Assigning the NMR spectra of aromatic amino acids in proteins: analysis of two Ets pointed domains

Carolyn M. Slupsky, Lisa N. Gentile, and Lawrence P. McIntosh

Abstract: The measurement of interproton nuclear Overhauser enhancements (NOEs) and dihedral angle restraints of aromatic amino acids is a critical step towards determining the structure of a protein. The complete assignment of the resonances from aromatic rings and the subsequent resolution and identification of their associated NOEs, however, can be a difficult task. Shown here is a strategy for assigning the ^{1}H, ^{13}C, and ^{15}N signals from the aromatic side chains of histidine, tryptophan, tyrosine, and phenylalanine using a suite of homo- and hetero-nuclear scalar and NOE correlation experiments, as well as selective deuterium isotope labelling. In addition, a comparison of NOE information obtained from homonuclear NOE spectroscopy (NOESY) and ^{13}C-edited NOESY – heteronuclear single quantum correlation experiments indicates that high-resolution homonuclear two-dimensional NOESY spectra of selectively deuterated proteins are invaluable for obtaining distance restraints to the aromatic residues.

Key words: NMR assignment, aromatic residue, transcription factor, NOE, dihedral angle.

Résumé : La détermination des effets Overhauser nucléaires (NOE) entre protons et des contraintes d'angle dièdre des acides aminés aromatiques est une étape importante lors de la détermination de la structure d'une protéine. Cependant, l'attribution complète des résonances des noyaux aromatiques, ainsi que l'identification et la résolution des NOE associés peuvent être des tâches ardues. Dans cet article, nous indiquons une stratégie pour attribuer les signaux de ^{1}H, ^{13}C et ^{15}N des chaînes aromatiques latérales de l'histidine, du tryptophane, de la tyrosine et de la phénylalanine à l'aide d'une série d'expériences de corrélation de NOE scalaires homo- et hétéro-nucléaires, ainsi qu'à un marquage sélectif au deutérium. De plus, la comparaison des mesures de NOE obtenues au cours d'expériences de spectroscopie NOESY homonucléaire et NOESY-HSQC hétéronucléaire ^{13}C-éditée indique que les spectres NOESY bidimensionnels homonucléaires à haute résolution de protéines marquées sélectivement au deutérium sont d'une valeur inestimable pour déterminer les contraintes de distance des résidus aromatiques.

Mots clés : attribution de RMN, résidu aromatique, facteur de transcription, NOE, angle dièdre.

[Traduit par la Rédaction]

Introduction

Aromatic residues are, for the most part, hydrophobic in nature and tend to lie in the interior of proteins (Creighton 1993). These residues are of considerable interest to the biological NMR spectroscopist for several reasons: they have ^{1}H, ^{13}C, and ^{15}N chemical shifts distinct from those of the aliphatic amino acids; they increase the chemical shift dispersion of neighboring nuclei owing to ring current effects; and, at least in principle, they provide a wealth of structural and dynamic information about a protein. Unfortunately, the spectra of aromatic residues are often difficult to assign because the aromatic ^{1}H spin system is uncoupled to that of the main chain, and because of the possibility of flipping of tyrosine and phenylalanine rings on the slow, intermediate, or fast time scales. Furthermore, limited chemical shift dispersion, frequently observed with phenylalanine side chains, often precludes the unambiguous identification of nuclear Overhauser enhancement (NOE) interactions that are necessary for accurate structure determination.

Traditional methods for assigning ring protons in unlabelled proteins include using the double quantum filtered correlated spectroscopy (DQF-COSY) and total correlation spectroscopy (TOCSY) experiments to connect the reso-

Received February 6, 1998. Revised April 6, 1998. Accepted April 14, 1998.

Abbreviations: NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; DQF-COSY, double quantum filtered correlated spectroscopy; HMBC, heteronuclear multiple bond correlation; HSQC, heteronuclear single quantum correlation; CT, constant time; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; His, histidine; Trp, tryptophan; Tyr, tyrosine; Phe, phenylalanine.

C.M. Slupsky, L.N. Gentile, and L.P. McIntosh.[1] Department of Biochemistry and Molecular Biology, Department of Chemistry, and the Protein Engineering Network of Centres of Excellence, 2146 Health Sciences Mall, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.

[1]Author to whom all correspondence should be addressed (e-mail: mcintosh@otter.biochem.ubc.ca).

 © 1998 NRC Canada

nances from ring protons, followed by a NOE spectroscopy (NOESY) experiment to provide a link to the main-chain atoms (Wüthrich 1986). Recently, the $C_\beta H_\delta$ and $C_\beta H_\epsilon$ experiments were developed to correlate the chemical shift of the $^{13}C_\beta$ nucleus in a uniformly ^{13}C-labelled aromatic amino acid to that of its $^1H_\delta$ or $^1H_\epsilon$, respectively, via heteronuclear scalar couplings (Yamazaki et al. 1993). A related approach for linking the $^{13}C_\beta$ to the aromatic ring carbons and protons using audio-frequency NMR in the nutating frame was also described by Grzesiek and Bax (1995). Additional experiments that aid in the assignment of aromatic resonances include the three-dimensional ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC (Zerbe et al. 1996), which correlates the chemical shift of a given aromatic proton with that of the directly attached carbon as well as all vicinal protons. A similar approach, the (HC)C-(C)CH-TOCSY experiment, connects the $^{13}C_\gamma$ resonance with the $^{13}C_\beta/^1H_\beta$, $^{13}C_\delta/^1H_\delta$, or the remaining aromatic carbon and proton resonances (Löhr and Rüterjans 1996). However, in a protein that contains many phenylalanine residues, poor ^{13}C and ^{1}H chemical shift dispersion of the strongly coupled carbon spin systems often makes it very difficult to resolve and assign the resonances from these residues. Because of this problem, a reverse labelling strategy, wherein natural abundance (^{12}C/^{14}N)-Phe is incorporated into an otherwise fully ^{13}C/^{15}N labelled protein, combined with a set of ^{13}C-edited and ^{12}C-filtered NOESY experiments, has proven successful for the assignment of the resonances from these aromatic side chains and the subsequent identification of their NOE interactions (Vuister et al. 1994).

To assign the NMR spectra of aromatic residues in proteins and to obtain useful distance and dihedral angle restraints, we have developed a simple strategy employing a combination of homo- and hetero-nuclear scalar and NOE correlation experiments, combined with selective deuterium isotope labelling. This approach is illustrated with the pointed domains from two members of the Ets family of transcription factors, Ets-1 and GABPα (Graves and Petersen 1997). NMR studies of these domains, which are postulated to mediate intra- or inter-molecular protein–protein interactions, will help provide a structural foundation for understanding the regulation of gene expression by eukaryotic transcription factors.

Materials and methods

Protein samples

The gene encoding the pointed domain (residues 29–138) of murine Ets-1 was PCR-amplified from the full-length Ets-1 cDNA (Gunther et al. 1990) as an *Nde*I–*Hin*dIII fragment and cloned into a T7-based expression system (pET22b; Invitrogen, La Jolla, Calif.). The GABPα pointed domain was cloned into the same vector from GABPα cDNA using an *Nde*I–*Eco*RI fragment containing the coding sequence for residues 168–254. Each protein was expressed in *Escherichia coli* BL21(λDE3) cells. Protocols for the preparation and purification of unlabelled and uniformly ^{13}C/^{15}N labelled Ets-1 or GABPα will be reported elsewhere manuscripts in preparation. Samples of the Ets-1 pointed domain containing ^{2}H-[$\gamma^{1,2}$,$\delta^{1,2}$,ζ]-Phe and the GABPα pointed domain containing either ^{2}H-[$\gamma^{1,2}$,$\delta^{1,2}$,ζ]-Phe and ^{2}H-[$\gamma^{1,2}$,$\delta^{1,2}$]-Tyr or ^{2}H-[$\gamma^{1,2}$,$\delta^{1,2}$,ζ]-Phe and ^{2}H-[δ^1,ζ^2,ζ^3,ϵ^3,η^2]-Trp were prepared as previously described (McIntosh and Dahlquist 1990; McIntosh et al. 1990). The protein samples were in 10 mM K_2HPO_4, 10 mM KCl, 10 mM dithiothreitol, pH 6.5 (Ets-1 pointed domain), or 20 mM Na_2HPO_4, 20 mM NaCl, 10 mM dithiothreitol, pH 7.2 (GABPα pointed domain).

NMR spectroscopy

Spectra of the Ets-1 and GABPα pointed domains were collected at 30°C on a Varian Unity 500 NMR spectrometer equipped with a pulsed-field gradient accessory. NMR data were processed using NMRpipe (Delaglio et al. 1995) and analyzed using the interactive peak-picking program PIPP (Garrett et al. 1991). Two-dimensional DQF-COSY, TOCSY (MLEV-17 τ_{mix} = 60 ms), and NOESY (τ_{mix} = 75 ms) spectra were obtained for unlabelled or selectively deuterated proteins in either D_2O or H_2O buffer. Two-dimensional ^{1}H-^{15}N heteronuclear multiple bond correlation (HMBC) in mixed mode with a single transfer delay of 22 ms (Bax and Marion 1988; Pelton et al. 1993), two-dimensional ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC), and three-dimensional ^{15}N-edited NOESY-HSQC (τ_{mix} = 75 ms; Zhang et al. 1994) spectra were recorded for uniformly ^{15}N-labelled proteins in H_2O buffer. Two-dimensional constant time ^{1}H-^{13}C HSQC (Santoro and King 1992; Vuister and Bax 1992), two-dimensional $C_\beta H_\delta$ and $C_\beta H_\epsilon$ (Yamazaki et al. 1995), two-dimensional ^{13}C′-{$^{13}C_\gamma$} and ^{15}N-{$^{13}C_\gamma$} spin echo (Hu et al. 1997), and three-dimensional ^{13}C-edited NOESY-HSQC (τ_{mix} = 75 ms; Pascal et al. 1994) spectra were acquired with uniformly ^{13}C/^{15}N-labelled samples in H_2O buffer. Typical spectral widths were 7200 Hz in ^{1}H, 4000 Hz in ^{13}C, and 10 000 Hz (HMBC) or 1230 Hz (HSQC or three-dimensional ^{15}N-edited NOESY-HSQC) in ^{15}N for Ets-1. Chemical shifts are referenced to an external sample of sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), as described by Wishart et al. (1995*a*).

Results and discussion

Assignment strategy: overview

The strategy used here is a combination of traditional two-dimensional ^{1}H methods for the assignment of the resonances from the aromatic ring protons of histidine, tryptophan, tyrosine, and phenylalanine, combined with modern ^{13}C and ^{15}N edited methods to exploit the chemical shift information and the added dispersion of the heterospin. Selective deuteration is utilized in parallel to simplify the aromatic region of the ^{1}H-NMR spectra (McIntosh et al. 1990; Johnson et al. 1996). For example, with a protein sample containing ring-deuterated phenylalanine and tryptophan, the only protons observed downfield of ~6 ppm arise from the tyrosine and histidine side chains (assuming that D_2O buffer is utilized to eliminate signals from the labile amide groups).

Central to this assignment strategy is the constant time ^{1}H-^{13}C HSQC experiment, illustrated in Fig. 1 for the pointed domains of Ets-1 and GABPα. The advantages of this experiment are threefold. First, just as the ^{1}H-^{15}N HSQC spectrum is a fingerprint of the amide backbone of a protein, the ^{1}H-^{13}C HSQC spectrum is a fingerprint of the aromatic side chains. That is, the experiment correlates the chemical shifts of each ^{13}C-^{1}H pair, and thus the spectrum should contain two peaks for each histidine, five for each tryptophan, and depending on the rate of ring flipping, three or five and two or four for each phenylalanine and tyrosine, respectively. Second, the ring carbons of the aromatic amino acids have diagnostic ^{13}C chemical shifts, thus providing excellent peak dispersion in the ^{1}H-^{13}C HSQC spectrum (Richarz and Wüthrich 1978; Wishart et al. 1995*b*). For example, the

© 1998 NRC Canada

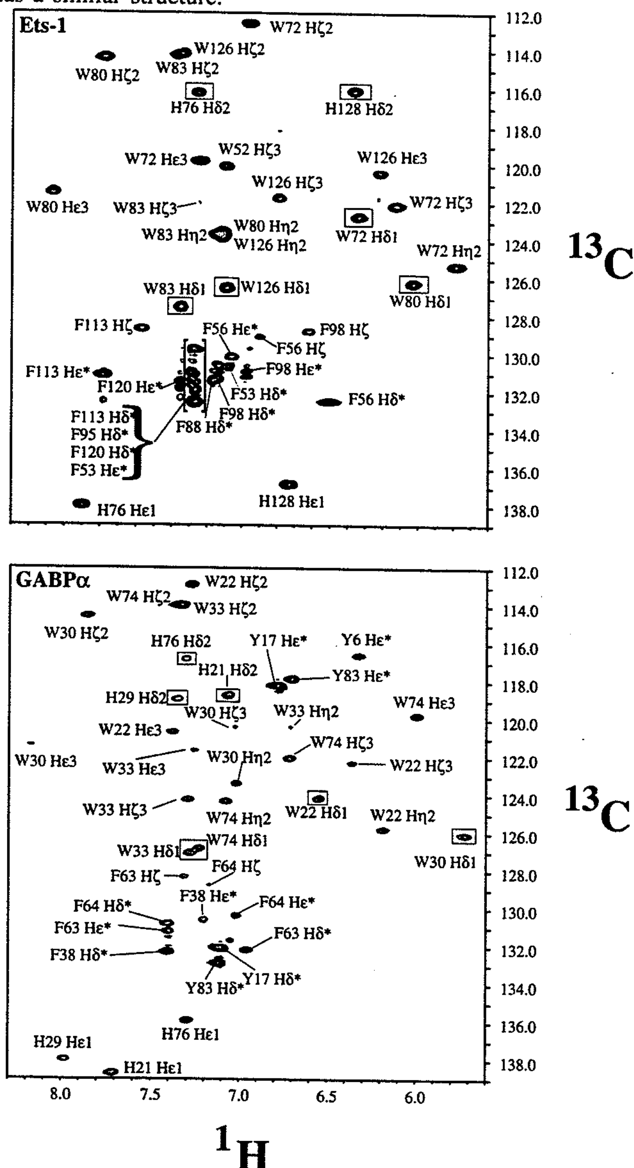


Fig. 1. Constant time ^{1}H-^{13}C HSQC spectra of the Ets-1 and GABPα pointed domains. Negative peaks from the tryptophan $^{13}C_{\delta1}/^{1}H_{\delta1}$ and histidine $^{13}C_{\delta2}/^{1}H_{\delta2}$ resonances are boxed. Peaks from the phenylalanine rings are distorted owing to strong coupling effects. Ets-1 pointed domain numbering starts at the N-terminus of the full-length Ets protein, whereas the numbering for the GABPα pointed domain starts from the first residue of the cloned sequence (GABPα residue 168). It is interesting to note the similarities in the relative dispersion of peaks from corresponding residues in the spectra of the two homologous proteins. For instance, Trp-72 in Ets-1 corresponds to Trp-22 in GABPα. Trp-126, -83, and -80 in Ets-1 correspond to Trp-74, -33, and -30, respectively, in GABPα. This implies that each protein has a similar structure.

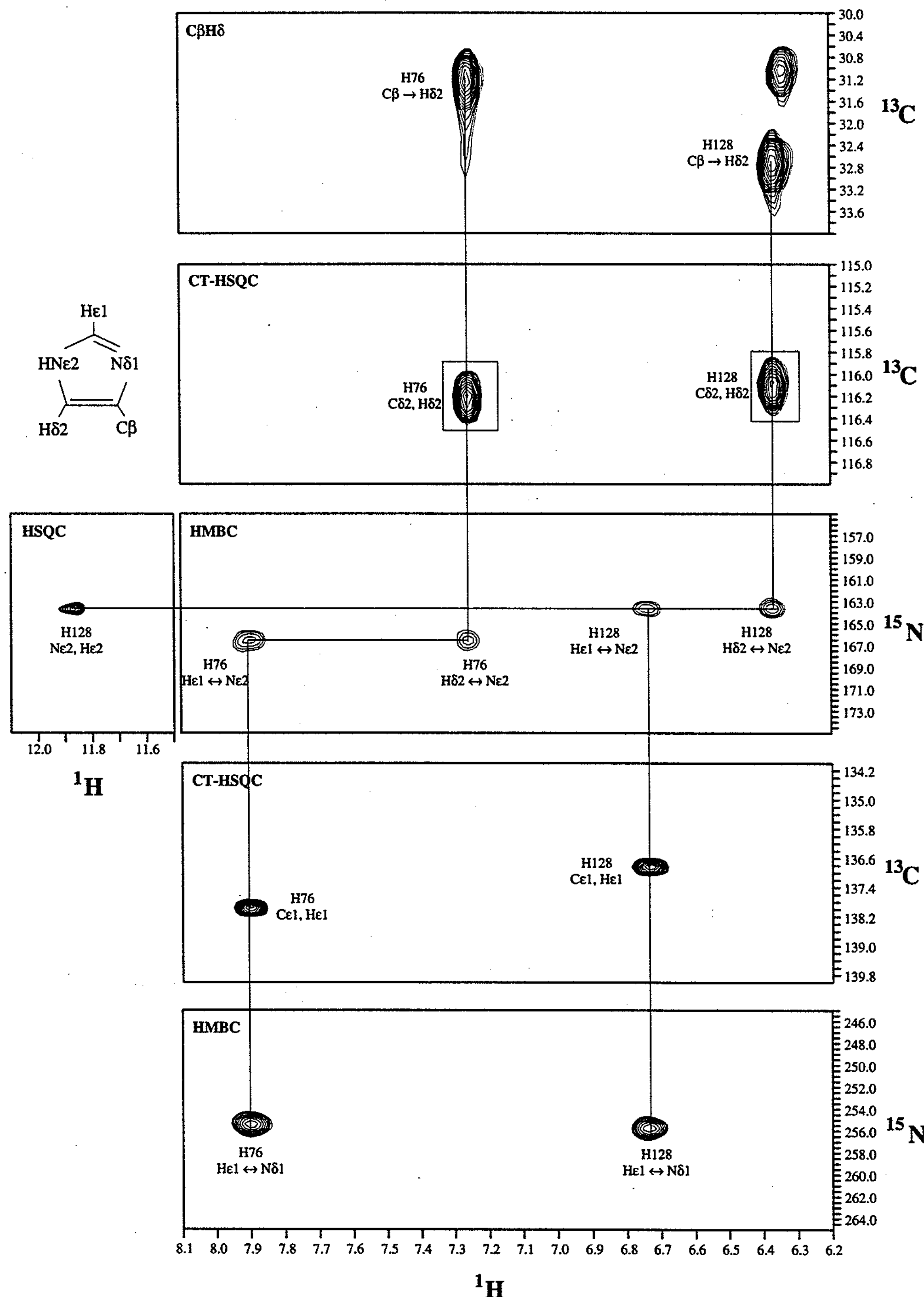
histidine $^{13}C_{\delta2}$ signals fall between 114.0 and 121.0 ppm, whereas the $^{13}C_{\epsilon1}$ signals have shifts between 135.0 and 139.0 ppm. The exact chemical shifts are dependent upon the ionization state of the imidazole ring (Yu and Fesik 1994). Tryptophan $^{13}C_{\zeta2}$ peaks are found between 112.0 and 116.0 ppm, the $^{13}C_{\zeta3}$ and $^{13}C_{\epsilon3}$ peaks are situated between 118.0 and 123.0 ppm, the $^{13}C_{\eta2}$ peaks are set between 122.0 and 126.0 ppm, and the $^{13}C_{\delta1}$ peaks are positioned between 122.0 ppm and 128.0 ppm. Tyrosine $^{13}C_{\delta}$ peaks have chemical shifts between 131.0 and 134.0 ppm, and the $^{13}C_{\epsilon}$ peaks are located between 116.0 and 120.0 ppm. Phenylalanine $^{13}C_{\zeta}$ peaks are situated between 128.0 ppm and 130.0 ppm and the $^{13}C_{\delta}$ and $^{13}C_{\epsilon}$ peaks are set between 129.0 and 134.0 ppm. Third, the constant time mode of acquisition allows for the discrimination of signals on the basis of the topology of the ^{13}C spin system (Santoro and King 1992; Vuister and Bax 1992). Specifically, using a total constant time delay of $1/(^{1}J_{CC}) = 1/(60\ Hz)$, ^{13}C-^{1}H pairs with one ^{13}C neighbor are inverted relative to those with zero or two neighbors. Thus, peaks corresponding to the $^{13}C_{\delta1}/^{1}H_{\delta1}$ correlations of tryptophan indole rings and the $^{13}C_{\delta2}/^{1}H_{\delta2}$ correlations of histidine imidazole rings are readily identified by their negative sign.

Using the constant time ^{1}H-^{13}C HSQC spectrum as a reference, connections between protons within an aromatic ring are provided by conventional two-dimensional DQF-COSY and TOCSY experiments (Figs. 2–5). Assignment of the resonances from the carbons directly bonded to these protons is then obtained by careful alignment of the homonuclear correlation spectra with the ^{1}H dimension of the constant time ^{1}H-^{13}C HSQC spectrum. Finally, the ^{13}C and ^{1}H assignments of each aromatic ring can be verified on the basis of intraresidue NOEs resolved in a three-dimensional ^{13}C-edited NOESY-HSQC spectrum (Figs. 6 and 7). Clearly it is important that these series of spectra be measured under very similar experimental conditions. In general, we record two-dimensional DQF-COSY and TOCSY spectra with an unlabelled protein sample in D_2O buffer and have not found any problem with their alignment to the ^{1}H-^{13}C HSQC spectrum of the corresponding sample of ^{13}C/^{15}N-labelled protein in H_2O. To reduce the possibility of misaligning peaks, one could certainly examine a single labelled sample and apply ^{13}C decoupling during the t1 and t2 evolution periods of the homonuclear correlation experiments, or alternatively, record a natural abundance ^{1}H-^{13}C HSQC on the unlabelled protein. As discussed previously, an alternative tack for assigning the spectra of the aromatic rings involves the three-dimensional ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC (Zerbe et al. 1996) or (HC)C-(C)CH-TOCSY (Löhr and Rüterjans 1996) experiments. However, on the basis of our experience, the high resolution and signal-to-noise ratio of two-dimensional ^{1}H DQF-COSY and TOCSY experiments, relative to their three-dimensional counterparts, combined with the spectral simplification afforded by selective deuteration, permit the confident assignment of aromatic amino acids in proteins.

The final step of the assignment strategy involves recording $C_{\beta}H_{\delta}$ and $C_{\beta}H_{\epsilon}$ experiments, which correlate the chemical shifts of the ring protons to that of the corresponding side-chain $^{13}C_{\beta}$. Using a plethora of three-dimensional ^{1}H-^{13}C-^{15}N correlation experiments, it is now routine to specifically assign the resonances from the main-chain and aliphatic side-chain nuclei of a protein (Clore and Gronenborn 1994). The $C_{\beta}H_{\delta}$ and $C_{\beta}H_{\epsilon}$ experiments therefore provide sequence-specific identification of the aromatic

© 1998 NRC Canada

Fig. 2. Assignment of the ^{1}H, ^{13}C, and ^{15}N resonances from the two histidine rings in the Ets-1 pointed domain. The individual experiments utilized are named in the upper left-hand corner of each box. Negative $^{13}C_{\delta2}/^1H_{\delta2}$ peaks in the ^{1}H-^{13}C CT-HSQC spectrum are boxed. Both histidine residues are in the neutral $N_{\epsilon2}$ tautomeric form on the basis of their ^{15}N chemical shifts and patterns of two- and three-bond ^{1}H-^{15}N couplings. The labile $H_{\epsilon2}$ of the buried His-128 is detected in a ^{1}H-^{15}N HSQC spectrum recorded without perturbation of water magnetization.



spin systems. These experiments rely exclusively on through-bond scalar correlations and thus avoid the potential ambiguities associated with more traditional NOE-based methods of assigning aromatic ring protons (Wüthrich 1986).

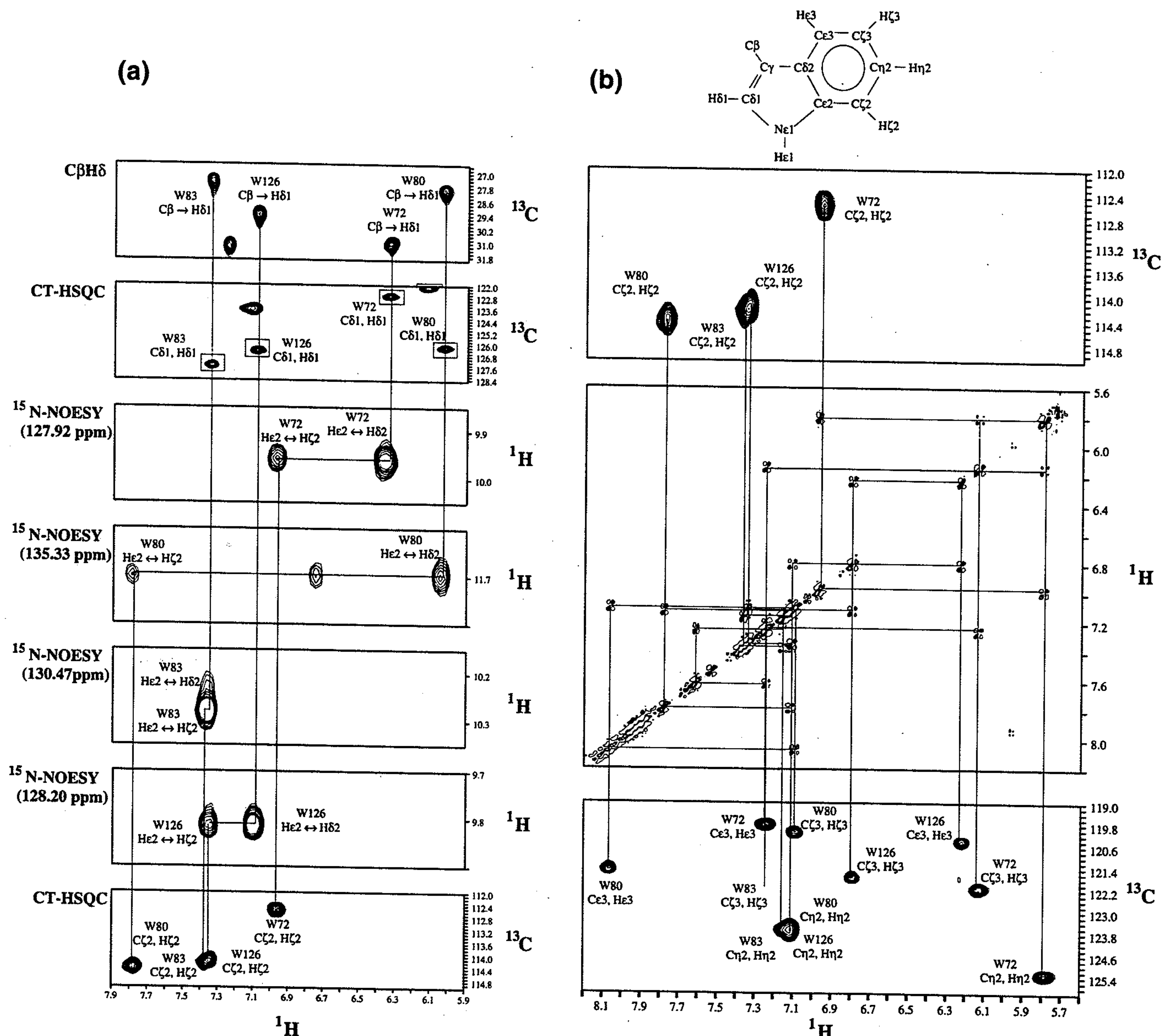
Specific examples of this assignment strategy applied to the aromatic rings of the pointed domains of Ets-1 and GABPα are discussed in the following sections.

Histidine

The assignment of the ^{1}H, ^{13}C, and ^{15}N resonances from the two histidine rings found in the Ets-1 pointed domain is presented in Fig. 2. Using the $C_\beta H_\delta$ experiment, the $^1H_{\delta2}$ assignments of His-76 and His-128 were made, linking the rings to the backbone of the protein. The corresponding $^{13}C_{\delta2}$ assignments were extracted by careful alignment of the

© 1998 NRC Canada

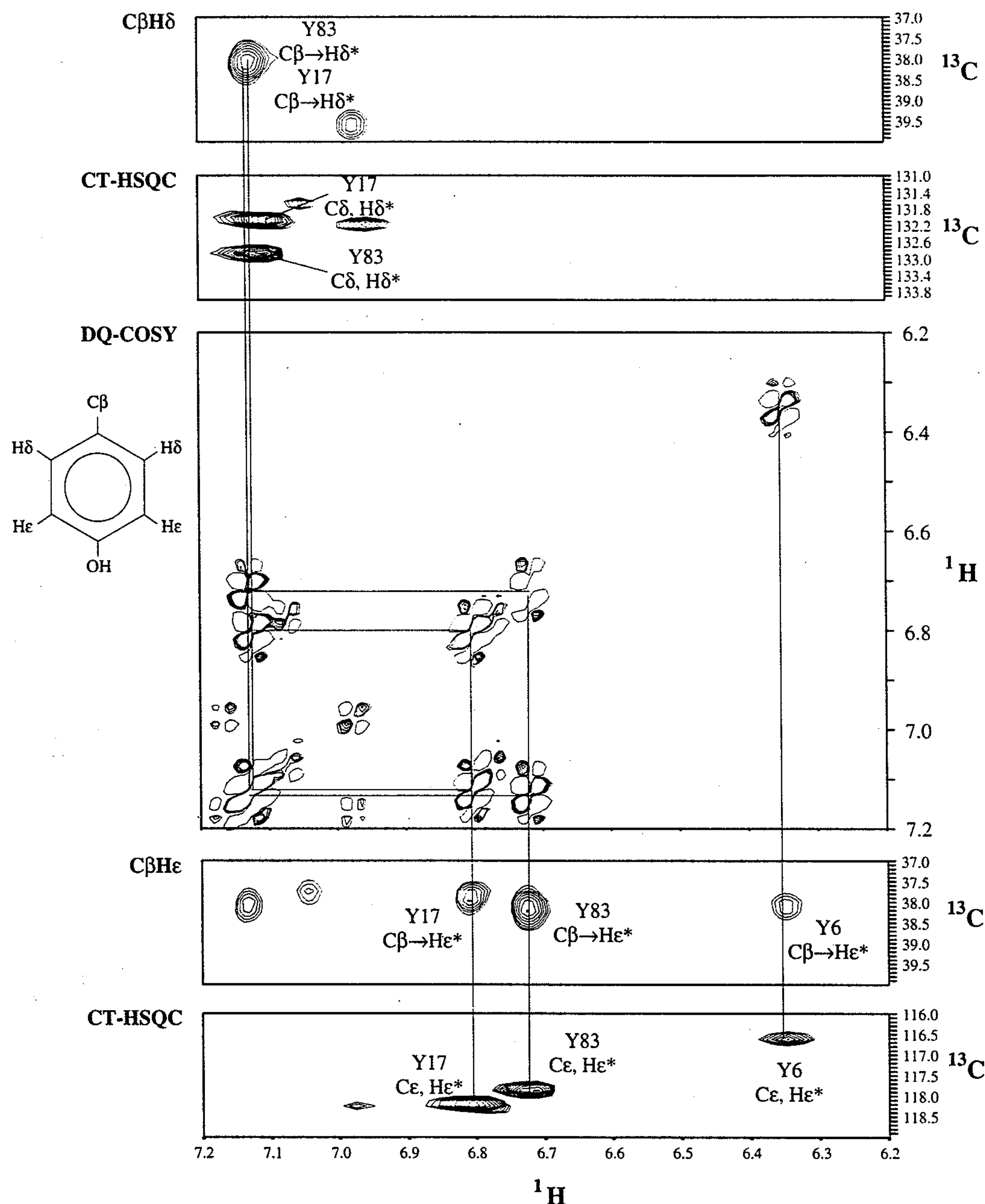
Fig. 3. Assignment of the ^{1}H, ^{13}C, and ^{15}N resonances from the four tryptophan rings in the Ets-1 pointed domain. The individual experiments utilized are indicated beside the upper left-hand corner of each box. Negative $^{13}C_{\delta1}/^{1}H_{\delta1}$ peaks in the ^{1}H-^{13}C CT-HSQC spectrum are boxed, and negative contours beside the DQF-COSY spectrum are indicated by a single contour. (*a*) Assignments of resonances from the $^{13}C_{\delta1}$, $^{1}H_{\delta1}$, $^{15}N_{\epsilon1}$, and $^{1}H_{\epsilon1}$ nuclei are provided by scalar correlations detected in the $C_{\beta}H_{\delta}$, ^{1}H-^{13}C CT-HSQC, and ^{1}H-^{15}N HSQC spectra. Connections to the $^{1}H_{\zeta2}$ and hence $^{13}C_{\zeta2}$ resonances are provided by the ^{15}N-edited NOESY-HSQC experiment. (*b*) Assignments of the resonances from the $^{1}H_{\eta2}$, $^{1}H_{\zeta3}$, and $^{1}H_{\epsilon3}$ nuclei, and their directly bonded carbons, are obtained from a DQF-COSY experiment, followed by alignment with the ^{1}H-^{13}C CT-HSQC spectrum. The DQF-COSY spectrum was recorded with a selectively ^{2}H-[$\gamma^{1,2}$,$\delta^{1,2}$,ζ]-Phe labelled sample in D_2O buffer, and thus only signals from the tryptophan rings are observed. (Note that histidine protons are detected in the corresponding TOCSY spectrum.)



$C_{\beta}H_{\delta}$ spectrum to the constant time ^{1}H-^{13}C HSQC, in which these peaks are negative. The ^{1}H-^{15}N HMBC experiment provided the information to link the $^{1}H_{\delta2}$ protons to the $^{15}N_{\epsilon2}$ and $^{15}N_{\delta1}$, and accordingly, $^{1}H_{\epsilon1}$ nuclei. These assignments were verified by a weak four bond $^{1}H_{\epsilon2}$ - $^{1}H_{\epsilon1}$ coupling that is often detected in a DQF-COSY or TOCSY spectrum (Wüthrich 1986). Finally, the $^{13}C_{\epsilon1}$ assignments were obtained from alignment with the constant time ^{1}H-^{13}C HSQC spectrum. For a slowly exchanging labile proton on $^{15}N_{\epsilon2}$ (such as the buried His-128 in Ets-1), the $^{1}H_{\epsilon2}$ assignment was made from a ^{1}H-^{15}N HSQC recorded without perturbation of the water magnetization. This $^{1}H_{\epsilon2}$, with a chemical shift downfield of 12 ppm, can also be detected in a one-dimensional ^{1}H spectrum of a protein recorded without saturation of water (Markley 1975; Plesniak et al. 1996; Connelly and McIntosh 1998).

© 1998 NRC Canada

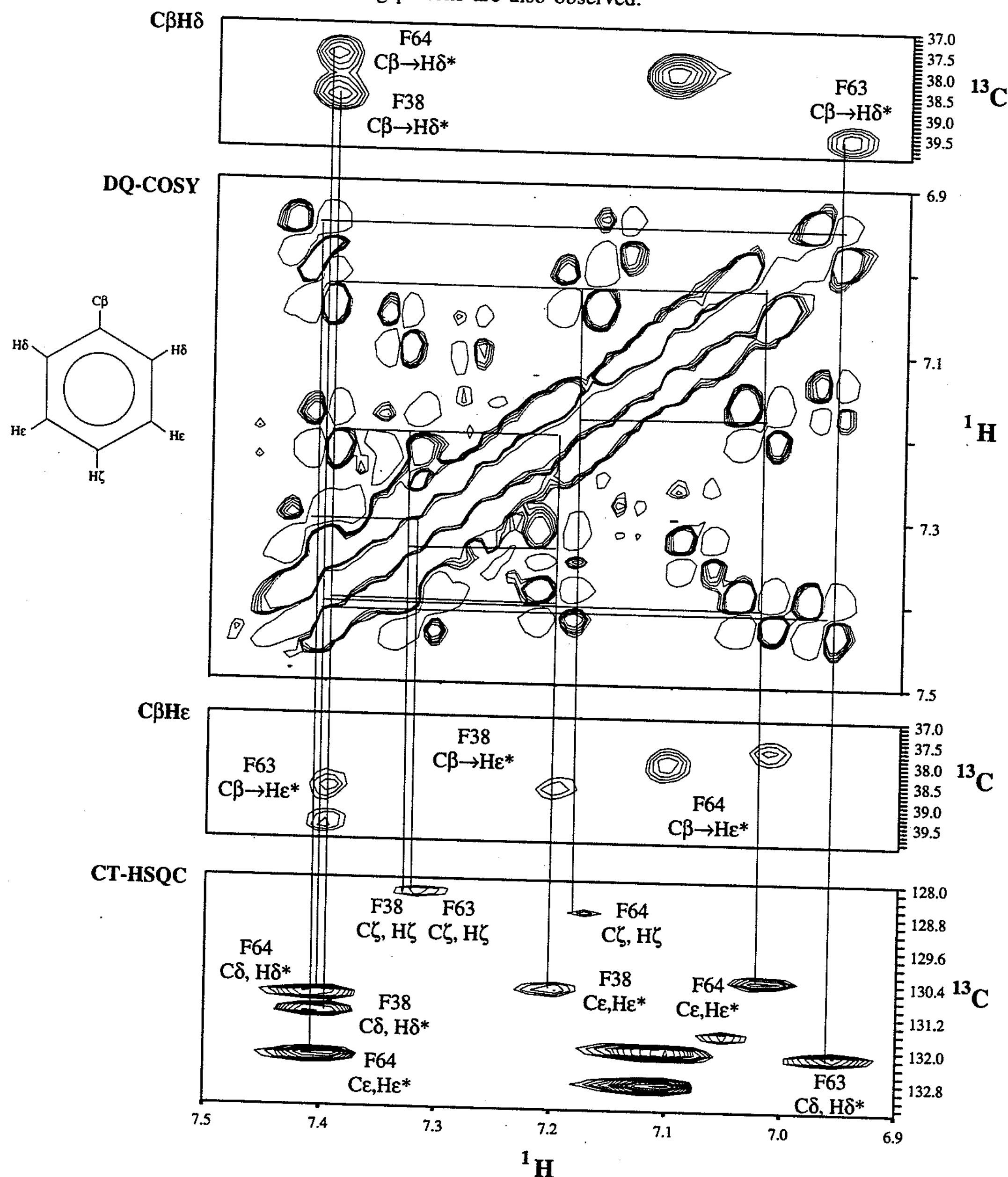
Fig. 4. Assignment of ^1H and ^{13}C resonances from the three tyrosine rings in the GABPα pointed domain. The individual experiments are indicated beside the upper left-hand corner of each box, and the negative peaks in the DQF-COSY spectrum are drawn with one contour. The DQF-COSY spectrum was recorded with a sample of ^2H-[$\gamma^{1,2}$,$\delta^{1,2}$,ζ]-Phe and ^2H-[δ^1,ζ^2,ζ^3,ϵ^3,η^2]-Trp labelled protein in D_2O buffer. This allows for the unambiguous identification of crosspeaks from the tyrosine phenolic protons. The weak peak near 7 ppm arises from an alternate conformation of the GABPα pointed domain.



Information on the ionization state and tautomeric form of each histidine residue in a protein is also provided by the ^1H-^{15}N HMBC experiment (Bachovchin 1986; Pelton et al. 1993; Plesniak et al. 1996). In the fully charged form, it is expected that both $^{15}N_{\epsilon2}$ and $^{15}N_{\delta1}$ nuclei will have chemical shifts clustered between 170 and 180 ppm. In contrast, the shift of the deprotonated ^{15}N in the neutral form of histidine will fall near 250 ppm. In the case of the Ets-1 and GABPα pointed domains, all histidines have ^{15}N shifts near 250 and 168 ppm, indicating that they are in the neutral form at pH 6.5 and 7.2, respectively (Fig. 2). The ^1H-^{15}N HMBC experiment is also used to determine the tautomeric state of the neutral imidazole ring. For the $N_{\delta1}$-H tautomer, it is expected that two strong correlations ($^{15}N_{\epsilon2}$-$^1H_{\epsilon1}$ and $^{15}N_{\epsilon2}$-$^1H_{\delta2}$) will be present near 250 ppm. For the $N_{\epsilon2}$-H tautomer, one strong correlation ($^{15}N_{\delta1}$-$^1H_{\epsilon1}$) at approximately 250 ppm and two slightly weaker correlations ($^{15}N_{\epsilon2}$-$^1H_{\epsilon1}$ and $^{15}N_{\epsilon2}$-$^1H_{\delta2}$) at approximately 168 ppm should be present. In the case of the Ets-1 pointed domain both histidines (His-76 and His-128) are in the $N_{\epsilon2}$-H tautomeric state because only one single strong peak ($^{15}N_{\delta1}$-$^1H_{\epsilon1}$) is present near 250 ppm (Fig. 2). A similar situation is found for the three histidines

© 1998 NRC Canada

Fig. 5. Assignment of ^1H and ^{13}C resonances from the three phenylalanine rings in the GABPα pointed domain. The individual experiments are indicated beside the upper left-hand corner of each box, and the negative peaks in the DQF-COSY spectrum are drawn with one contour. The DQF-COSY spectrum was recorded with an unlabelled sample of protein in D_2O buffer, and thus the signals from the tryptophan and tyrosine aromatic ring protons are also observed.



in the GABPα pointed domain.

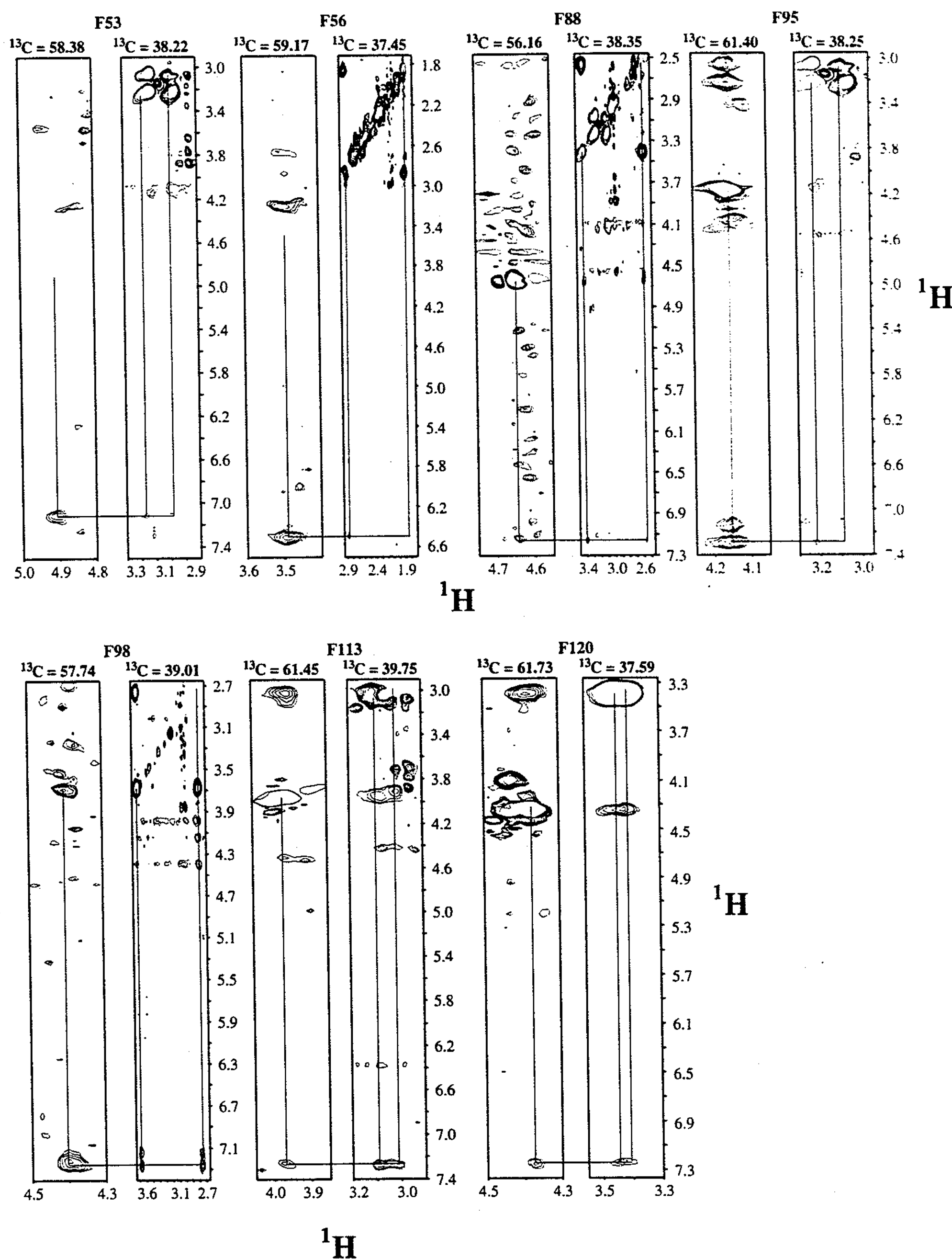
Tryptophan

The assignment of the ^1H, ^{13}C, and ^{15}N resonances from the four tryptophan residues in the Ets-1 pointed domain is illustrated in Fig. 3. From the $C_\beta H_\delta$ experiment, the $^1H_{\delta1}$ assignments were obtained. The corresponding $^{13}C_{\delta1}$ assignments were deduced from the corresponding negative peaks in an aligned constant time ^1H-^{13}C HSQC spectrum. By surveying the two-dimensional ^1H-^{15}N HSQC and HMBC, three-dimensional ^{15}N-edited TOCSY-HSQC and NOESY-HSQC, and two-dimensional homonuclear DQF-COSY and TOCSY experiments recorded in H_2O buffer (data not shown), the $^{15}N_{\epsilon1}$ and $^1H_{\epsilon1}$ resonances were readily assigned by virtue of their scalar and NOE coupling to the $^1H_{\delta1}$ nucleus. Connections to the six-membered ring were provided by NOEs between $^1H_{\epsilon1}$ and $^1H_{\zeta2}$ in a ^{15}N-edited NOESY-HSQC or two-dimensional homonuclear NOESY spectrum recorded in H_2O buffer (Fig. 3*a*). Referring back to the constant time ^1H-^{13}C HSQC, the $C_{\zeta2}$ assignments were obtained by alignment of the ^1H shift (Fig. 3*a*).

To complete the assignment of the tryptophan ring $H_{\zeta2}$, $H_{\eta2}$, $H_{\zeta3}$, and $H_{\epsilon3}$ resonances and DQF-COSY and TOCSY spectra were recorded with a protein sample in D_2O (Fig. 3*b*). To simplify this task, we biosynthetically

© 1998 NRC Canada

Fig. 6. The assignments of the $^1H_{\delta*}$ resonances from the seven phenylalanine residues in the Ets-1 pointed domain were confirmed on the basis of the observation of NOEs from the corresponding $^1H_\alpha$ and $^1H_{\beta\beta'}$ protons. Shown are strips from a three-dimensional ^{13}C-edited NOESY-HSQC spectrum taken at the $^{13}C_\alpha$ (left) and $^{13}C_\beta$ (right) frequencies of the phenylalanines. The carrier frequency was set at 67 ppm to allow simultaneous excitation of aromatic and aliphatic carbons.



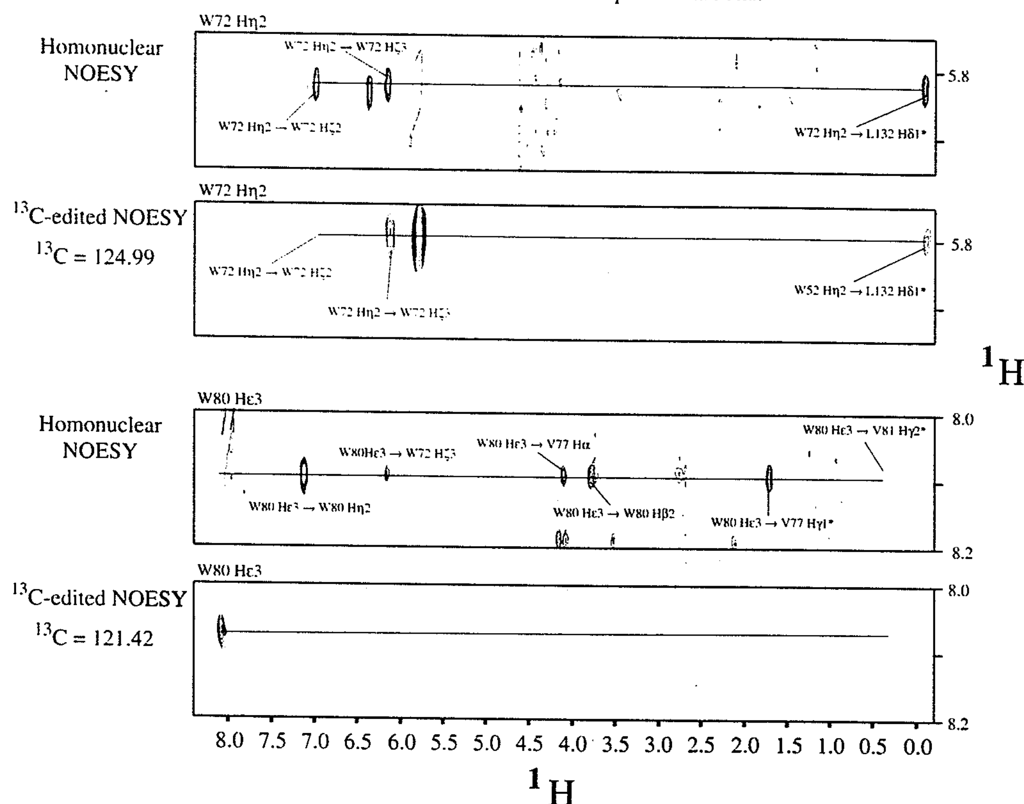
deuterated the phenylalanine residues (the Ets-1 pointed domain does not contain any tyrosine residues). Therefore, only the ring protons from the tryptophan indole and histidine imidazole rings were observed in these two-dimensional correlation spectra. Lastly, the corresponding ^{13}C assignments were extracted from the constant time ^1H-^{13}C HSQC and confirmed by inspection of a three-dimensional ^{13}C-edited NOESY-HSQC spectrum (data not shown).

Tyrosine

The assignment of the ^1H and ^{13}C resonances from the three tyrosine rings in the GABPα pointed domain is illustrated in Fig. 4. Starting again with the $C_\beta H_\delta$ experiment, the $^1H_{\delta*}$ from Tyr-83 and Tyr-17 were assigned. The chemical shifts of their corresponding $^{13}C_{\delta*}$ partners were obtained from the constant time ^1H-^{13}C HSQC spectrum. The $^1H_{\delta*}$ resonance of Tyr-6 was not detected in the $C_\beta H_\delta$ spectrum. The $^1H_{\epsilon*}$ protons were easily identified in a DQF-COSY spectrum of the GABPα pointed domain in which the phenylalanine and tryptophan residues were specifically deuterated. This eliminated confusion owing to overlap from the signals of these aromatic amino acids. The $^1H_{\epsilon*}$ were also identified in a $C_\beta H_\epsilon$ experiment and their corresponding attached carbons were identified in the constant

© 1998 NRC Canada

Fig. 7. A two-dimensional homonuclear NOESY spectrum of the unlabelled Ets-1 pointed domain in D_2O buffer provides significantly greater signal-to-noise and resolution than a three-dimensional ^{13}C-edited NOESY-HSQC spectrum of the ^{13}C/^{15}N labeled protein in H_2O. Shown are comparisons of the NOE data for two tryptophan protons provided by these experimental approaches (τ_{mix} = 75 ms). The ^{13}C carrier was set at 67 ppm for excitation of both the aromatic and aliphatic carbons.



time ^{1}H-^{13}C HSQC spectrum. The tyrosine rings in the GABPα pointed domain undergo fast flipping on the chemical shift time scale and thus exhibit degenerate chemical shifts for the δ1 and δ2 carbons–protons, and the ε1 and ε2 carbons–protons. A possible exception is Tyr-6, for which the unidentified $^{1}H_{\delta*}$ resonances may be absent owing to intermediate exchange broadening.

Phenylalanine

The final, and perhaps most difficult, task is the assignment of the resonances from the phenylalanine ring. Strong ^{13}C-^{13}C J couplings combined with limited shift dispersion (the non-first order AB spin system) distort peaks from these residues in a regular or constant time ^{1}H-^{13}C HSQC spectrum, making their resolution difficult. Figure 5 illustrates the assignment of the ^{1}H and ^{13}C resonances from the three phenylalanine residues (Phe-38, -63, and -64) in the GABPα pointed domain. Starting with the $C_\beta H_\delta$ experiment, the $^{1}H_{\delta*}$ protons were identified. Using a DQF-COSY in conjunction with a TOCSY spectrum (data not shown), the assignments of the $^{1}H_{\delta*}$, $^{1}H_{\epsilon*}$, and H_ζ were made. The $^{1}H_{\epsilon*}$ assignments were confirmed using the $C_\beta H_\epsilon$ experiment, and the $^{13}C_{\epsilon*}$ and $^{13}C_\zeta$ assignments extracted from the constant time ^{1}H-^{13}C HSQC. All of the phenylalanine rings in the GABPα and Ets-1 pointed domains undergo fast flipping on the chemical shift time scale.

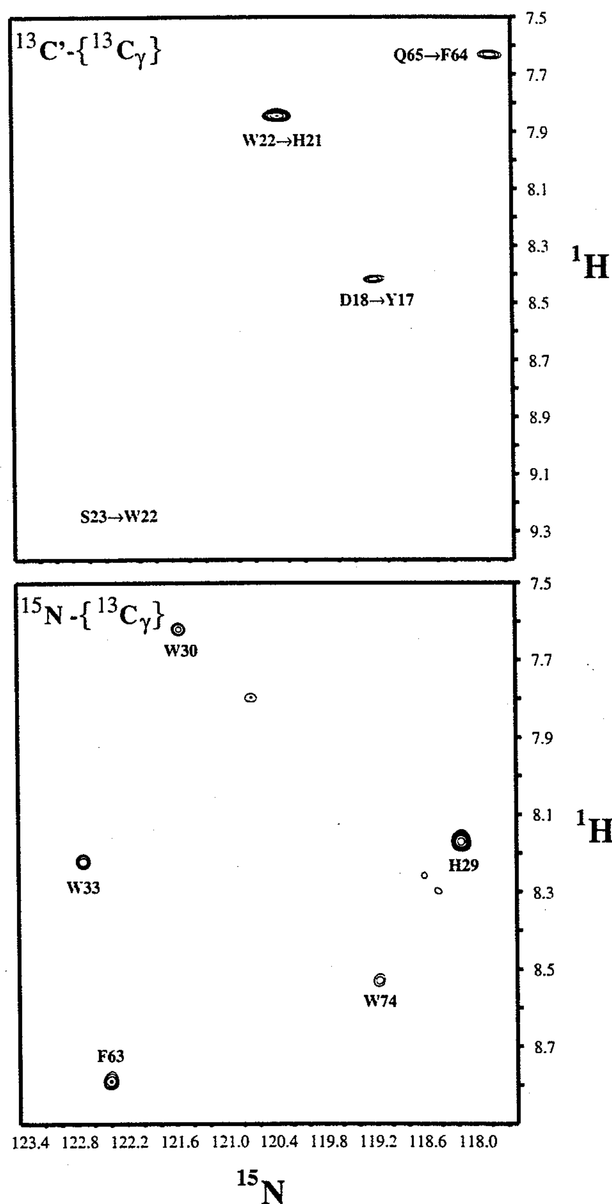
The assignment of the ^{1}H and ^{13}C resonances from the seven phenylalanine rings in the Ets-1 pointed domain was relatively difficult owing to their poor dispersion in the ^{1}H dimension (Fig. 1). The $^{1}H_{\delta*}$ assignments, obtained from the $C_\beta H_\delta$ experiment, were confirmed on the basis of the observation of intraresidue NOEs to the $^{1}H_\alpha$ and $^{1}H_{\beta\beta'}$ resonances resolved in a three-dimensional ^{13}C-edited NOESY-HSQC spectrum (Fig. 6). In four cases (Phe-53, -56, -88, and -98), the $^{13}C_{\delta*}$ assignments could be deduced by alignment to the ^{1}H-^{13}C constant time (CT)-HSQC spectrum (Fig. 1), and by the observation of symmetry NOEs between the aliphatic and aromatic protons in the ^{13}C-edited NOESY spectrum. Using DQF-COSY and TOCSY spectra, recorded with an unlabelled sample of the Ets-1 pointed domain, selected $^{13}C_{\epsilon*}/^{1}H_{\epsilon*}$ (Phe-56, -98, -113, -120) and $^{13}C_\zeta/^{1}H_\zeta$ (Phe-56, -98, -112) protons and carbons were assigned. Although it may have been possible to obtain more complete assignments of the phenylalanine spin systems by preparing a sample of this protein with selectively deuterated tryptophan indole rings, the severe overlap of the signals from these residues would have still precluded the resolution of NOEs needed for structural analysis.

Structural restraints

The determination of the structure of a protein by NMR spectroscopy is dependent primarily on the measurement of three-bond scalar couplings and NOE interactions to provide dihedral angle and interproton distance restraints, respectively. Specific considerations of these measurements in the case of aromatic residues are discussed in the following two sections.

© 1998 NRC Canada

Fig. 8. Measurement of the χ_1 dihedral angles of aromatic residues in the GABPα pointed domain by qualitative analysis of the $^{13}C'$-{$^{13}C_\gamma$} and ^{15}N-{$^{13}C_\gamma$} spin-echo experiments according to a staggered rotamer model. In the $^{13}C'$-{$^{13}C_\gamma$} difference spectrum, a peak at the frequency of the amide ^1H-^{15}N of residue (i + 1) indicates that the aromatic residue (i) has a χ_1 angle of –60°. In the ^{15}N-{$^{13}C_\gamma$} experiment, a peak at the amide frequency of aromatic residue (i) indicates that it has a χ_1 angle of 180°.



Dihedral angle restraints

Expanding upon the elegant idea of quantitative J-correlation spectroscopy (Bax et al. 1994), Hu et al. (1997) developed the two-dimensional $^{13}C'$-{$^{13}C_\gamma$} and ^{15}N-{$^{13}C_\gamma$} spin echo experiments to provide the $^3J_{C'C\gamma}$ and $^3J_{NC\gamma}$ coupling constants for aromatic residues, respectively. These experiments also aid in the assignment of the ^1H-^{15}N spectrum of a protein. When analyzed according to a staggered rotamer model, if a ^1H-^{15}N peak from residue (i + 1) is present in the $^{13}C'$-{$^{13}C_\gamma$} difference spectrum, then the χ_1 of the preceding aromatic residue (i) is –60°. Conversely, if a peak from the amide of aromatic amino acid (i) is observed in the ^{15}N-{$^{13}C_\gamma$} difference spectrum, then the χ_1 of that residue is 180°. In the case of the GABPα pointed domain, Tyr-17, His-21, Trp-22, and Phe-64 were found to have a χ_1 of –60°, whereas His-29, Trp-30, Trp-33, Phe-63, and Trp-74 have χ_1 angles of 180° (Fig. 8). These dihedral angles were also confirmed by the combined use of the three-dimensional HNHB, short mixing time ^{15}N-edited TOCSY-HSQC, and $^{13}C/^{15}N$-edited NOESY-HSQC experiments (Powers et al. 1993; Clore and Gronenborn 1994).

NOE-based distance restraints

Three- or four-dimensional ^{13}C-edited NOESY spectra are invaluable for the resolution and assignment of interproton NOE interactions (Clore and Gronenborn 1994). However, as illustrated in Fig. 7, few NOEs to aromatic protons are actually detected in the ^{13}C-edited NOESY-HSQC spectrum of a protein such as the Ets-1 pointed domain. This is attributed to a loss in signal owing to ^{13}C-induced dipolar relaxation, peak broadening resulting from strong ^{13}C-^{13}C scalar couplings (~60 Hz), and the difficulty of uniformly exciting the full ^{13}C spectrum of a protein as required to record simultaneously the NOE interactions between all aromatic and aliphatic protons. Note that the latter problem can be offset by the use of broadband adiabatic WURST pulses (Zwahlen et al. 1997), or by recording a ^{13}C-edited NOESY-HSQC centered only about the aromatic carbons.

In contrast, we have found that a standard two-dimensional homonuclear NOESY experiment provides the necessary signal-to-noise ratio and, in combination with selective deuteration, resolution to confidently measure NOE interactions involving aromatic protons (Donaldson et al. 1996; manuscripts in preparation). For example, six strong NOEs to Trp-80 $H_{\epsilon3}$ were observed in the homonuclear NOESY spectrum of the unlabelled Ets-1 pointed domain, whereas none were detected in the ^{13}C-edited version of this experiment recorded with a $^{13}C/^{15}N$ enriched protein (Fig. 7). In addition to aiding with spectral assignments, selective labelling with deuterated aromatic amino acids has proven to be a highly effective method for overcoming potential problems caused by chemical shift degeneracies in the analysis of the homonuclear NOESY spectrum of a protein. Of course, a parallel NOESY spectrum recorded on an unlabelled sample in D_2O buffer is necessary to detect possible NOEs between different types of aromatic amino acids. Combining the information contained in both homonuclear and ^{13}C-edited NOESY-HSQC spectra of the Ets-1 and GABPα pointed domains, 153 and 124 restraints, respectively, involving aromatic protons were measured for the structural determination of these two protein fragments.

In conclusion, we have described a strategy for the assignment of the ^1H, ^{13}C, and ^{15}N resonances from the aromatic residues in proteins using a suite of homo- and hetero-nuclear scalar and NOE correlation experiments, combined with selective deuterium labelling. This strategy also allows for the measurement of dihedral angle and NOE-based distance restraints for aromatic amino acids, as required to increase the accuracy and precision of protein structure determination by NMR methods.

© 1998 NRC Canada

Acknowledgments

The authors are grateful to Dr. Lewis Kay and Dr. Jack Skalicky for providing invaluable NMR advice and many of the pulse sequences used in this article. This research was supported by grants from the National Cancer Institute of Canada with funds from the Canadian Cancer Society (to L.P.M.), the Leukemia Research Fund of Canada (to C.M.S.), and the Jane Coffin Child's Memorial Fund for Medical Research (to L.N.G.). Instrument support was provided by the Protein Engineering Network Centres of Excellence (to L.P.M.).

References

Bachovchin, W.W. 1986. ^{15}N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. Biochemistry, **25**: 7751–7759.

Bax, A., and Marion, D. 1988. Improved resolution and sensitivity in ^{1}H-detected multiple-bond correlation spectroscopy. J. Magn. Reson. **78**: 186–191.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R., and Zhu, G. 1994. Measurement of homo- and heteronuclear J couplings from quantitative J correlation. Methods Enzymol. **239**: 79–105.

Clore, G.M., and Gronenborn, A.M. 1994. Structures of larger proteins, protein–ligand complexes, and protein–DNA complexes by multidimensional heteronuclear NMR. Protein Sci. **3**: 372–390.

Connelly, G.P., and McIntosh, L.P. 1998. Characterization of a buried neutral histidine in *Bacillus circulans* xylanase: internal dynamics and interaction with a bound water molecule. Biochemistry, **37**: 1810–1818.

Creighton, T. E. 1993. Proteins: structures and molecular properties. 2nd ed. W.H. Freeman & Company, New York.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR, **6**: 277–293.

Donaldson, L.W., Petersen, J.M., Graves, B.J., and McIntosh, L.P. 1996. Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. EMBO J. **15**: 125–134.

Garrett, D.S., Powers, R., Gronenborn, A.M., and Clore, G.M. 1991. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. J. Magn. Reson. **95**: 214–220.

Graves, B.J., and Petersen, J.M. 1997. Specificity within the Ets family of transcription factors. *In* Advances in cancer research. *Edited by* George Vande Woude and George Klein. Academic Press, New York.

Grzesiek, S., and Bax, A. 1995. Audi-frequency NMR in the nutating frame. Application to the assignment of phenylalanine residues in isotopically enriched proteins. J. Am. Chem. Soc. **117**: 6527–6531.

Gunther, C.V., Nye, J.A., Bryner, R.S., and Graves, B.J. 1990. Sequence-specific DNA binding of the proto-oncoprotein ets-1 defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. Genes Dev. **4**: 667–679.

Hu, J.S., Grzesiek, S., and Bax, A. 1997. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. J. Am. Chem. Soc. **119**: 1803–1804.

Johnson, P.E., Tomme, P., Joshi, M.D., and McIntosh, L.P. 1996. Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 2. NMR and ultraviolet absorption spectroscopy. Biochemistry, **35**: 13 895 – 13 906.

Löhr, F., and Rüterjans, H. 1996. Novel pulse sequences for the resonance assignment of aromatic side chains in ^{13}C-labeled proteins. J. Magn. Reson. **112**: 259–268.

Markley, J.L. 1975. Observation of histidine residues in proteins by means of NMR spectroscopy. Acc. Chem. Res. **8**: 70–80.

McIntosh, L.P., and Dahlquist, F.W. 1990. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q. Rev. Biophys. **23**: 1–38.

McIntosh, L.P., Wand, A.J., Lowry, D.F., Redfield, A.G., and Dahlquist, F.W. 1990. Assignment of the backbone ^{1}H and ^{15}N NMR resonances of bacteriophage T4 lysozyme. Biochemistry, **29**: 6341–6362.

Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. 1994. Simultaneous acquisition of ^{15}N and ^{13}C-edited NOE spectra of proteins dissolved in H_2O. J. Magn. Reson. **103**: 197–201.

Pelton, J.G., Torchia, D.A., Meadow, N.D., and Roseman, S. 1993. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. Protein Sci. **2**: 543–558.

Plesniak, L.A., Connelly, G.P., Wakarchuk, W.W., and McIntosh, L.P. 1996. Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignments, pH titration and hydrogen exchange. Protein Sci. **5**: 2319–2328.

Powers, R., Garrett, D., March, C., Frieden, E., Gronenborn, A., and Clore, G. 1993. The high-resolution three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Biochemistry, **32**: 6744–6762.

Richarz, R., and Wüthrich, K. 1978. Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers, **17**: 2133–2141.

Santoro, J., and King, G.C. 1992. A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. J. Magn. Reson. **97**: 202–207.

Vuister, G.W., and Bax, A. 1992. Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling. J. Magn. Reson. **98**: 428–435.

Vuister, G.W., Kim, S.J., and Bax, A. 1994. 2D and 3D NMR study of phenylalanine residues in proteins by reverse isotopic labeling. J. Am. Chem. Soc. **116**: 9206–9210.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. 1995*a*. ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR, **6**: 135–140.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995*b*. ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigation of nearest-neighbor effects. J. Biomol. NMR, **5**: 67–81.

Wüthrich, K. 1986. NMR of proteins and nucleic acids. John Wiley & Sons, New York.

Yamazaki, T., Forman-Kay, J.D., and Kay, L. 1993. Two-dimensional NMR experiments for correlating $^{13}C_\beta$ and $^{1}H_{\delta/\epsilon}$

© 1998 NRC Canada

chemical shifts of aromatic residues in ^{13}C-labelled proteins via scalar couplings. J. Am. Chem. Soc. **115**: 11 054 – 11 055.

Yu, L., and Fesik, S.W. 1994. pH titration of the histidine residues of cyclophilin and FK506 binding protein in the absence and presence of immunosuppressant ligands. Biochim. Biophys. Acta, **1209**: 24–32.

Zerbe, O., Szyperski, T., Ottiger, M., and Wüthrich, K. 1996. Three-dimensional ^{1}H-TOCSY-relayed ct[^{13}C-^{1}H]-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins. J. Biomol. NMR, **7**: 99–106.

Zhang, O., Kay, L.E., Olivier, J.P., and Forman-Kay, J.D. 1994. Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domains of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J. Biomol. NMR, **4**: 845–858.

Zwahlen, C., Legault, P., Vincent, S.F.J., Greenblatt, J., Konrat, R., and Kay, L.E. 1997. Methods for measurement of intermolecular NOE's by multi-nuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. J. Am. Chem. Soc. **119**: 6711–6721.

© 1998 NRC Canada