity of the ζ-carbon correlations, and is presumably the reason they are not observed in Figure 1B.

Signal cancellation in the spectrum of the fractionally labeled sample can be circumvented by setting the CT period to an integral multiple of $2/J_{CC}$ (36 ms). This CT period affords an ca. five-fold increase in digital resolution compared to a non-CT spectrum of a 99% ^{13}C labeled sample, where the effects of strong ^{13}C-^{13}C coupling are suppressed by a short ^{13}C evolution period (ca. 8 ms). The gain in spectral resolution will greatly facilitate the assignment of aromatic signals except, in rare cases, when they are all well separated. In addition, the long CT period allows one to take advantage of the increase in T_2 that results from cross-correlation (cancellation) of the ^{13}C-^{1}H dipolar and ^{13}C CSA interactions (Goldman, 1984; Pervushin et al., 1998) to selectively observe the signal of the downfield component of the ^{1}H-^{13}C doublet, Figures 2A–C. The spectra shown in Figure 2 were obtained simply by retaining ^{1}H-^{13}C coupling evolution during the 36 ms ($2/J_{CC}$) CT period. In the 499.53 MHz spectrum of the 10% fractionally labeled sample, Figure 2A, the C^{δ} and $C^{\varepsilon}H$ correlations have the same sign, as expected. However, only two $C^{\zeta}H$ correlations are seen, and these signals are barely above the noise level. At 749.0 MHz, where the TROSY effect is enhanced, the 36 ms CT TROSY spectrum of the fractionally labeled sample reveals at least five of the six expected $C^{\zeta}H$ correlations. In other respects, the spectrum is similar to Figure 2A. As expected, the 36 ms CT-TROSY spectrum of the 98% uniformly ^{13}C labeled sample, Figure 2C, has better signal-to-noise than the spectrum of the fractionally labeled sample, Figure 2B; however, strong

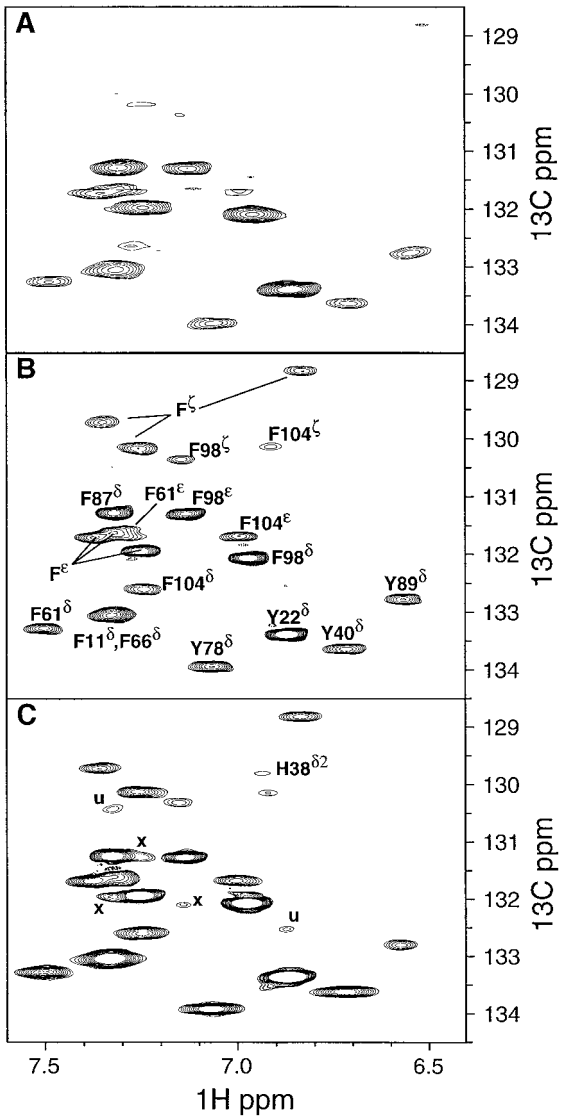

Figure 2. Comparison of Phe aromatic regions of CT-TROSY [^{13}C-^{1}H] HSQC spectra recorded with a 36 ms CT period. All spectra were acquired with a sweep width of 26 ppm in the ^{13}C dimension and 256 scans per fid. (A) Ten percent ^{13}C fractionally labeled c-VIAF recorded at 499.53 MHz on a high-sensitivity cold probe with 112 complex t1 points; (B) 10% ^{13}C fractionally labeled c-VIAF recorded at 749.0 MHz with 139 complex t1 points; (C) 98% ^{13}C uniformly labeled c-VIAF recorded at 749.0 MHz with 138 complex t1 points. All spectra were recorded without proton decoupling during t1 (^{13}C evolution). Typically only the downfield component of each ^{13}C-^{1}H doublet is observed in the F1 dimension because the signal of upfield (anti-TROSY) component decays rapidly. In (B) the three Phe C^{ε}-H and C^{ζ}-H correlations labeled F^{ε} and F^{ζ}, respectively, have been assigned to F11, F66 and F87. These three Phe residues have identical $^{1}H^{\delta}$ shifts, complicating sequential assignment of their ε- and ζ-signals, which is underway. The F61 $^{1}H^{\zeta}$ has been assigned at 7.23 ppm and assignment of its attached carbon is in progress. In (C), (u) identifies weak anti-TROSY signals, and (x) identifies strong coupling artifacts.

coupling artifacts remain even at the high external field, peaks marked x, in Figure 2C. In addition, while the Tyr89 $C^{\delta}H$ correlation is observed, its intensity is much weaker than that of the three other Tyr89 $C^{\delta}H$ correlations. Its weak intensity would make it difficult to identify in a more crowded spectrum where it would be less likely to be well resolved from the intense signals.

Note that our discussion of the dependence of the signs of Tyr and Phe signal intensities on the CT period applies only in the weak coupling condition. Evidently, as a consequence of fractional ^{13}C labeling, the signal intensities of all c-VIAF Phe and Tyr residues are in accord with predictions based upon weak coupling. Clearly, this need not be the case for an aromatic residue experiencing strong coupling, that is not reduced by fractional labeling. The effects of strong coupling can be eliminated, while retaining good resolution, by recording an HSQC spectrum of a ^{13}C natural abundance sample. However the signal sensitivity is significantly less than obtained using a biosynthetically directed 10% ^{13}C labeled sample, and Phe $^{13}C^{\delta}$ and $^{13}C^{\varepsilon}$ correlations cannot be distinguished.

The 2D [^{13}C,^{1}H]-HSQC spectra in Figures 1B and 2B together with the aromatic ^{1}H natural abundance TOCSY spectrum yields identification of all six Phe C^{δ}-H and C^{ε}-H correlations. The proton chemical shifts of the five well resolved C^{ζ}-H correlations observed in Figure 2B are confirmed by proton chemical shifts observed in the ^{1}H TOCSY spectrum. Only the F61 C^{ζ}-H correlation remains to be identified. The ^{1}H TOCSY spectrum indicates that the Phe 61 H^{ζ} chemical shift is at 7.23 ppm. It seems likely that the F61 C^{ζ}-H correlation overlaps another C^{ζ}-H correlation in Figure 2B, but its assignment at this time remains uncertain. The chemical shifts of six Phe and four Tyr correlations $C^{\delta}H$ are in agreement with Phe and Tyr $C^{\beta}H$-$C^{\delta}H$ cross-peaks observed in 3D- and 4D ^{13}C separated NOESY spectra, and the proton positions of the six Phe $C^{\varepsilon}H$ correlations are in agreement with the signals observed in the ^{1}H TOCSY spectrum. We are currently determining the structure of c-VIAF and will present the final set of signal assignments when the structure determination is completed.

Conclusions

The 10% ^{13}C fractionally labeled protein sample used to make stereospecific methyl assignments can facilitate analysis of the aromatic region of CT 2D [^{13}C,^{1}H]-HSQC spectra. The effects of strong J-coupling on CT spectra are reduced as compared with the fully labeled sample, while retaining the benefits of high resolution in the carbon dimension. In addition, when fractionally ^{13}C labeled aromatic rings are weakly J-coupled, the characteristic signal intensities observed in CT 2D [^{13}C,^{1}H]-HSQC spectra provide a simple means to distinguish and identify δ-, ε- and ζ-correlations. It can be easily appreciated that in larger proteins, where spectral overlap often precludes assignment of aromatic residues, signal assignments will be greatly facilitated by the high-resolution approach described herein. The method is expected to be applicable to any 10% ^{13}C enriched protein derived from an aerobic *E. coli* growth, using the standard labeling protocol (Neri et al., 1989; Senn et al., 1989).

Acknowledgements

We thank Bettina Richter and Colin Duckett for making the hVIAF clone used in this study available to us at an early stage of their investigation. We also thank Frank Delaglio and Dan Garrett for processing software, Dusty Baber for hardware support, Michelle Markus for the sample of protein S4 Δ41, and Andrew Hinck and Rieko Ishima for helpful comments. This work was supported by the Intramural AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

References

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

Goldman, M. (1984) *J. Magn. Reson.*, **60**, 437–452.

Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry*, **28**, 7510–7516.

Pervushin, K., Braun, D., Fernandez, C. and Wüthrich, K. (2000) *J. Biomol. NMR*, **17**, 195–202.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1998) *J. Am. Chem. Soc.*, **120**, 6394–6400.

Senn, H., Werner, B., Messerle, B.A., Weber, C., Traber, R. and Wüthrich, K. (1989) *FEBS Lett.*, **249**, 113–118.

Szyperski, T. (1995) *Eur. J. Biochem.*, **232**, 433–448.

Szyperski, T., Neri, D., Leiting, B., Otting, G. and Wüthrich, K. (1992) *J. Biomol. NMR.*, **2**, 323-334.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.

Vuister, G.W., Kim, S.J., Wu, C. and Bax, A. (1994) *J. Am. Chem. Soc.*, **116**, 9206–9210.

Wang, H., Janowick, D.A., Schkeryantz, J.M., Liu, X.H. and Fesik, S.W. (1999) *J. Am. Chem. Soc.*, **121**, 1611–1612.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.