Journal of Biomolecular NMR, **13:** 381–385, 1999.
KLUWER/ESCOM

A novel NMR experiment for the sequential assignment of proline residues and proline stretches in $^{13}C/^{15}N$-labeled proteins

M.J. Bottomley[#], M.J. Macias[#], Z. Liu & M. Sattler*
European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Received 1 December 1998; Accepted 19 January 1999

Key words: proline, proline-rich, selective pulses, sequential assignment, triple resonance

Abstract

A new pulse sequence is described for the sequential assignment of proline residues in $^{13}C/^{15}N$-labeled proteins by correlating C^{δ} and C^{α} chemical shifts of proline residues with the H^{α} chemical shift of the preceding residue. Notably, the experiment can provide the sequential connectivities in poly-proline stretches, which cannot be determined using standard triple resonance experiments. Excellent solvent suppression is achieved by coherence selection via a heteronuclear gradient echo. The new pulse sequence has been successfully applied to the 11 kDa HRDC domain.

Proline residues play important structural and functional roles in proteins. For example, due to the cyclic side chain, the conformational space of proline residues is restricted compared to other amino acids. Furthermore, proline residues cannot act as hydrogen bond donors due to the lack of an amide proton. As a result of these properties, prolines often induce helix bending or are part of tight turns in three-dimensional (3D) structures of proteins (Williamson, 1994; Vanhoof et al., 1995). Proline-rich sequences have a high tendency to adopt a left-handed poly-proline type II helix (Adzhubei and Sternberg, 1993), and many proteins containing poly-proline sequences have been shown to play important roles in cellular signaling (Alexandropoulos et al., 1995; Chan et al., 1996; Chen et al., 1997; Bedford et al., 1998). Several interesting interactions of such peptides with SH3 or WW protein domains have been structurally characterized (Feng et al., 1994; Yu et al., 1994; Macias et al., 1996; Andreotti et al., 1997).

The determination of the 3D solution structures of complexes between proline-rich peptides and their binding partners requires the complete resonance assignment of both species. Standard heteronuclear NMR techniques rely on correlating the H^N chemical shift of each residue to the C^{α} and C^{β} chemical shifts of the same and the preceding residue using, for example, HNCA, CBCA(CO)NH and CBCANH experiments (Clore and Gronenborn, 1998; Sattler et al., 1999). Therefore, these experiments cannot be used to sequentially connect proline residues to the preceding residue or to assign stretches of prolines. As a consequence, the sequential assignment for proline residues has to rely on the observation of NOEs to neighboring residues. However, this is not desirable because the conformational dependence of NOEs may introduce misassignments.

For the identification of residues preceding a proline, an elegant proline-edited HACA(CO)(N) experiment has been proposed (Olejniczak and Fesik, 1994). In this experiment, the H^{α} and C^{α} signals of residues preceding prolines are observed exclusively, based on the absence of an amide proton. However, as a potential drawback, this experiment does not provide any chemical shift information for the proline residue. Thus, although the chemical shifts of neighboring residues (even prolines) are obtained, the sequential connectivity cannot be established.

Here we describe a novel NMR experiment, CDCA(NCO)CAHA, which is designed for the sequential assignment of proline residues and especially

# These authors contributed equally.
*To whom correspondence should be addressed. E-mail: sattler@EMBL-Heidelberg.de

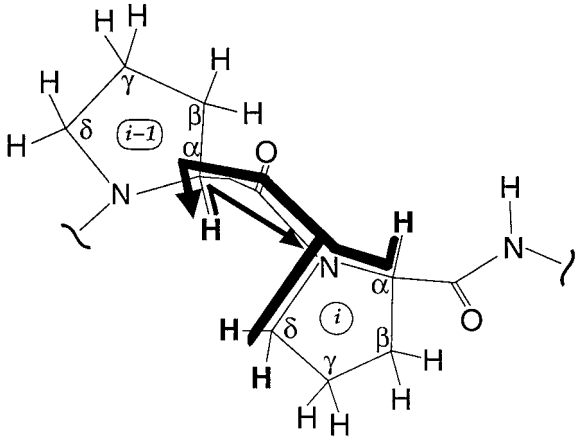

Figure 1. The spin system of two neighboring proline residues. The magnetization transfer pathways observed with the new pulse sequence presented are indicated with arrows. The thick arrow represents pathways utilizing the $^1J(C^{\alpha/\delta},N)$ coupling, whereas the thin arrow represents the less efficient pathway via $^2J(C^{\alpha},N)$. Protons directly involved in the transfer pathways are shown in bold font.

proline-rich sequences in ^{13}C,^{15}N-labeled proteins. The new pulse sequence correlates the $C^{\alpha}(i)$ and $C^{\delta}(i)$ chemical shifts of proline (i) to the $H^{\alpha}(i-1)$ and $C^{\alpha}(i-1)$ chemical shifts of the preceding residue (i−1) via ^1J and ^2J couplings. Since H^{α} rather than H^{N} chemical shifts are observed, even neighboring proline residues can be sequentially assigned. If the experiment is applied to an H_2O solution of the protein, signals originating from non-proline residues can be suppressed by the use of a purging scheme.

In case of a spin system where residue (i) is a proline (Figure 1), the CDCA(NCO)CAHA experiment employs the following magnetization transfer pathways:

$$H^{\delta}(i) \rightarrow C^{\delta}(i) \xrightarrow{^1J(C^{\delta},N)} N(i) \rightarrow C'(i-1) \rightarrow C^{\alpha}(i-1) \rightarrow H^{\alpha}(i-1) \quad (1a)$$

$$H^{\alpha}(i) \rightarrow C^{\alpha}(i) \xrightarrow{^1J(C^{\alpha},N)} N(i) \rightarrow C'(i-1) \rightarrow C^{\alpha}(i-1) \rightarrow H^{\alpha}(i-1) \quad (1b)$$

$$H^{\alpha}(i-1) \rightarrow C^{\alpha}(i-1) \xrightarrow{^2J(C^{\alpha},N)} N(i) \rightarrow C'(i-1) \rightarrow C^{\alpha}(i-1) \rightarrow H^{\alpha}(i-1) \quad (1c)$$

The pulse sequence for the CDCA(NCO)CAHA experiment (Figure 2) is briefly described below. During the initial INEPT step (point a to b in Figure 2) magnetization is transferred from H^{δ} and H^{α} protons to the directly bound carbon spins. This transfer is followed by a constant-time delay 2T (point b to c) to create carbon/nitrogen antiphase coherence ($2N_yC_z^{\alpha}$ and $2N_yC_z^{\delta}$) via $^1J(C^{\alpha/\delta},N)$ and $^2J(C^{\alpha},N)$ couplings. The delay 2T is also used for constant-time evolution of $^{13}C^{\alpha/\delta}$ chemical shifts during t_1. This gives rise to three cross peaks corresponding to the magnetization transfer pathways given in Equations 1a–c and indicated in Figure 1.

During the delay $2\tau_N$ (points c to d) the carbon/nitrogen antiphase coherence ($2N_yC_z^{\alpha/\delta}$) is converted into nitrogen carbonyl antiphase coherence ($2N_yC_z'$) utilizing the $^1J(N,C')$ coupling. Consecutively, magnetization is then transferred from $C' \rightarrow C^{\alpha} \rightarrow H^{\alpha}$ via INEPT steps. For the constant-time evolution of C^{α} chemical shifts during t_2 (point e to f) a $^{13}C^{\alpha}$-selective 180° pulse (phase ϕ_7) is used in order to refocus the $^1J(C^{\alpha},C^{\beta})$ couplings (see below). The final $C^{\alpha} \rightarrow H^{\alpha}$ transfer is implemented as COS-HSQC (Cavanagh et al., 1991; Kay et al., 1992) optimized for the $C^{\alpha} \rightarrow H^{\alpha}$ spin system (Schleucher et al., 1994). When the latter is combined with a heteronuclear gradient echo, excellent suppression of the solvent signal is achieved, which is required for the observation of H^{α} resonances at or near the solvent resonance in H_2O solution.

The sensitivity of the CDCA(NCO)CAHA experiment is mainly determined by the transverse relaxation of $^{13}C^{\alpha}$ and $^{13}C^{\delta}$ magnetization which is active during the delays 2T and 2δ. In order to optimize the signal-to-noise, $^{13}C^{\alpha/\delta}$ selective 180° pulses are applied during these delays (pulses shown as unfilled boxes in Figure 2) as has been described previously (McCoy, 1995; Matsuo et al., 1996). For example, a $^{13}C^{\alpha/\delta}$ selective 180° pulse is applied between b and c in order to refocus the $^1J(C^{\alpha},C^{\beta})$ and $^1J(C^{\delta},C^{\gamma})$ couplings (180° ^{13}C pulse, phase ϕ_3 in Figure 2). This allows the delay 2T to be optimized with respect to the coherence transfer efficiency ($C^{\alpha/\delta} \rightarrow N$) and the transverse relaxation of the $^{13}C^{\alpha/\delta}$ spins. Especially for the short, transverse relaxation times of ^{13}C spins, the signal-to-noise for this implementation is greatly improved because a shorter delay 2T can be used. For $^{13}C^{\alpha/\delta}$ $T_2 = 15$ ms the implementation using the selective pulse ($2T = 14$ ms) yields >50% better signal-to-noise than the version with a non-selective 180° pulse, where the delay should be set to $2T \approx 28$ ms in order to refocus the homonuclear $^1J(C,C)$ couplings. Note that when using a $^{13}C^{\alpha/\delta}$ selective pulse the t_1 FID is modulated by $\cos(\pi\, ^1J(C,C)\, t_1)$. Thus, for $t_1^{max} = 2T \approx 1/(2^1J(C,C))$, the FID is enveloped by a cosine window function and additional apodization of the time-domain data is not required during data processing. Based

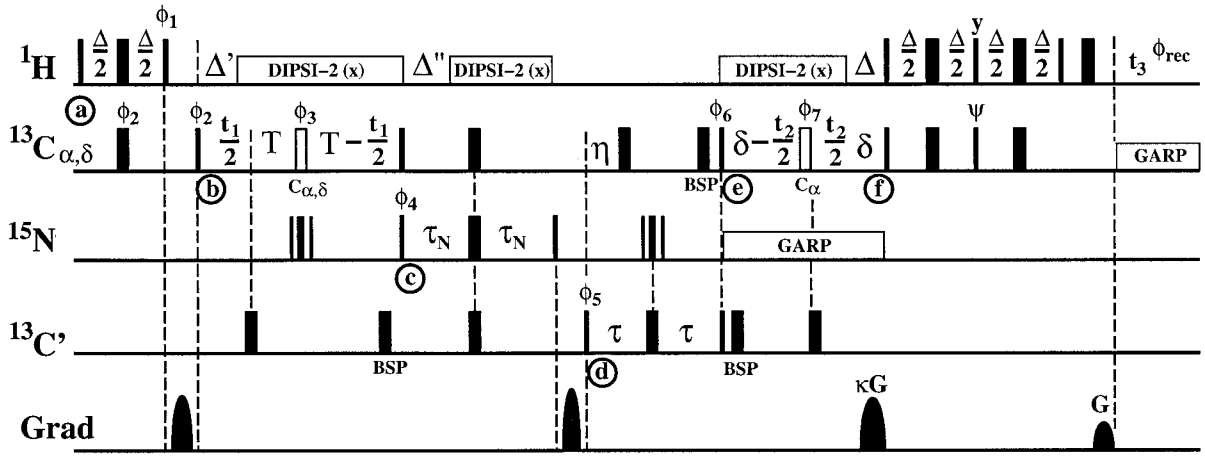

Figure 2. Pulse sequence for the proline-edited CDCA(NCO)CAHA experiment. ^{13}C 90°/180° pulses are applied as shaped G4/G3 pulses (Emsley and Bodenhausen, 1990) of 400/250 μs duration, respectively. The pulses shown as unfilled boxes correspond to $^{13}C^{\alpha/\delta}$ selective 180° pulses, implemented as 1.25 ms RE-BURP pulses (Geen and Freeman, 1991) (256 points) for a ^{13}C frequency of 150 MHz. Pulses applied to compensate for Bloch-Siegert phase shifts are denoted 'BSP'. All other pulses are rectangular. Carrier offsets are 4.7, 118, 56 and 175 ppm for ^{1}H, ^{15}N, $^{13}C^{\alpha/\delta}$ and ^{13}C′, respectively. Delay durations are: $\Delta = 3.5$ ms, $\Delta' = 2.0$ ms, $\Delta'' = 5.4$ ms, $T = 7$ ms, $\tau_N = \tau = 14$ ms, $\eta = 4.5$ ms, $\delta = 3.5$ ms. Composite-pulse decoupling fields are applied with strengths of 3.125 kHz, 2.5 kHz and 1.25 kHz for ^{1}H, ^{13}C and ^{15}N, respectively. The gradient amplitudes and durations are g1 = 1.4 ms, 40.0 G/cm; g2 = 1.4 ms, 35.0 G/cm; g3 = 1.4 ms, ± 45.0 G/cm and g4 = 350 μs, 45.2 G/cm. For each t_2 value the gradient g3 and phase ψ are inverted and the corresponding FIDs are stored separately in order to select for echo- and antiecho pathways (Kay et al., 1992; Schleucher et al., 1993). The phase cycling is: $\psi = \pm y$; $\phi_1 = y$; $\phi_2 = x + TPPI(t_1)$; $\phi_3 = 4(x), 4(y), 4(-x), 4(-y)$; $\phi_4 = x, -x$; $\phi_5 = 2(x), 2(-x)$; $\phi_6 = x$; $\phi_7 = 4(x), 4(-x), 4(y), 4(-y)$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x, x, 2(x), -x, x, 2(-x), x$. The pulse program in Bruker DRX format is available at http://www.NMR.EMBL-Heidelberg.DE/sattler.

on the transfer amplitudes between points b and d (assuming $T_2(^{13}C^{\alpha/\delta}) = 20$ ms, $T_2(^{15}N) = 80$ ms, $^1J(C^{\alpha/\delta}, N) = 11$ Hz, $^2J(C^\alpha, N) = 7$ Hz, $2T = 14$ ms, $2\tau_N = 28$ ms), the relative signal-to-noise for the cross peaks involving the C^α, C^δ and $C^\alpha(i-1)$ spins is 1, 0.8 and 0.3, respectively.

In order to edit for proline residues the pulse sequence has to be applied to an H_2O solution of the protein. Due to the evolution of ^{1}J(H,N) couplings, transverse magnetization of non-proline nitrogen spins will be antiphase with respect to the directly bound amide proton at the end of the delay $\Delta'' = 1/(2^1J(H, N))$. DIPSI-2 decoupling following Δ'' will then eliminate all antiphase product operators $2N_{x,y}H_z$ originating from non-proline residues. Alternatively, the pulse sequence could be applied to a sample in D_2O. In this case, the proton decoupling should be continued during Δ'', and ^{2}H-decoupling should be applied between points c and d in order to suppress quadrupolar relaxation of the amide deuteron spins during $2\tau_N$ (Wang et al., 1995). With a pulse sequence modified accordingly, additional cross peaks would be observed for non-proline residues.

The CDCA(NCO)CAHA experiment is applicable to small and medium-sized proteins with sufficiently long $^{13}C^{\alpha,\delta}$ T_2 relaxation times (>15 ms). In principle, the sensitivity of the experiment could be optimized for detection of cross peaks involving only C^α spins by setting $\Delta' = 3.8$ ms and utilizing multiple quantum line narrowing by evolution of C^α, H^α double- and zero-quantum coherences (Grzesiek and Bax, 1995; Grzesiek et al., 1995) during the $C^{\alpha/\delta} \rightarrow N$ transfer (point b to c). However, cross peaks from methylene carbons are then suppressed and the C^δ chemical shifts cannot be obtained. In practice, this is disadvantageous because the observation of at least two cross peaks, $C^\alpha(i)$ and $C^\delta(i)$, is helpful for identifying the proline spin system.

We have applied the novel pulse sequence to the HRDC domain of the yeast helicase Sgs1p (11 kDa) (Morozov et al. 1997). The HRDC domain contains five proline residues, all of which are observed in the 3D version of the pulse sequence (Figure 2) within 36 h experimental time. The proline spin systems can be readily identified as pairs of C^δ and C^α chemical shifts centered around 48 ppm and 60 ppm, respectively (Figure 3). The sequential connectivity between the neighboring residues Pro 35 and Pro 36 is established based on the C^α chemical shift of Pro 35 which is observed at the $C^\alpha(i)$ cross peak in the Pro 35 strip and at the $C^\alpha(i-1)$ cross peak in the Pro 36 strip. The $C^\alpha(i-1)$ chemical shift can be extracted from the

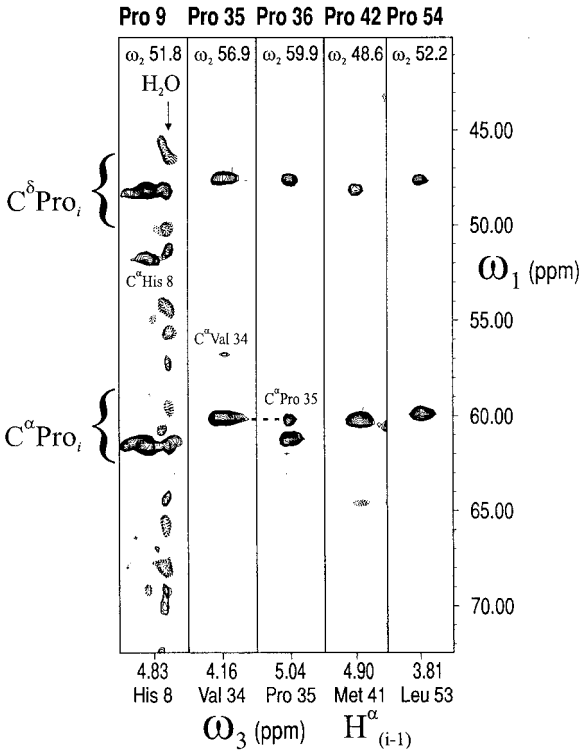

Figure 3. The sequential assignment of the five proline residues in the HRDC domain of the Sgs1p protein using the new pulse sequence. The strips are labeled at the top by the corresponding proline residue (i) and the ω_2 chemical shift (in ppm) of the $C^{\alpha}(i-1)$ plane from which the strip was taken. The name of the $(i-1)$ residue and the corresponding $H^{\alpha}(i-1)$ chemical shift which is observed in ω_3 are given at the bottom of each strip. For Pro 9, Pro 35 and Pro 36 the diagonal cross peak involving the $C^{\alpha}(i-1)$ chemical shift is visible. Thus, for these residues the $C^{\alpha}(i-1)$ chemical shift can be observed both from the peak and from the chemical shift of the ω_2 plane itself. Approximate regions containing the proline C^{α} and C^{δ} resonances are indicated on the left. Note the excellent solvent suppression, which is achieved by the use of a heteronuclear gradient echo. This allows for the observation of the Pro 9 cross peaks at the H^{α} chemical shift of His 8 which is close to the water resonance. The 3D experiment was recorded at 295 K for 36 h on a Bruker DRX spectrometer operating at a ^1H frequency of 600 MHz and equipped with a triple-axis gradient unit. The protein sample was prepared to a final concentration of 1.8 mM in a solution of 50 mM NaCl, 0.02% NaN_3, 20 mM sodium phosphate buffered H_2O (10% (v/v) D_2O) at pH 6.5. The pulse program was executed with 32 scans per increment and 32, 28 and 512 complex points were recorded in the $C^{\alpha/\delta}$ (t_1), the C^{α} (t_2) and the H^{α} (t_3) dimensions with acquisition times of 5.7, 6.4 and 64 ms, respectively. Spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed using XEASY (Bartels et al., 1995). For the t_1 FID forward-backward linear prediction was applied. Prior to Fourier transformation, time-domain data were multiplied by a 90° shifted, sine-squared shaped window function and zero-filled to final sizes of 128, 128 and 1024 points for the processed data in the ω_1, ω_2 and ω_3 dimensions, respectively.

corresponding cross peak in the Pro 36 strip or from the $^{13}C^{\alpha}$ chemical shift of the $H^{\alpha}(i-1)/C^{\alpha}(i-1)$ plane where the Pro 36 strip is taken from. In principle, the sequential connectivity can also be established from a 2D version of the experiment with chemical shift evolution during t_1 and t_3. However, as discussed above, the signal-to-noise of the $C^{\alpha}(i-1)$ cross peak is much lower than the signal-to-noise of the $C^{\alpha/\delta}(i)$ cross peaks. Therefore, since the determination of the $C^{\alpha}(i-1)$ chemical shift is required for the sequential assignment, it is preferable to record a 3D version by evolving the chemical shifts of $C^{\alpha}(i-1)$ during t_2. In any case, recording the 3D rather than the 2D experiment does not lead to a lower signal-to-noise because constant-time chemical shift evolution is employed for the t_2 evolution time and COS-HSQC is used for the $C^{\alpha} \rightarrow H^{\alpha}$ back transfer.

In summary, we have developed a triple resonance pulse sequence which enables the sequential assignment of prolines, and especially of neighboring proline residues. The pulse sequence is applicable to medium-sized proteins and has been successfully applied to the 11 kDa HRDC domain. The experiment will be especially useful for the study of proline-rich sequences, where the standard H^N-based through-bond sequential assignment strategy cannot be employed.

Acknowledgements

This work has been supported by the DFG (SA 823/1-1). M.J.B. is grateful to EMBO for a long-term postdoctoral fellowship. Z.L. acknowledges an EMBL postdoctoral fellowship.

References

Adzhubei, A.A. and Sternberg, M.J. (1993) *J. Mol. Biol.*, **229**, 472–493.

Alexandropoulos, K., Cheng, G. and Baltimore, D. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3110–3114.

Andreotti, A.H., Bunnell, S.C., Feng, S., Berg, L.J. and Schreiber, S.L. (1997) *Nature*, **385**, 93–97.

Bartels, C., Xia, T.-H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Bedford, M.T., Reed, R. and Leder, P. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 10602–10607.

Cavanagh, J., Palmer III, A.G., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **91**, 429–436.

Chan, D.C., Bedford, M.T. and Leder, P. (1996) *EMBO J.*, **15**, 1045–1054.

Chen, H.I., Einbond, A., Kwak, S.J., Linn, H., Koepf, E., Peterson, S., Kelly, J.W. and Sudol, M. (1997) *J. Biol. Chem.*, **272**, 17070–17077.

Clore, G.M. and Gronenborn, A.M. (1998) *Tibtech*, **16**, 22–34.
Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
Emsley, L. and Bodenhausen, G. (1990) *Chem. Phys. Lett.*, **165**, 469–476.
Feng, S., Chen, J.K., Yu, H., Simon, J.A. and Schreiber, S.L. (1994) *Science*, **266**, 1241–1247.
Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.
Grzesiek, S. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 335–339.
Grzesiek, S., Kuboniwa, H., Hinck, A.P. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 5312–5315.
Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. and Oschkinat, H. (1996) *Nature*, **382**, 646–649.
Matsuo, H., Kupce, E., Li, H. and Wagner, G. (1996) *J. Magn. Reson.*, **B113**, 91–96.
McCoy, M. (1995) *J. Magn. Reson.*, **B107**, 270–273.
Morozov, V., Mushegian, A.R., Koonin, E.V. and Bork, P. (1997) *Trends Biochem. Sci.*, **22**, 417–418.
Olejniczak, E.T. and Fesik, S.W. (1994) *J. Am. Chem. Soc.*, **116**, 2215–2216.
Sattler, M., Schleucher, J. and Griesinger, C. (1999) *Prog. NMR Spectrosc.*, in press.
Schleucher, J., Sattler, M. and Griesinger, C. (1993) *Angew. Chem. Int. Ed. Engl.*, **32**, 1489–1491.
Schleucher, J., Schwendinger, M.G., Sattler, M., Schmidt, P., Glaser, S.J., Sørensen, O.W. and Griesinger, C. (1994) *J. Biomol. NMR*, **4**, 301–306.
Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. and Scharpe, S. (1995) *FASEB J.*, **9**, 736–744.
Wang, A.C., Grzesiek, S., Tschudin, R., Lodi, P.J. and Bax, A. (1995) *J. Biomol. NMR*, **5**, 376–382.
Williamson, M.P. (1994) *Biochem. J.*, **297**, 249–260.
Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. and Schreiber, S.L. (1994) *Cell*, **76**, 933–945.