

Article

Automated resonance assignment of proteins: 6D APSY-NMR

Francesco Fiorito[†], Sebastian Hiller[†], Gerhard Wider & Kurt Wüthrich*

Institute for Molecular Biology and Biophysics, ETH Zurich, Schafmattstr. 20, 8093 Zurich, Switzerland

Received 11 January 2006; Accepted 8 March 2006

Key words: 6D APSY-seq-HNCOCANH experiment, automatic backbone NMR assignment, GAPRO software, GARANT software, protein structure determination

Abstract

The 6-dimensional (6D) APSY-seq-HNCOCANH NMR experiment correlates two sequentially neighboring amide moieties in proteins via the C' and C^α nuclei, with efficient suppression of the back transfer from C^α to the originating amide moiety. The automatic analysis of two-dimensional (2D) projections of this 6D experiment with the use of GAPRO (Hiller et al., 2005) provides a high-precision 6D peak list, which permits automated sequential assignments of proteins with the assignment software GARANT (Bartels et al., 1997). The procedure was applied to two proteins, the 63-residue 434-repressor(1–63) and the 115-residue TM1290. For both proteins, complete sequential assignments for all NMR-observable backbone resonances were obtained, and the polypeptide segments thus identified could be unambiguously located in the amino acid sequence. These results demonstrate that APSY-NMR spectroscopy in combination with a suitable assignment algorithm can provide fully automated sequence-specific backbone assignments of small proteins.

Introduction

Projection NMR spectroscopy provides access to high-dimensional NMR experiments which would not, in practice, be available using conventional NMR, due to prohibitively long measurement times even at low digital resolution (Brutscher et al., 1995; Freeman and Kupče, 2004; Kim and Szyperski, 2004). In higher-dimensional experiments, peak overlap is reduced, and the information content increases, enabling more efficient and reliable data analysis (Szyperski et al., 2002). In the context of the present work it is of particular interest that higher dimensionality enables longer-range correlations to be established along poly-

peptide chains. Automated projection spectroscopy (APSY) data sets analyzed with the algorithm GAPRO (Hiller et al., 2005) enable the generation of accurate and virtually artifact-free peak lists of high-dimensional NMR experiments from low-dimensional projection spectra, without the need of reconstruction of the high-dimensional spectrum. Here, such high-quality peak lists have been used to achieve fully automated resonance assignment of proteins, by combining an appropriate APSY-NMR experiment with a suitable automatic assignment algorithm.

The 6D APSY-seq-HNCOCANH experiment connects two sequentially neighboring amide moieties in polypeptide chains via the carbonyl and C^α atoms, with the back transfer from C^α to the originating amide moiety suppressed in this sequential (seq) experiment (Figure 1). Thus, each peak from the 6D peak list generated by GAPRO correlates

*To whom correspondence should be addressed. Email: kurt.wuthrich@mol.biol.ethz.ch.

[†]Both authors contributed equally to this work.

the resonance frequencies of two sequentially adjacent amide moieties. Clearly, this information will be adequate for the sequential assignment of the backbone resonances, provided that there is sufficient dispersion of the amide group chemical shifts. In this paper, we used the 6D peak lists thus obtained as input for the program GARANT (Bartels et al., 1997) in order to obtain NMR assignments for the two proteins 434-repressor(1–63) (Neri et al., 1992) and TM1290 (Etezady-Esfarjani et al., 2004), in an approach which does not involve any further human interaction after starting the APSY-NMR experiments.

Materials and methods

NMR measurements

The 6D APSY-seq-HNCOCANH experiment with a 0.9 mM solution of 434-repressor(1–63) was recorded at 30°C on a Bruker DRX 750 MHz spectrometer equipped with a triple-resonance probehead with a z-gradient coil. The spectral widths were 3000 Hz, 1600 Hz, 1900 Hz and 5700 Hz in the $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$ and $^{13}\text{C}^{\alpha}$ dimensions, respectively. The interscan delay was 1.0 s, and 32 transients were accumulated for each increment of the combined evolution time. 1024 complex points were recorded in the acquisition

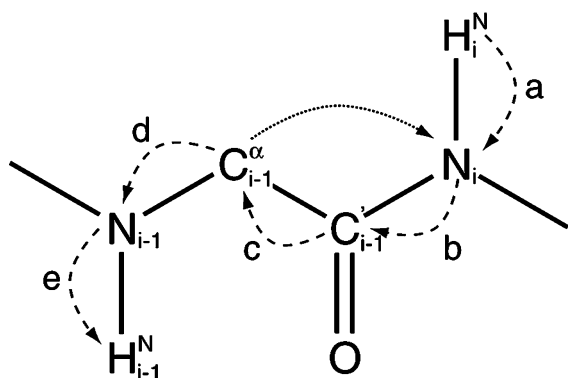


Figure 1. Magnetization transfer in the 6D APSY-seq-HNCOCANH experiment. The pathway represented by the dashed arrows a to e (Equation (5)) leads to a single peak in the 6-dimensional frequency space. The dotted arrow indicates an undesired magnetization back transfer, which is suppressed by the experimental scheme shown in Figure 2 (see text).

dimension, with a sweep width of 11.0 ppm. In the indirect dimension, 64 complex points were measured. The total spectrometer time used for the recording of 25 2D projections (Table 1) was 40 h.

The corresponding experiment with a 3.0 mM solution of TM1290 was recorded at 35°C on a Bruker DRX 500 MHz spectrometer equipped with a triple-resonance cryogenic probehead with a z-gradient coil. The spectral widths were 2000 Hz, 1650 Hz, 1500 Hz, and 3800 Hz in the $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$ and $^{13}\text{C}^{\alpha}$ dimensions, respectively. The interscan delay was 1.0 s, and 8 transients were accumulated. 1024 complex points were recorded in the acquisition dimension, with a sweep width of 12.0 ppm. In the indirect dimension, 128 complex points were measured. The total spectrometer time used for the recording of 25 2D projections (Table 1) was 19.6 h.

Recording of projection spectra

For a 2D projection of the 6D APSY-seq-HNCOCANH experiment with the four projection angles α , β , γ and δ , the time domain was sampled along a straight line defined by the unit vector \vec{p}_1 ,

$$\vec{p}_1 = \begin{pmatrix} \sin\delta \\ \sin\gamma \cdot \cos\delta \\ \sin\beta \cdot \cos\gamma \cdot \cos\delta \\ \sin\alpha \cdot \cos\beta \cdot \cos\gamma \cdot \cos\delta \\ \cos\alpha \cdot \cos\beta \cdot \cos\gamma \cdot \cos\delta \\ 0 \end{pmatrix}. \quad (1)$$

The spectral width, SW , of each 2D projection was calculated from

$$SW = \sum_{i=1}^5 p_1^i \cdot SW_i, \quad (2)$$

where p_1^i represents the coordinates of the vector \vec{p}_1 (see Equation (1)), and SW_i are the spectral widths of the five indirect dimensions. The dwell time for the recording of discrete data points, Δ , was calculated as

$$\Delta = 1/SW, \quad (3)$$

and the resulting increments for the evolution times t_i , Δ_i , in the five indirect dimensions are given by

$$\Delta_i = p_1^i \cdot \Delta. \quad (4)$$

Quadrature detection was achieved using hypercomplex Fourier transformation on pure sine and cosine terms of corresponding positive and negative projection angles, which were obtained using the trigonometric addition theorem (Brutscher et al., 1995; Freeman and Kupče, 2004). In the 6D-seq-HNCOCANH experiment (Figure 2), the five evolution time periods t_i ($i = 1, \dots, 5$) were performed as semi-constant time or as constant time periods, depending on the value of the increment Δ_i , and on the number of complex points, n , recorded in the indirect dimension. The initial values for t_i^a , t_i^b and t_i^c are given in the caption to Figure 2. For $\Delta_i/2 > t_i^c/n$, semi-constant time evolution periods were applied, with t_i^c being decremented by $\Delta_i^c = t_i^c/n$, and t_i^a and t_i^b incremented by $\Delta_i^a = \Delta_i/2$ and $\Delta_i^b = \Delta_i^a - \Delta_i^c$, respectively. In all other situations, constant time evolution periods were used, with t_i^c decremented by $\Delta_i^c = \Delta_i/2$, t_i^a incremented by $\Delta_i^a = \Delta_i/2$, and t_i^b maintaining its initial value.

Processing of the 2D projection spectra

All data processing steps were fully automated. Zero-order phase correction for the direct dimension was determined with PROSA (Güntert et al., 1992). The use of constant-time or semi-

constant time evolution periods and proper phase settings in the pulse program (Figure 2) provides absorptive in-phase signals for the indirect dimension, without need for phase correction. In both dimensions, the FID was multiplied with a 75°-shifted sine bell prior to Fourier transformation (De Marco and Wüthrich, 1976), and zero-filled to the next power of two complex points. The baseline was corrected using the IFLAT method (Bartels et al., 1995) in the direct dimension, and polynomials in the indirect dimension.

APSY analysis

The 2D projection spectra were automatically peak picked with the peak picking routine of ATNOS (Herrmann et al., 2002), with $R_{\min} = 4.0$ for 434-repressor(1–63) and $R_{\min} = 5.0$ for TM1290. The resulting peak lists of the projections were analyzed with the algorithm GAPRO (Hiller et al., 2005), which identifies all projected peaks arising from the same 6D peak, and then calculates the final 6D peak list. For 434-repressor(1–63), GAPRO was applied with the parameters $j = 25$, $S_{\min,1} = S_{\min,2} = 5$, $k = 50$, $w = 200$, $r_{\min} = 50$ Hz, $\Delta v_{\min} = 7.5$ Hz. The calculation time on a standard LINUX PC with a 2.8 GHz Pentium 4 processor

Table 1. Values of the projection angles α , β , γ and δ (see text), and of the spectral widths (SW) in the dimension ω_{1-5}^a used here for the recording of 25 2D projections of the 6D APSY-seq-HNCOCANH experiment, the resulting linear combination of frequencies are given in column LC.

α	β	γ	δ	SW [Hz]	LC
0°	0°	0°	0°	1650	ω_5
90°	0°	0°	0°	3800	ω_4
0°	90°	0°	0°	1500	ω_3
0°	0°	90°	0°	1650	ω_2
0°	0°	0°	90°	2000	ω_1
0°	0°	90°	$\pm 30^\circ$	2429	$\omega_2 \cos(30^\circ) \pm \omega_1 \sin(30^\circ)$
0°	90°	0°	$\pm 60^\circ$	2482	$\omega_3 \cos(60^\circ) \pm \omega_1 \sin(60^\circ)$
90°	0°	0°	$\pm 30^\circ$	4290	$\omega_4 \cos(30^\circ) \pm \omega_1 \sin(30^\circ)$
0°	0°	0°	$\pm 60^\circ$	2557	$\omega_5 \cos(60^\circ) \pm \omega_1 \sin(60^\circ)$
0°	90°	$\pm 30^\circ$	0°	2124	$\omega_3 \cos(30^\circ) \pm \omega_2 \sin(30^\circ)$
90°	0°	$\pm 60^\circ$	0°	3329	$\omega_4 \cos(60^\circ) \pm \omega_2 \sin(60^\circ)$
0°	0°	$\pm 30^\circ$	0°	2254	$\omega_5 \cos(30^\circ) \pm \omega_2 \sin(30^\circ)$
90°	$\pm 60^\circ$	0°	0°	3199	$\omega_4 \cos(60^\circ) \pm \omega_3 \sin(60^\circ)$
0°	$\pm 30^\circ$	0°	0°	2179	$\omega_5 \cos(30^\circ) \pm \omega_3 \sin(30^\circ)$
$\pm 30^\circ$	0°	0°	0°	3329	$\omega_5 \cos(30^\circ) \pm \omega_4 \sin(30^\circ)$

^a The dimension ω_{1-5} consists of projections of the five indirect dimensions $\omega_1(^1\text{H}^N)$, $\omega_2(^{15}\text{N})$, $\omega_3(^{13}\text{C}')$, $\omega_4(^{13}\text{C}^\alpha)$, $\omega_5(^{15}\text{N})$ (see column LC and the text).

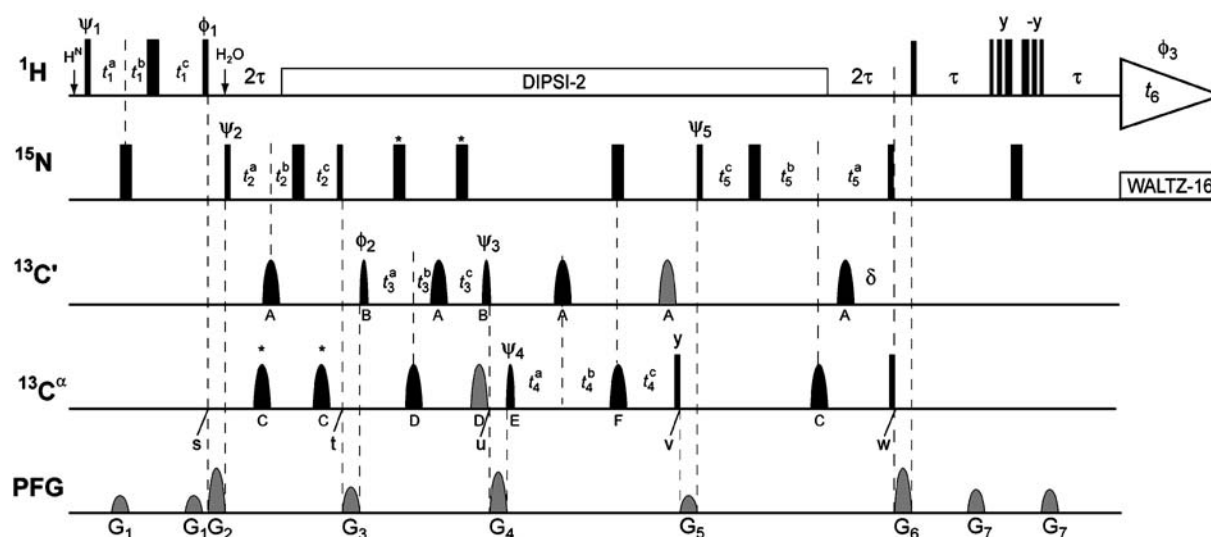


Figure 2. 6D-APSY-seq-HNCOCANH experimental scheme, which correlates pairs of sequentially adjoining amide protons (Figure 1). The alternative out-and-back magnetization transfer pathway, which ends at the starting amide moiety, is efficiently suppressed (see text). Radio-frequency pulses were applied at 118.0 ppm for ^{15}N , 173.0 ppm for carbonyl carbons ($^{13}\text{C}'$), and 54.0 ppm for $^{13}\text{C}^\alpha$. At the outset of the pulse sequence (“ $^1\text{H}^{\text{N}}$ ” on the line ^1H), the proton carrier frequency was set in the amide proton region at 8.2 ppm, and at the time point “ H_2O ” the carrier was set to 4.7 ppm. Narrow and wide rectangular bars represent non-selective 90° and 180° pulses, respectively. Thin sine bells are shaped 90° pulses, and fat sine bells are shaped 180° pulses. Individual shaped pulses are identified with the capital letters A to F. The actual shapes depend on the purpose of the pulses, and the durations depend on the shape and the spectrometer frequency (here: 500 MHz): A, Gaussian shape, 150 μs ; B, Gaussian, 180 μs ; C, I-burp (Geen and Freeman, 1991), 300 μs ; D, Gaussian, 120 μs ; E, E-burp (Geen and Freeman, 1991), 400 μs ; F, RE-burp (Geen and Freeman, 1991), 525 μs . The last six hard pulses on the ^1H line represent a 3-9-19 WATERGATE element (Sklenar et al., 1993). The four pulses marked with an asterisk are continuously centered with respect to the time periods $t_2^a + t_2^b$, t_2^c , $t_3^a + t_3^b$, and t_3^c , respectively. Grey pulses were applied for compensation of off-resonance effects (Bloch and Siegert, 1948; McCoy and Mueller, 1992). Decoupling using DIPSI-2 on ^1H (Shaka et al., 1988) and WALTZ-16 on ^{15}N (Shaka et al., 1983) is indicated by white rectangles. The triangle with t_6 represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped, pulsed magnetic field gradients applied along the z-axis, with the following durations and strengths: G_1 , 700 μs , 13 G/cm; G_2 , 1000 μs , 35 G/cm; G_3 , 800 μs , 20 G/cm; G_4 , 800 μs , 32 G/cm; G_5 , 800 μs , 13 G/cm; G_6 , 1000 μs , 35 G/cm; G_7 , 800 μs , 18 G/cm. The following phase cycling was used: $\phi_1 = y, -y$, $\phi_2 = x, x, -x, -x$, $\phi_3 = x, -x, -x, x$, $\psi_1 = x$, $\psi_2 = x$, $\psi_3 = y$, $\psi_4 = x$, $\psi_5 = x$, and all pulses without indication of a phase above the pulse symbol were applied with phase x . The time points s, t, u, v , and w are discussed in the text. The following initial delays were used: $t_1^a = t_1^c = 2.7$ ms, $t_2^a = t_2^c = 14.0$ ms, $t_3^a = t_3^c = 4.75$ ms, $t_4^a = 4.75$ ms, $t_4^b = 20.25$ ms, $t_4^c = 25.0$ ms, $t_5^a = t_5^c = 14.0$ ms, and $t_1^b = t_2^b = t_3^b = t_5^b = 0$ ms. The delay $\tau = 2.7$ ms was invariant during the experiment, and the delay δ was continuously adjusted to $\delta = (t_5^a + t_5^b - t_5^c)/2$. In the five indirect evolution periods, constant-time or semi-constant-time periods were applied (see text). Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem to obtain pure cosine and sine terms for a subsequent hypercomplex Fourier transformation (Brutscher et al., 1995; Kupče and Freeman, 2004). The pulse phases ψ_1 to ψ_5 were used for this purpose for the dimensions ω_1 to ω_5 , respectively, and only the pulse phases of the evolution periods which are part of the given projection (Table 1) are phase cycled. For consecutive FIDs, ψ_1, ψ_2, ψ_3 , and ψ_4 were simultaneously incremented in 90° steps, and ψ_5 was decremented in 90° steps.

was approximately 10 min. For TM1290, GAPRO was applied with the parameters $j=25$, $S_{\text{min},1} = S_{\text{min},2} = 5$, $k=100$, $w=400$, $r_{\text{min}} = 50$ Hz, $\Delta v_{\text{min}} = 5.0$ Hz. The calculation time was approximately 30 min.

Protocol and parameters used for automated assignment with GARANT

The precise and artifact-free 6D peak lists obtained for the two proteins from the 6D-APSY-

seq-HNCOCANH experiment were used as input for the assignment algorithm GARANT (Bartels et al., 1997), which had been modified by increasing the maximally allowed dimension from 4 to 6. GARANT was applied to the 6D peak lists using a standard annealing protocol (Bartels et al., 1997). For each peak list, 30 resonance assignment calculations were performed, using as input the same 6D-APSY-seq-HNCOCANH peak list but different, randomly chosen starting conditions. The 30 resulting sequence-specific resonance assignments

were merged into one list. An assignment for a given atom was accepted if the same result, within 0.02 ppm and 0.4 ppm for protons and heavy atoms, respectively, was obtained in at least 50% of the calculations. Otherwise, the backbone atom was considered to remain unassigned.

Results

The 6D APSY-seq-HNCOCANH NMR experiment

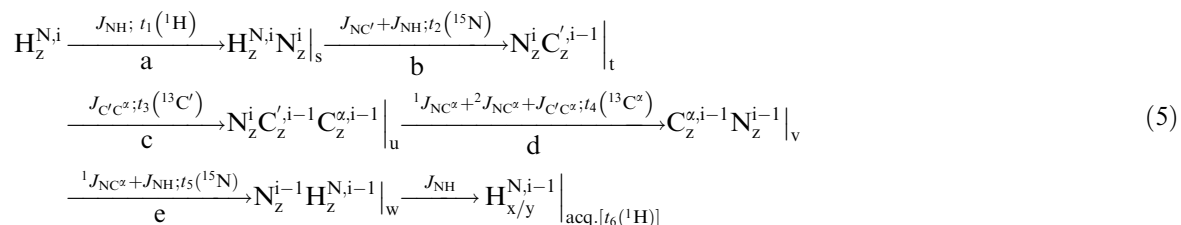
This experiment connects sequentially adjacent amide moieties. Previous experimental schemes providing corresponding information include 3D and 4D HN(COCA)NH (Bax and Grzesiek, 1993), which also connect sequential amide moieties via C^α and the carbonyl carbon. With APSY, the full potential of this magnetization transfer pathway is exploited in a six-dimensional experiment (Figure 1). The existing 3D and 4D versions give rise to an intraresidual and a sequential peak (Grzesiek et al., 1993; Matsuo et al., 1996; Panchal et al., 2001). Since the intraresidual peak resulting from the ‘‘back transfer’’ magnetization pathway would not contribute any new information, it is efficiently suppressed in the 6D APSY-seq-HNCOCANH experiment, and only one peak for each pair of sequentially adjacent amide moieties appears in the 6D frequency space. In the coherence transfer scheme of Equation (5), the transfer steps of Figure 1 are represented by Cartesian product operators (Sørensen et al., 1983), where chemical shift evolution and transfer efficiencies are not represented, and only relevant magnetization components are retained.

nucleus in parentheses). Below the arrows, the letters a to e relate the magnetization transfer steps to the corresponding steps in Figure 1. The characters s, t, u, v and w at the end of the individual product operator expressions indicate time points in the experimental scheme shown in Figure 2. The suppression of the back transfer pathway (dotted line in Figure 1) is implemented between the time points u and v, where the one-bond scalar coupling ${}^1J_{NC^\alpha}$ as well as the two-bond coupling ${}^2J_{NC^\alpha}$ are active, leading to the terms $C_Z^{\alpha,i-1}N_Z^{i-1}$ and $C_Z^{\alpha,i-1}N_Z^i$ at time v. The term $C_Z^{\alpha,i-1}N_Z^{i-1}$ is associated with the sequential pathway and eventually leads to $H_{x/y}^{N,i-1}$, which is the desired signal measured during the acquisition period. The term $C_Z^{\alpha,i-1}N_Z^i$ leads to back transfer and ends up as undesired magnetization on the amide proton from which the experiment was started, $H_{x/y}^{N,i-1}$ (Panchal et al., 2001). The transfer efficiencies for the terms $C_Z^{\alpha,i-1}N_Z^{i-1}$ and $C_Z^{\alpha,i-1}N_Z^i$, respectively, are proportional to the expressions (6) and (7):

$$\sin(\pi \cdot T \cdot {}^1J_{NC^\alpha}) \cdot \sin(\pi \cdot T \cdot {}^2J_{NC^\alpha}) \quad (6)$$

$$\cos(\pi \cdot T \cdot {}^1J_{NC^\alpha}) \cdot \cos(\pi \cdot T \cdot {}^2J_{NC^\alpha}). \quad (7)$$

The values of $1/(2 \cdot {}^1J_{NC^\alpha})$ and $1/(2 \cdot {}^2J_{NC^\alpha})$ are similar, and adjusting the time period $T = t_4^a + t_4^b + t_4^c$ between u and v (Figure 2) to the average of these two values results in strong attenuation of the back transfer peak (Equations (5) to (7)). During the period T , the magnetization evolves also due to the scalar coupling between α - and β -carbons, and values of T close to multiples of 27 ms are therefore preferred for optimal sen-



Above the arrows which connect subsequent states, we indicate the active J coupling constants and the evolution periods t_i , with the active

sitivity. Overall, $T = 50$ ms was found to be a good compromise. Considering the variations of ${}^1J_{NC^\alpha}$ and ${}^2J_{NC^\alpha}$ along a polypeptide chain, the back

transfer of magnetization (dotted line in Figure 1) can then reach at most 10% of the intensity of the corresponding desired peak (Brutscher, 2002). In the 6D APSY-seq-HNCOCANH experiment the undesired resonance was very well suppressed with $T=50$ ms, retaining high sensitivity for the observation of the desired sequential signal (Figure 3).

APSY-NMR with the proteins 434-repressor(1–63) and TM1290

In the context of this paper, 6D APSY-seq-HNCOCANH experiments (Figure 2) were recorded with the two proteins TM1290 and 434-repressor(1–63). For each protein, 25 projections were measured, using the projection angles given in Table 1. 20 projection angles were chosen such

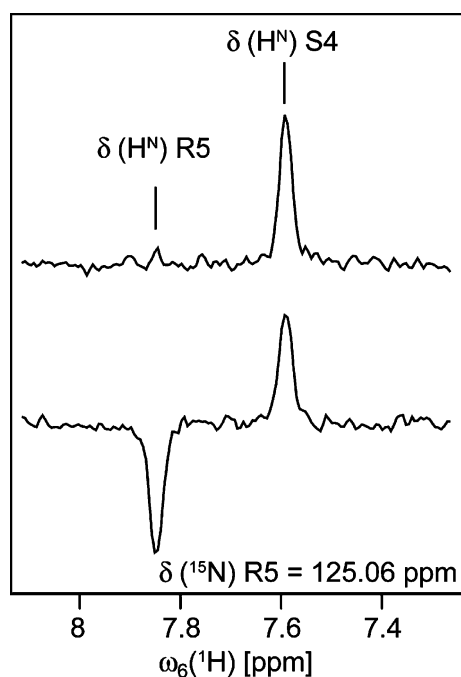


Figure 3. Suppression of the $C_{i-1}^{\alpha} \rightarrow {}^{15}N_i$ back transfer (Figure 1) in the 6D-APSY-seq-HNCOCANH spectrum of [$^{13}C, {}^{15}N$]-labeled 434-repressor(1–63) (protein concentration = 0.9 mM, 20 mM sodium phosphate at pH 6.5, $T=15^{\circ}C$). Shown are 1D cross sections through the projection spectra recorded with the projection angles $(\alpha, \beta, \gamma, \delta) = (0^{\circ}, 90^{\circ}, 0^{\circ}, 0^{\circ})$, which were taken at the Arg 5 ${}^{15}N$ -frequency of 125.06 ppm and cover the ${}^1H^N$ chemical shift range containing the intraresidual correlation to Arg 5 and the sequential correlation to Ser 4. The time period $T = t_4^a + t_4^b + t_4^c$ (Figure 2) was set to 50 ms (top) and 27 ms (bottom), resulting in experiments with (top) and without (bottom) suppression of back transfer.

that all possible pairs of evolution dimensions are combined in two projections. In addition, 5 “direct” projections along a single evolution dimension were measured. These projections fully exploit the six-dimensionality of the experiment, even though none of the selected individual projection includes chemical shift information of more than 3 nuclei (Table 1). Combinations of more than two evolution dimensions could readily be recorded and analyzed with the present setup. The price to pay would be that these projection experiments have reduced sensitivity, since the sensitivity is proportional to $(1/\sqrt{2})^{n-1}$, where n is the number of combined evolution dimensions.

The projection spectra were automatically peak picked, and the 2D peak lists were used as input for the GAPRO algorithm in order to compute the 6D peak positions (Hiller et al., 2005). Since peak picking uses the individual projections, the sensitivity exploited in this step is in principle that of the individual 2D projections. However, GAPRO makes further use of the fact that the NMR peak positions in different projections are correlated, so that noise peaks can be efficiently eliminated. The signal-to-noise ratio can therefore be as low as 3–5 without generating artifacts in the final results from the peak picking (S/N ratio = peak maximum divided by the standard deviation of the noise).

For the two protein samples presented here, sensitivity was not a limiting factor. In both applications, a complete 6D APSY peak list was obtained for the residues with observable ${}^{15}N$ - 1H signals (see below). Since the 6D APSY-seq-HNCOCANH scheme correlates two sequential amide protons, the N-terminal residue, the Pro residues, and possibly other residues without observable ${}^1H^N$ signal are not contained in this peak list.

For 434-repressor(1–63), the peak list contained 56 out of 57 peaks that would be expected from the chemical structure of the molecule, and there were no artifacts. The connection between residues 2 and 3 was missing, probably because local dynamics lead to a reduction of the signal below the noise level in all projection spectra.

For TM1290, all but three of the peaks expected from the previous, interactively determined assignment (Etezady-Esfarjani et al., 2003) were contained in the final 6D peak list. TM1290 contains dynamically disordered regions, which

makes the sequential $^1\text{H}^{\text{N}}\text{-}^1\text{H}^{\text{N}}$ connectivities in the regions 46–52 and 72–73 NMR-unobservable (Etezady-Esfarjani et al., 2003). Three additional peaks were missing in the GAPRO peak list, i.e., the connections 19–20, 20–21 and 71–72, because the signal intensities were below the noise level in the 6D APSY-seq-HNCOCANH data set but not in the conventional triple resonance experiments (Etezady-Esfarjani et al., 2003). The final GAPRO peak list thus contained 98 sequential peaks, and there were no artifacts.

In Figure 4, the 6D peak list obtained from 25 2D projections (Table 1) of the 6D APSY-seq-HNCOCANH experiment with TM1290 is projected onto two experimental 2D projections, visualizing the excellent agreement between the calculated and measured peak positions. The precision of the chemical shifts can be directly measured in the 6D APSY-seq-HNCOCANH data set, since the resonance of each amide moiety is part of two different 6D peaks. In one of these two peaks, the amide proton chemical shift is measured in the direct dimension (ω_6) during acquisition, and in the second peak the same $^1\text{H}^{\text{N}}$ proton shift contributes to the indirect dimension (ω_1); the corresponding ^{15}N chemical shift is also measured twice, in ω_5 and ω_2 . There are 93 amide moieties that contribute to two peaks in the 6D peak list of TM1290, and from these the chemical shift differences $\omega_1(^1\text{H}) - \omega_6(^1\text{H})$ and $\omega_2(^{15}\text{N}) - \omega_5(^{15}\text{N})$ were calculated. The resulting standard deviation for the proton chemical shifts was 0.0014 ppm (0.72 Hz), with a maximal deviation of 0.005 ppm (2.5 Hz), and the standard deviation for the nitrogen chemical shifts was 0.0137 ppm (0.69 Hz), with a maximal deviation of 0.048 ppm (2.4 Hz).

Backbone resonance assignment of 434-repressor (1–63)

The software GARANT had previously been shown to yield reliable automated sequential NMR assignments, provided that an input of precise, artifact-free peak lists was available (Bartels et al., 1997). We therefore chose GARANT to perform sequential assignments based on the presently obtained 6D APSY peak lists.

For the 434-repressor(1–63), 30 GARANT calculations all converged to the correct sequence-specific resonance assignment for all the

resonances contained in the 6D peak list. This assignment procedure based on the 6D APSY-seq-HNCOCANH experiment yielded three fragments of length 39, 3 and 17 residues, which could unambiguously be fitted to the sequence of 434-repressor(1–63) in the positions 3–41, 43–45, and 47–63, respectively. These fragments are separated by the two prolines in positions 42 and 46. For the residues 1 and 2, the $^{15}\text{N}\text{-}^1\text{H}$ NMR signals were not observed. A plot corresponding to Figure 5 therefore contains a continuous thick line at the level $N=30$ for the polypeptide segments 3–41, 43–45, and 47–63 (not shown).

Backbone resonance assignment for TM1290

30 GARANT calculations made with the 6D APSY-peak list of TM1290 did not all converge to the correct solution, but for all the resonances observed in the 6D experiment the correct sequence-specific assignment was found in at least half of the calculations. Erroneous assignments of residues for which no NMR signals had been observed were obtained only in a small number of the calculations (Figure 5). This result can be rationalized based on previous work with TM1290, which had shown that the proline residue in positions 6, and other residues with NMR-unobservable amide moieties divide the TM1290 sequence into sequentially connected segments of residues 2–5, 7–45, 53–71 and 74–115 (Etezady-Esfarjani et al., 2003). As mentioned above, three signals expected from the results of this earlier work could not be detected by the 6D APSY experiment, so that overall the 6D correlation NMR signals with the residues 1, 6, 20, 46–52 and 72–73 were absent.

Discussion

The results obtained for the 434-repressor(1–63) and for TM1290 (Figures 4 and 5) document that 6D APSY-seq-HNCOCANH data sets can provide the information needed for obtaining *de novo* backbone resonance assignments for small and medium-size proteins. Here we want to further investigate the robustness of the procedure with respect to deterioration of the 6D data sets, and to replacement of 6D APSY-NMR with lower-dimensional APSY-NMR experiments.

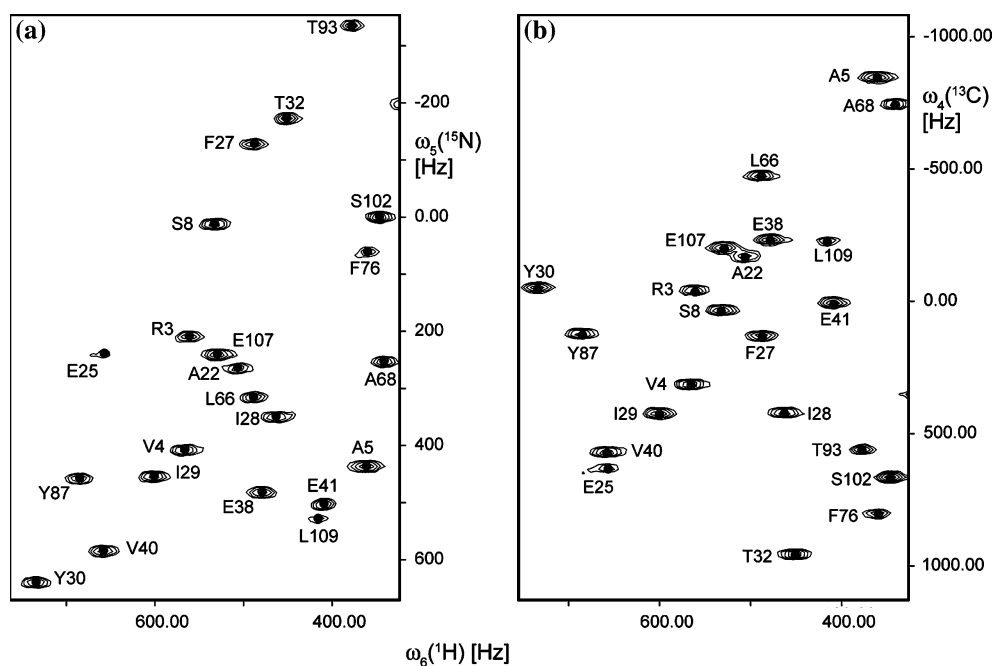


Figure 4. Visualization of part of the result of the fully automated sequence-specific backbone resonance assignment of the [^{13}C , ^{15}N]-labeled protein TM1290 (protein concentration 3.0 mM, 20 mM phosphate buffer at pH 6.0, $T = 35^\circ\text{C}$). (a) and (b) are regions of two orthogonal projection spectra from a 6D APSY-seq-HNCACONH experiment with projection angles $(\alpha, \beta, \gamma, \delta) = (0^\circ, 0^\circ, 0^\circ, 0^\circ)$ and $(90^\circ, 0^\circ, 0^\circ, 0^\circ)$, respectively. The 6D APSY peak list is projected onto the spectra by black dots. The sequence-specific resonance assignment, as found by GARANT, is written next to each peak, using the one-letter amino acid code and the sequence position.

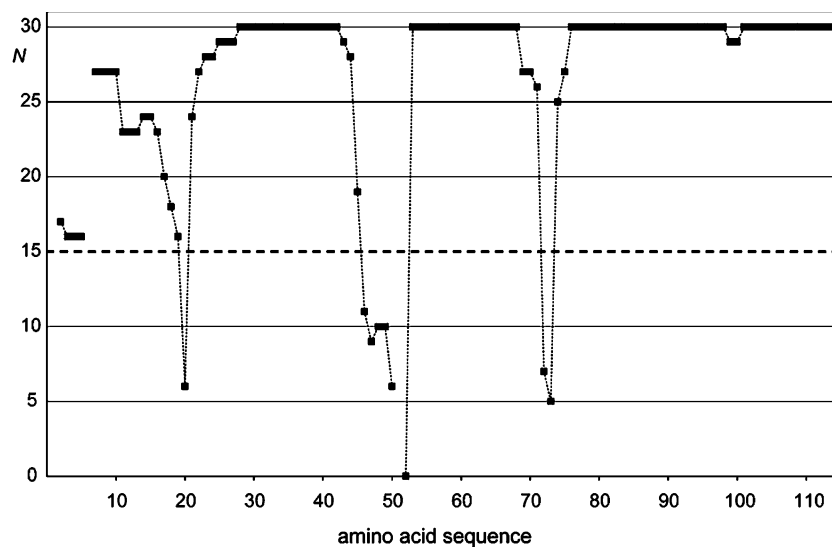


Figure 5. Automated sequence-specific resonance assignment for the protein TM1290 obtained with GARANT (Bartels et al., 1997), using as input the peak lists from a 6D APSY-seq-HNCOCANH experiment calculated with the program GAPRO. GARANT was run 30 times with the same input list of 6D peaks but with different, randomly chosen starting conditions. For each residue the most frequent ^{15}N assignment was counted and plotted as a square along the vertical N-axis (thick black lines correspond to a sequence of squares). If $N \geq 15$ (dashed horizontal line), the sequence-specific assignment was accepted. All other ^{15}N were considered to remain unassigned. The amide protons of residues 1, 6, 20, 46–52 and 72–73, which include the prolines 6 and 51, are not contained in the 6D APSY-seq-HNCOCANH resonances. With the acceptance criterion used here, all these amide protons remained unassigned.

Model calculations on the impact of type and quality of APSY-NMR data on the resonance assignments

Peak lists from 6D APSY-seq-HNCOCANH have outstandingly high information content, with six chemical shifts per peak, high accuracy of the chemical shifts, and no or very few artifacts. The influence of variable quality of the input on the automated assignment procedure was evaluated by computationally generating lower-dimensional peak lists, which in their combinations had, in principle, the same information content as the 6D peak list. We thus generated peak lists for combinations of the two 4D spectra HNCOCA and HNCACO, and the four 3D spectra HNCA, HN(CO)CA, HN(CA)CO and HNCO from the 6D APSY-NMR peak list of TM1290. These peak lists contain overlapping peaks that could not be resolved in the 3D or 4D spectra, but are resolved in the 6D APSY-HNCOCANH spectrum. The two 4D and the four 3D peak lists, respectively, were used as input for 30 GARANT calculations (Figure 6, c and d). Although these computationally generated peak lists had the same chemical shift precision as the experimental 6D peak list, 4 and 15 of the previously assigned residues, respectively, were not assigned by GARANT when using the 4D and 3D data sets, and the 4D data sets led to two erroneous assignments.

The influence of the accuracy of the chemical shifts, and of the presence of artifacts in the input peak lists on the automated assignment procedure was tested by randomly varying the peak positions in the 6D APSY-NMR peak list of TM1290, and by adding artifacts to it. In one of the “deteriorated” 6D peak lists, the peak positions in the dimensions $\omega_6(^1\text{H})$ and $\omega_5(^{15}\text{N})$ were randomly changed within ranges of 0.02 ppm and 0.4 ppm, respectively (Figure 6e), ensuring that all connectivities could be found within the tolerance range used by GARANT. In another modification of the 6D peak list, 10 artifacts were added at random positions within the experimental spectral ranges (Figure 6f). Finally, chemical shift variation of individual peaks and the addition of artifacts were combined in a single list (Figure 6g). Each type of deteriorated list was generated 10 times, with different, randomly chosen chemical shift changes and/or additions of artifacts at different random positions. For each of the resulting 30 peak lists,

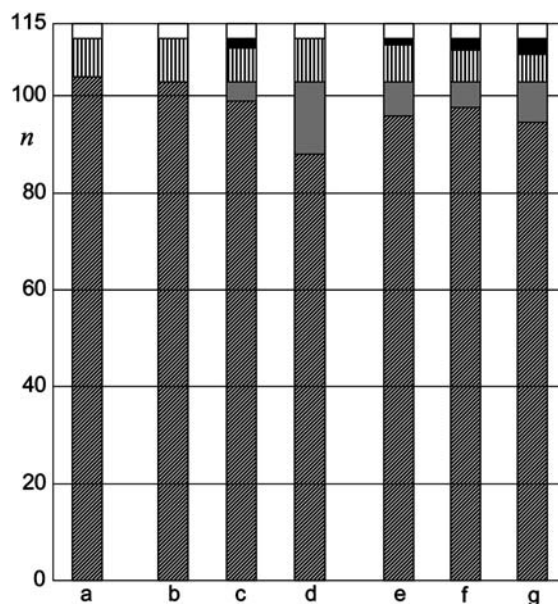


Figure 6. Extent of the automated backbone assignment of the protein TM1290 achieved with GARANT when using input peak lists of different quality. The height of the diagonally dashed areas along the vertical axis indicates the number of $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ moieties for which both $^1\text{H}^{\text{N}}$ and ^{15}N were correctly assigned in the final chemical shift list; gray areas indicate the number of amide group resonances that remained unassigned; vertically dashed areas represent the number of NMR-unobservable amide groups (see text); black areas indicate the number of erroneous assignments, which all occur for residues for which no NMR signals were observed; the white area at the top indicates the residues for which no $^1\text{H}^{\text{N}}$ signals are expected, i.e., the N-terminal residue and the two prolines. The interpretation of this data can be based on the following properties of TM1290, which are visualized in (a): the protein contains 115 residues with 112 backbone amide groups that should normally be observable by NMR. Of these, 8 were not detected in conventional 2D and 3D NMR experiments, so that sequence-specific assignments were obtained for 104 amide groups (Etezady-Esfarjani et al., 2003). (b) to (g) represent the extent of assignments obtained with GARANT when using the following six different input data sets: (b) final 6D peak list from the 6D APSY-seq-HNCOCANH measurement (same as Figure 5); (c) Two 4D peak lists computed from the original 6D peak list for the experiments 4D APSY-HNCOCA and 4D APSY-HNCACO; (d) Four 3D peak lists computed from the original 6D peak list for the experiments 3D APSY-HNCA, 3D APSY-HN(CO)CA, 3D APSY-HNCO and 3D APSY-HN(CA)CO; (e) Ten 6D peak lists generated by randomly varying the peak positions in the experimental 6D peak list (b) along $\omega_6(^1\text{N})$ and $\omega_5(^{15}\text{N})$ within 0.02 and 0.4 ppm, respectively; (f) Ten 6D peak lists generated by adding 10 artifacts at random positions to the experimental 6D peak list (b); (g) Combination of (e) and (f). For each data set, the GARANT analysis was repeated 30 times with different, random starting conditions (Bartels et al., 1997), and the results were analyzed as described in Material and methods. For each of the columns (e) to (g), the average of the results obtained with the 10 deteriorated peak lists is plotted in the figure (see text).

30 GARANT calculations were performed and analyzed as described above for the experimentally obtained 6D APSY-seq-HNCOCANH peak list. The results for the 10 lists with random chemical shift changes, the 10 lists with artifact peaks, and the 10 lists containing both manipulations, respectively, were then averaged. Figure 6, e to g, shows the average number of correctly assigned residues (diagonally dashed areas), the average number of unassigned residues for which an assignment is expected (gray areas), the average number of “correctly unassigned residues”, i.e., residues for which the ^{15}N - ^1H NMR signals could not be detected (vertically dashed areas), and the average number of erroneously assigned residues (black areas).

Overall, these model calculations revealed an encouraging robustness of the APSY-NMR-based assignment, since throughout at least 85% of the NMR-observable residues were assigned, with a correctness rate above 97% (Figure 6). Nonetheless, deterioration of the high-quality peak list obtained from the 6D APSY-NMR experiment did lead to fewer residues being assigned, and to introducing some wrong assignments, which however always occur for residues for which no NMR signals were observed. Similar effects resulted from the replacement of 6D APSY-seq-HNCOCANH with, in principle, equivalent combinations of lower-dimensional APSY-NMR experiments, where the combination of four 3D data sets resulted in the lowest yield of correct assignments. The combination of two 4D data sets, in contrast, gave nearly the same extent of correct assignments as the 6D data set, albeit with introduction of erroneous assignments for two NMR-unobservable residues, which we presently attribute to an intrinsic weakness in the GARANT assignment protocol (see below).

Outlook

The use of a single 6D APSY experiment to perform the backbone resonance assignment avoids calibration problems between different spectra, the high dimensionality eliminates resonance overlaps, and the high information content of a 6D peak reduces the number of possible assignment combinations and thus makes the procedure of sequential backbone assignments efficient and robust. Based on the large dispersion of the ^{15}N

chemical shifts in polypeptide chains (Braun et al., 1994; Wüthrich, 1986), the 6D-APSY-seq-HNCOCANH experiment should in principle be widely applicable.

For practical applications, it will be of interest to further optimize the sensitivity of APSY-NMR experiments. Obviously, minor but not negligible improvements may result from adding water flip-back pulses (Grzesiek and Bax, 1993) and sensitivity enhancement schemes (Palmer et al., 1991; Kay et al., 1992). Furthermore, part of the projections could be measured with individually optimized, shortened pulse sequences, which would improve the sensitivity for these projections (Kupče and Freeman, 2003). For example, direct projection of the $\omega_5(^{15}\text{N})$ -dimension could be replaced by a standard [^{15}N , ^1H]-HSQC experiment. In this context, the outcome of the model studies with 3D and 4D APSY-NMR data sets (Figure 6, c and d) is also highly encouraging. For these intrinsically more sensitive APSY experiments, which are based on combining HNCACO and HNCOCA elements, there is a clear promise for use with larger proteins. Current work in our laboratory is focused on generating combinations of lower-dimensional APSY-NMR experiments suitable for use with larger proteins.

The software GARANT used here to derive automated backbone resonance assignments had been developed to work with peak lists obtained from “conventional” heteronuclear correlation and triple resonance NMR spectra. It allows multiple alternative assignments in situations of peak overlap, and it has been laid out to deal with the presence of artifacts. High-dimensional APSY-NMR data sets typically contain no or at most few peak overlaps and artifacts. Substituting GARANT with a new assignment algorithm that is tailored specifically for the needs to analyze APSY-NMR data for automated resonance assignment, promises to further enhance the potentialities of APSY-NMR-based automated resonance assignment for proteins.

In conclusion, APSY-NMR combined with a suitable assignment algorithm promises to enable fully automated sequence-specific backbone resonance assignments for small and medium-sized proteins. Eliminating the extensive human interactions typically needed so far for obtaining NMR assignments will further add to making high-resolution NMR with proteins an attractive tool in

structural biology and structural genomics. The algorithm GAPRO and the pulse sequence used in this work are available from the authors.

Acknowledgements

The authors thank Drs. Fred F. Damberger and Touraj Etezady-Esfarjani for gifts of 434-repressor(1–63) and TM1290, respectively. This work was supported by the Schweizerischer Nationalfonds and the ETH Zürich through the National Center for Competence in Research (NCCR) Structural Biology.

References

- Bartels, C., Güntert, P., Billeter, M. and Wüthrich, K. (1997) *J. Comput. Chem.*, **18**, 139–149.
- Bartels, C., Güntert, P. and Wüthrich, K. (1995) *J. Magn. Reson. A.*, **117**, 330–333.
- Bax, A. and Grzesiek, S. (1993) *Accts. Chem. Res.*, **26**, 131–138.
- Bloch, F. and Siegert, A. (1948) *Phys. Rev.*, **57**, 522–527.
- Braun, D., Wider, G. and Wüthrich, K. (1994) *J. Am. Chem. Soc.*, **116**, 8466–8469.
- Brutscher, B. (2002) *J. Magn. Reson.*, **156**, 155–159.
- Brutscher, B., Morelle, N., Cordier, F. and Marion, D. (1995) *J. Magn. Reson. B.*, **109**, 238–242.
- De Marco, A. and Wüthrich, K. (1976) *J. Magn. Reson.*, **24**, 201–204.
- Etezady-Esfarjani, T., Herrmann, T., Peti, W., Klock, H.E., Lesley, S.A. and Wüthrich, K. (2004) *J. Biomol. NMR*, **29**, 403–406.
- Etezady-Esfarjani, T., Peti, W. and Wüthrich, K. (2003) *J. Biomol. NMR*, **25**, 167–168.
- Freeman, R. and Kupče, E. (2004) *Conc. Magn. Reson.*, **23**, 63–75.
- Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.
- Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 4369–4370.
- Grzesiek, S. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 12593–12594.
- Güntert, P., Dötsch, V., Wider, G. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 619–629.
- Herrmann, T., Güntert, P. and Wüthrich, K. (2002) *J. Biomol. NMR*, **24**, 171–189.
- Hiller, S., Fiorito, F., Wüthrich, K. and Wider, G. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 10876–10881.
- Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
- Kim, S. and Szyperski, T. (2004) *J. Biomol. NMR*, **28**, 117–130.
- Kupče, E. and Freeman, R. (2003) *J. Am. Chem. Soc.*, **125**, 13958–13959.
- Kupče, E. and Freeman, R. (2004) *J. Am. Chem. Soc.*, **126**, 6429–6440.
- Matsuo, H., Kupče, E., Li, H.J. and Wagner, G. (1996) *J. Magn. Reson. B.*, **111**, 194–198.
- McCoy, M.A. and Mueller, L. (1992) *J. Magn. Reson.*, **99**, 18–36.
- Neri, D., Billeter, M. and Wüthrich, K. (1992) *J. Mol. Biol.*, **223**, 743–767.
- Panchal, S.C., Bhavesh, N.S. and Hosur, R.V. (2001) *J. Biomol. NMR*, **20**, 135–147.
- Palmer, A.G., Cavanagh, J., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.
- Shaka, A.J., Keeler, J., Frenkiel, T. and Freeman, R. (1983) *J. Magn. Reson.*, **52**, 335–338.
- Shaka, A.J., Lee, C.J. and Pines, A. (1988) *J. Magn. Reson.*, **77**, 274–293.
- Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson. A.*, **102**, 241–245.
- Sørensen, O.W., Eich, G.W., Levitt, M.H., Bodenhausen, G. and Ernst, R.R. (1983) *Prog. Nucl. Magn. Reson. Spectrosc.*, **16**, 163–192.
- Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N.B. and Montelione, G.T. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 8009–8014.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* John Wiley & Sons, New York.