

APSY-NMR with proteins: practical aspects and backbone assignment

Sebastian Hiller · Gerhard Wider · Kurt Wüthrich

Received: 7 May 2008 / Accepted: 11 August 2008 / Published online: 8 October 2008
© Springer Science+Business Media B.V. 2008

Abstract Automated projection spectroscopy (APSY) is an NMR technique for the recording of discrete sets of projection spectra from higher-dimensional NMR experiments, with automatic identification of the multidimensional chemical shift correlations by the dedicated algorithm GAPRO. This paper presents technical details for optimizing the set-up and the analysis of APSY-NMR experiments with proteins. Since experience so far indicates that the sensitivity for signal detection may become the principal limiting factor for applications with larger proteins or more dilute samples, we performed an APSY-NMR experiment at the limit of sensitivity, and then investigated the effects of varying selected experimental parameters. To obtain the desired reference data, a 4D APSY-HNCOCA experiment with a 12-kDa protein was recorded in 13 min. Based on the analysis of this data set and on general considerations, expressions for the sensitivity of APSY-NMR experiments have been generated to guide the selection of the projection angles, the calculation of the sweep widths, and the choice of other acquisition and

processing parameters. In addition, a new peak picking routine and a new validation tool for the final result of the GAPRO spectral analysis are introduced. In continuation of previous reports on the use of APSY-NMR for sequence-specific resonance assignment of proteins, we present the results of a systematic search for suitable combinations of a minimal number of four- and five-dimensional APSY-NMR experiments that can provide the input for algorithms that generate automated protein backbone assignments.

Keywords Projection spectroscopy · Automated peak picking · Protein NMR · Sequence-specific NMR assignments

Introduction

In NMR spectroscopy with biological macromolecules, recent progress with data sampling, processing and analysis enables practical applications of novel multidimensional NMR experiments (Kim and Szyperski 2003; Orekhov et al. 2003; Kupce and Freeman 2004; Moseley et al. 2004; Eghbalian et al. 2005; Luan et al. 2005; Malmodin and Billeter 2005; Coggins and Zhou 2006; Kupce and Freeman 2006; Szyperski and Atreya 2006; Cornilescu et al. 2007; Lescop and Brutscher 2007; Mishkovsky et al. 2007; Jaravine et al. 2008; Zhang et al. 2008). One of these new approaches is automated projection spectroscopy (APSY; Hiller et al. 2005), which is based on the recording of projections from N -dimensional experiments at arbitrary projection angles (Nagayama et al. 1978; Brutscher et al. 1995; Kupce and Freeman 2004). The data sets thus obtained are analyzed by automated peak picking that provides the input for a geometric analysis algorithm, GAPRO, which generates a list of N -dimensional chemical

S. Hiller · G. Wider · K. Wüthrich (✉)
Institute of Molecular Biology and Biophysics, ETH Zürich,
8093 Zürich, Switzerland
e-mail: wuthrich@mol.biol.ethz.ch

Present Address:

S. Hiller
Department of Biological Chemistry and Molecular
Pharmacology, Harvard Medical School, 240 Longwood
Avenue, Boston, MA 02115, USA
e-mail: hiller.sebastian@gmail.com

K. Wüthrich
Department of Molecular Biology and Skaggs Institute of
Chemical Biology, The Scripps Research Institute, La Jolla, CA
92037, USA
e-mail: wuthrich@mol.biol.ethz.ch

shift correlations. Due to extensive redundancy in the input data for GAPRO, high precision of the chemical shift measurements is achieved (Hiller et al. 2005).

This paper evaluates practical aspects of APSY-NMR spectroscopy. Section "Methods" describes improvements of the GAPRO algorithm, which are based on the experience gained so far from using different experimental schemes with a variety of different proteins. Section "Results and discussion" presents a 4D APSY-HNCOCA experiment with a 116-residue protein that was recorded in 13 min. This data set, which is at the limit of sensitivity, is then used as a reference to discuss the selection of favorable experimental parameters for APSY-NMR spectroscopy, and to derive general recommendations for acquiring, recording and analyzing APSY-NMR experiments with proteins. This includes the results of a systematic search for favorable combinations of a minimal number of four- and five-dimensional APSY-NMR experiments that can provide the data needed as input for algorithms in use for automated sequence-specific backbone assignments in globular proteins (e.g., Bartels et al. 1997; Volk et al. 2008).

Materials and methods

New features of the GAPRO algorithm

GAPRO serves to evaluate the positions of the NMR peaks in the N -dimensional frequency space of the "parent experiment" from the ensemble of all measured projections in an APSY-NMR data set. Based on the experience with applications of APSY-NMR so far, we introduced several new features into this software. Here, we first describe modifications introduced into the core part of GAPRO. This is followed by accounts of the replacement of the peak picker with a new routine, and the addition of a "secondary peak filter".

In the GAPRO core, the parameter w (Table 1), which in the original version of GAPRO (Hiller et al. 2005) represents the number of calculation cycles used to evaluate the high-dimensional peak coordinates, has been replaced by an iterative procedure that converges when the coordinates have been calculated to a precision of 0.1 Hz. This convergence will always be reached, since the possible peak positions are sampled randomly and the geometric average of all calculations is taken. Furthermore, the geometric analysis algorithm has been modified so as to include aliased peaks in the experimental 2D projections into the calculations, such that support for a candidate peak in the high-dimensional frequency space can now also be evaluated directly from aliased peak positions.

The presently introduced implementation of GAPRO uses a new peak picking routine, the "GAPRO peak picker". For the determination of the spectral noise in this routine, the spectrum is divided into 64 equally sized submatrices (Fig. 1a). For each submatrix, the standard deviation of the intensity distribution is calculated. Using the assumption that at least one of the 64 submatrices contains only white noise, the spectral noise, η , is taken as the minimum of the standard deviations of all 64 submatrices. All local maxima of the spectrum are identified, and those with an intensity larger or equal to ηR_{\min} (Table 1) are considered to be peaks. The position of the peak maximum is interpolated by a symmetrization procedure involving the intensities of the two neighboring data points in each dimension (Fig. 1b). This interpolation determines the position of the peak maxima with a precision that is better than the digital resolution (Figs. 1c, 2), and therefore we recommend to use 1 digital point as the input value for Δv_{\min} and r_{\min} (Table 1). Compared to the previously used peak picker (Herrmann et al. 2002; Hiller et al. 2005), the results for the projection spectra

Table 1 Parameters used as part of the input for the software GAPRO

Subroutines	Parameters ^a	Description	Value used ^b
Peak picking	R_{\min}	Signal/noise threshold for peak picking	4.0
	$\Delta\omega_{\text{H}_2\text{O}}$	Frequency range along the acquisition dimension on both sides of the water resonance within which no peaks are picked	850 Hz
GAPRO core functions	S_{\min}	Minimal support value required for the identification of an N -dimensional peak	6
	Δv_{\min}	Intersection tolerance in the directly detected dimension	5.0 Hz
	r_{\min}	Intersection tolerance in the indirect dimensions	15.0 Hz
	k	Number of independent GAPRO calculations from which the final result is derived by a statistical analysis	200
	w	Number of GAPRO results averaged for the computation of the final peak coordinates.	n.a.
Secondary peak filter	R_{single}	Signal:noise threshold at the back-projected peak positions (see text)	0 ^c
	n	Allowed number of violations of R_{single}	0 ^c

^a The three most important GAPRO parameters (see text) are framed with thick solid lines

^b These parameter values were used for the analysis of the data in Figs. 3 and 4. n.a., not applicable in the version of GAPRO used here (see text)

^c $R_{\text{single}} = 0$ and $n = 0$ means that the signal level at a given peak position, ηR_{single} , cannot have a negative value in any of the projection spectra

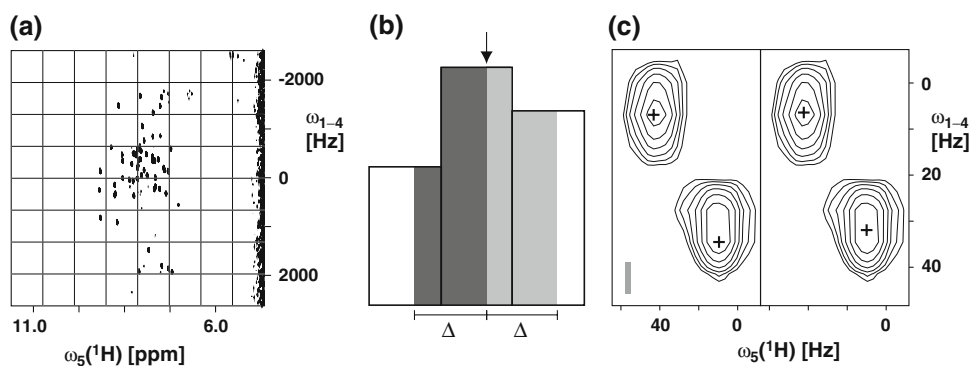


Fig. 1 The GAPRO peak picker. (a) For the determination of the thermal noise RMSD, the spectrum is divided into 64 submatrices of equal size. (b) Interpolation scheme used to determine the position of the peak maximum. Three digital data points are indicated as *open rectangles*, with the height denoting the intensity, and the *width* representing the digital resolution of the spectrum, Δ . The interpolated peak position (*arrow*) is defined such that the light and *dark gray* areas have the same size. (c) Impact of the interpolation scheme

(b) on the positions of two peaks in a 5D APSY-HACACONH experiment recorded with the same sample of the protein TM1290 as used in the experiment of Table 5. The peak positions based on local maxima identification only (*left*), or on local maxima identification plus interpolation (*right*) are indicated by black crosses. The digital resolution in the ω_{1-4} and ω_5 dimensions is indicated by the *gray rectangle* in the *lower left corner*

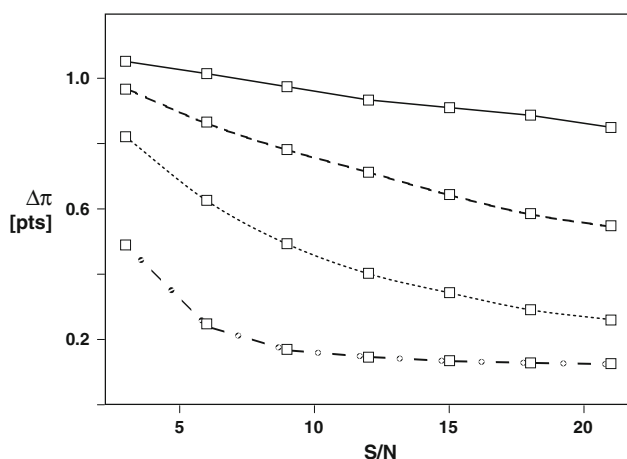


Fig. 2 Simulation of the peak picking precision achieved with the new GAPRO peak picker. The standard deviation of the difference between the real peak position and the picked peak position in units of the digital resolution (digital points, pts), $\Delta\pi$, is plotted versus the ratio of signal intensity and spectral noise (S/N). Each data point (*open squares*) represents an average over 100,000 simulated peak picking events (see text). The four data sets differ in the assumed widths of the resonance lines in digital points: *solid line* 21 pts, *dashed line* 11 pts, *dotted line* 6 pts, *dashed-dotted dashed line* 1 pt

are only marginally different. The main advantage of the new routine is that it requires less CPU time when used as part of the new version of GAPRO.

The precision of the GAPRO peak picker was tested with the following data simulation: A single peak picking of a one-dimensional Lorentzian resonance line with intensity S was digitized with its maximum positioned randomly with respect to the digital grid, random white Gaussian noise with a standard deviation G was added, and

the GAPRO peak picker was applied to the resulting data. The precision of each individual peak picking is then given by the difference between the correct peak position of the original Lorentzian line and the peak position determined by GAPRO. Each data point shown in Fig. 2 represents the standard deviation of the precision calculated from 100,000 such peak picking events.

Although the all-important discrimination between artifacts and noise is based on distinguishing between coherent and incoherent peaks in a sufficiently large number of independently recorded experimental projection spectra, experience so far indicates that it is worthwhile to add a secondary peak filter to be applied to the final result of the GAPRO calculation. Thereby the projections of the high-dimensional peak list, which was obtained from the combined analysis of all the experimental projection spectra, are “back-calculated”, and the signal intensities at the resulting positions in the experimental projections are read out and divided by the spectral noise. Based on user-defined criteria, the GAPRO result is then subjected to additional screening. For example, all peaks with an average sensitivity below a user-defined threshold value of R_{ave} can be deleted, or all peaks with more than n violations of the threshold R_{single} (Table 1) can be deleted, where n and R_{single} are user-defined threshold parameters. The secondary filter thus provides an efficient additional validation of the GAPRO result.

The new version of GAPRO has been implemented in C language, and the evaluation routines have been substantially simplified by optimizing the sorting and matrix inversion procedures. As a result, the computation time is about two orders of magnitude shorter than when using the previous version of GAPRO (Hiller et al. 2005) with the

same computer hardware (see Section “Results and discussion”).

Protein preparation and NMR equipment

The [u- ^{13}C , ^{15}N]-labeled protein TM1290 was produced as described (Etezady-Esfarjani et al. 2003). In the NMR sample the protein concentration was adjusted to 1.0 ± 0.05 mM, as determined by PULCON (Wider and Dreier 2006), and it contained 20 mM phosphate buffer at pH 6.0 and 0.1% NaN_3 in 95% $\text{H}_2\text{O}/5\%$ D_2O . The 4D APSY-HNCOCA experiment of Figs. 3 and 4 was performed at 25°C on a 600 MHz Bruker Avance III spectrometer with a z-gradient triple resonance room temperature probe. The spectra were recorded with the previously described APSY-NMR experimental scheme (Hiller et al. 2005).

The APSY-NMR experiments used to collect the input data for the automated backbone assignment of TM1290 were carried out with the same sample at 25°C on a 750 MHz Bruker Avance III spectrometer with a z-gradient triple resonance room temperature probe. The 5D APSY-HACACONH pulse sequence has been described (Hiller et al. 2005), the new 4D APSY-HACANH pulse sequence is presented in the Appendix 1 (Fig. A1), and the pulse sequence of the 5D APSY-CBCACONH experiment was

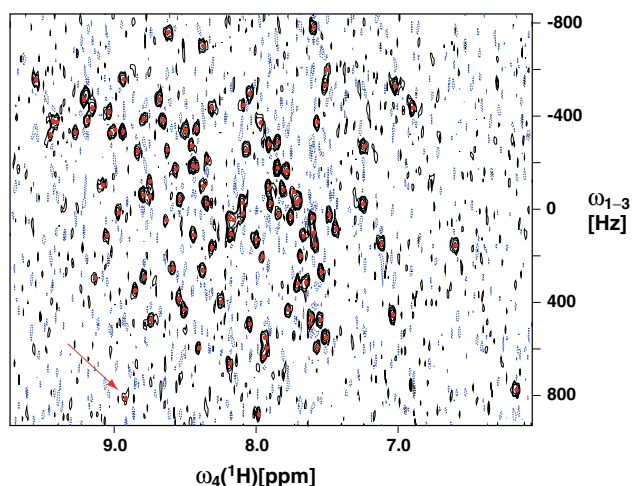


Fig. 3 Experimental 2D projection from a 4D APSY-HNCOCA experiment, with projection angles $(\alpha, \beta) = (20^\circ, -70^\circ)$ corresponding to the vector $(0.94, 0.12, 0.32)$ in the $(^{15}\text{N}, ^{13}\text{C}, ^{13}\text{C}^z)$ time domain. The data set was recorded on a Bruker AvanceIII 600 MHz spectrometer equipped with a room temperature probehead, using a 1 mM sample of TM1290 at 25°C. Thirteen projections were acquired with a total measuring time of 13 min. The *red dots* are the positions defined by projecting the final GAPRO-generated 4D APSY-HNCOCA peak list onto the experimental 2D projection spectrum. The *red arrow* indicates a projected peak position with low signal-to-noise ratio in the experimental spectrum, which is further discussed in Fig. 4

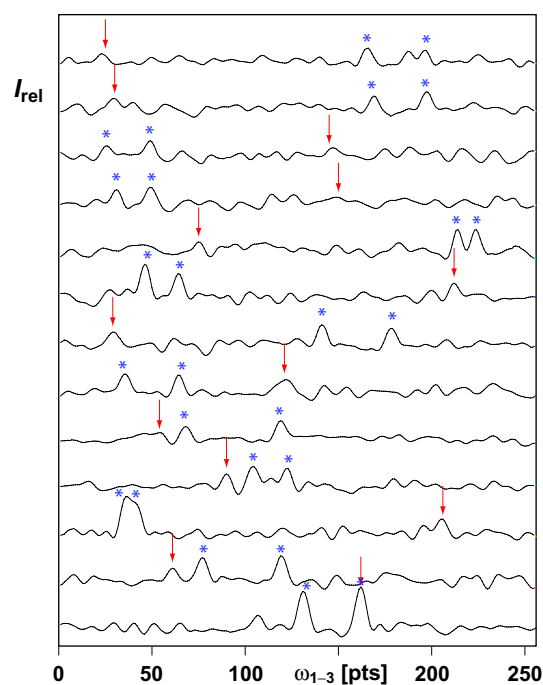


Fig. 4 One-dimensional cross-sections through the 13 2D projections of a 4D APSY-HNCOCA experiment recorded as described in the text and in the caption to Fig. 3. The cross-sections were taken at $\omega_4(^1\text{H}) = 8.957$ ppm, which is the chemical shift of the peak identified in Fig. 3 with an *arrow*. In each of the 13 cross-sections, the position of this peak is again indicated with a *red arrow*, and *blue asterisks* denote two other resonance lines that are present at this ω_4 -position. All other local maxima in the cross-sections arise from random spectral noise

derived in a straightforward way from the corresponding conventional experiment.

For the automated resonance assignment we used the program GARANT (Bartels et al. 1997). The input consisted of the peak lists resulting from the APSY-NMR experiments used, and of the amino acid sequence. Ten independent GARANT assignment runs from different randomly chosen starting conditions were performed, and the assignment acceptance criterion was 50% (Fiorito et al. 2006).

Results and discussion

The experience gained so far indicates that the sensitivity of detection of the NMR signals will in practice probably be a key limiting factor for applications of APSY. In this paper, we therefore chose to record a standard experiment with a medium-sized protein under conditions that are at the limit of signal detection, and then investigate the influence of individual user-selected experimental parameters on the resulting data set. In addition, the influence of the variable GAPRO parameters was investigated by

varying these parameters one by one, while leaving all other parameters at a set of “standard values.”

4D APSY-HNCOCA experiment with the protein TM1290

As an experimental reference, a 4D APSY-HNCOCA experiment was recorded with the 12 kDa protein TM1290, of which the sequence-specific resonance assignments are known (Etezady-Esfarjani et al. 2003, 2004). 13 2D projections were measured, with the projection angles $(\alpha, \beta) = (90^\circ, 0^\circ), (0^\circ, 0^\circ), (0^\circ, 90^\circ), (\pm 60^\circ, 0^\circ), (0^\circ, \pm 60^\circ), (90^\circ, \pm 60^\circ)$ and $(\pm 20^\circ, \pm 70^\circ)$. The interscan delay was set to 670 ms. In the direct dimension, 1,024 complex points were recorded with an acquisition time of 97 ms, and prior to Fourier transformation the FID was multiplied with a 75° -shifted sine bell (De Marco and Wüthrich 1976) and zero-filled to 2,048 complex points. In the indirect dimension, 36 complex points were measured, and the data was multiplied with a 75° -shifted sine bell and zero-filled to 256 complex points before Fourier transformation. The sweep widths were 1,750, 1,500 and 4,000 Hz for $\omega_1(^{15}\text{N})$, $\omega_2(^{13}\text{C}')$ and $\omega_3(^{13}\text{C}^\alpha)$, respectively. A single transient was recorded per time point in the indirect dimension, so that each 2D projection was recorded in 1 min. The baseline was corrected using the IFLAT method in the $\omega_4(^1\text{H})$ -dimension (Bartels et al. 1995), and with 5th order polynomials in the $\omega_{1-3}(^{15}\text{N}, ^{13}\text{C}', ^{13}\text{C}^\alpha)$ -dimension. Signals were identified using the new GAPRO peak picker with $R_{\min} = 4.0$, and the following set of GAPRO parameters: $S_{\min} = 6$, $\Delta v_{\min} = 5.0$ Hz, $r_{\min} = 15.0$ Hz and $k = 200$ (Table 1). The calculation time with the presently introduced new version of GAPRO on a standard PC, running Linux with a 2.8 GHz Pentium 4 processor, was approximately 10 s, as compared to about 10 min when using the original implementation of GAPRO (Hiller et al. 2005) with the same hardware.

From the amino acid sequence of the TM1290 protein, 112 correlations are expected in the 4D APSY-HNCOCA experiment. In the 4D peak list generated by the algorithm GAPRO from the 13 2D projections, 110 of these correlations were contained. The missing two correlations involve the amide groups of residues 48 and 49, which are located in a loop and for which the NMR peaks are not observable in conventional spectra either, due to broadening by intermediate exchange (Etezady-Esfarjani et al. 2003, 2004). With the selected short measuring time, the intensity of the weakest NMR signals in the projections is comparable to the intensity of the thermal noise (Figs. 3, 4). Nonetheless, even the weakest of the 110 correlation peaks, linking the amide group of Gly 51 with the carbonyl- and α -carbons of His 50, was recognized by GAPRO as a true 4D correlation, whereas no false 4D

correlations appear in the peak list generated by GAPRO. This nicely illustrates that APSY makes use of the combined sensitivity of all the projections in the input, and does not require unambiguous identification of the individual peaks in each projection. In the experiment of Figs. 3 and 4, the simultaneous analysis of the complete set of 13 2D projections thus enabled reliable recognition of the correlation peaks on the background of the thermal noise.

Design of the evolution elements in APSY-NMR pulse sequences

The most critical aspect in the design of the evolution elements is that the time domain origin should be included in the sampling. It is therefore necessary to start with zero evolution time for all evolution elements, which establishes a common reference time point for all the projections. As a consequence, no phase correction has to be applied in the indirect dimension and aliased peaks do not change their sign. For zero evolution time, one usually has to introduce 180° refocusing pulses into the evolution periods. Where this is possible, it is advisable to use symmetric pairs of pulsed field gradients flanking these 180° pulses, in order to allow the suppression of artifacts created by imperfect 180° pulses without additional phase cycling. For improved resolution in the projected dimension, constant-time periods may have to be extended to semi-constant time elements; this is generally advisable when working with slowly relaxing systems, such as denatured proteins (Hiller et al. 2007).

General aspects of the processing of APSY-NMR data sets

During the processing of the projection spectra, pairs of data sets are combined before Fourier transformation, in order to obtain pure sine and cosine terms of the modulations created by the simultaneous incrementation of evolution times (Brutscher et al. 1995; Kupce and Freeman 2004). This ensures that projection spectra are obtained, rather than reduced-dimensionality spectra (Szyperski et al. 1993a, b). As a prerequisite for the GAPRO calculations, the processed spectra have to be properly phased and placed correctly in the higher-dimensional space. Phasing can become cumbersome or even impossible in the projected dimension. For this reason it is recommended to sample the time domain origin (see above), so that the phase has to be corrected only in the acquisition dimension. However, the digital fast Fourier transformation algorithm usually results in an even number of points in the spectrum, so that the frequency domain origin (carrier offset frequency) in the center of a spectrum measured in quadrature mode cannot be represented by a digital point. Therefore,

the projection spectrum should not be placed symmetrically about the frequency domain origin, but shifted by $\frac{1}{2}$ digital point.

Sensitivity for signal detection in APSY-NMR experiments

The signal intensity of a given multi-dimensional correlation peak can vary widely in the individual projections of an APSY experiment (Fig. 4). In the following we present expressions which describe the dependence of the signal-to-noise ratio of a given resonance on the projection angles, and we discuss how the insights obtained with this formalism can be used to optimize the performance of APSY-NMR experiments.

We consider an N -dimensional APSY-NMR experiment m , which is acquired at the frequency of a nuclear species A . Adapting the notation by Ernst et al. (1987) for APSY-NMR, each 2D projection of this N -dimensional experiment is described by an $(N-2)$ -dimensional vector of projection angles, $\vec{\varphi} = (\alpha, \beta, \gamma, \dots)$. We then make use of the fact that a 2D projection spectrum is the Fourier transform of a 2D cross-section through the time domain of the N -dimensional experiment m (Nagayama et al. 1978; Kupce and Freeman 2004), where it is oriented by two N -dimensional Cartesian vectors for the indirect and the direct dimensions, $\vec{p}_1(\vec{\varphi})$ and \vec{p}_2 . The N -dimensional time domain, (t_1, t_2, \dots, t_N) , is thus sampled at discrete points on the hyperplane $t \cdot \vec{p}_1(\vec{\varphi}) + t_N \cdot \vec{p}_2$ (Fig. 5a), where t is a time parameter along $\vec{p}_1(\vec{\varphi})$ in the $(N-1)$ -dimensional time domain subspace of the indirect dimensions, and t_N is the time axis for the direct detection. This sampling yields a complex time domain signal, $s_m(\vec{\varphi}, t, t_N)$. Since all cross-sections start at the time domain origin (Fig. 5a), the signal amplitude at the time origin of a given experiment m ,

$s_m(0)$, is independent of $\vec{\varphi}$ (Fig. 5b). It is a key feature of the following procedure that angle-dependent factors are separated from those that are angle-independent.

We factorize the time domain signal of a single resonance into expressions for the direct and the indirect dimensions. For a 2D projection of an APSY-NMR experiment m with projection angles $\vec{\varphi}$, we thus have

$$s_m(\vec{\varphi}, t, t_N) = s_m(0) \cdot s_m^e(\vec{\varphi}, t) \cdot \exp\{i\Omega(\vec{\varphi})t\} \cdot s_A^e(t_N) \cdot \exp\{i\Omega_A t_N\}. \quad (1)$$

In Eq. 1 the term $\exp\{i\Omega_A t_N\}$ describes the oscillation at the frequency Ω_A in the direct dimension, and $\exp\{i\Omega(\vec{\varphi})t\}$ is the oscillation at the frequency $\Omega(\vec{\varphi})$ in the indirect dimension. (The projection angle-dependence of the NMR frequencies is exploited by the algorithm GAPRO, which determines the resonance frequencies along the individual dimensions from the Fourier-transformed signal). The functions $s_A^e(t_N)$ and $s_m^e(\vec{\varphi}, t)$ are the signal envelopes in the direct and indirect dimensions, respectively. The envelope function in the direct dimension is usually an exponential decay with the relaxation rate of the acquisition nucleus A . The envelope function in the indirect dimension is a projection angle-dependent combination of the relaxation envelopes along the indirect dimension axes. The signal envelopes have a strong impact on the signal intensity in the projection spectra (Fig. 5b, c).

Before Fourier transformation, the time domain signal $s_m(\vec{\varphi}, t, t_N)$ is typically multiplied with a window function, $h_m(\vec{\varphi}, t, t_N)$, which can be individually adapted for each experiment m . Subsequent Fourier transformation with respect to t and t_N yields a resonance peak with frequencies $\Omega(\vec{\varphi})$ and Ω_A in the 2D projection spectrum (Ernst et al. 1987). Its intensity at the peak maximum, $S_m(\vec{\varphi})$, is given by

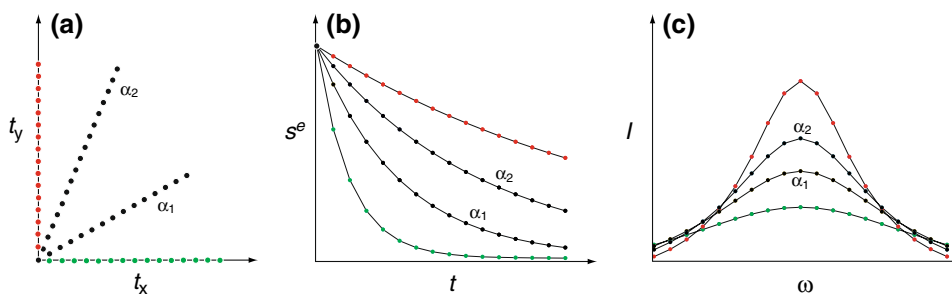


Fig. 5 Schematic representation of data sampling and processing for projections of two indirect dimensions x and y . (a) The time domain is sampled along cross-sections through the origin. The green and red dots are sample points for projections along the time dimensions t_x and t_y , respectively. The black dots are sampling points for two projections at angles α_1 and α_2 , respectively. (b) Signal envelopes $s^e(t)$

along the time domain cross-sections. The relaxation rate of a given signal depends on its projection angle (see text). (c) After Fourier transformation, the NMR spectrum in the frequency space, $I(\omega)$, is obtained. The intensity at the peak maximum depends on the relaxation rate of the signal (see (b)) and on the processing parameters

$$S_m(\vec{\varphi}) = s_m(0) \cdot n(\vec{\varphi}) \frac{M(\vec{\varphi})M_A}{t_{\max}(\vec{\varphi})t_{\max,A}} \int_0^{t_{\max}(\vec{\varphi})} dt \int_0^{t_{\max,A}} dt_N s_m^e(\vec{\varphi}, t) \cdot s_A^e(t_N) \cdot h_m(\vec{\varphi}, t, t_N). \tag{2}$$

In Eq. 2, $M(\vec{\varphi})$ and M_A are the numbers of complex points in the indirect and direct dimensions, respectively, $t_{\max}(\vec{\varphi})$ and $t_{\max,A}$ are the maximal evolution times in the indirect and direct dimensions, respectively, and $n(\vec{\varphi})$ is the number of scans. $M(\vec{\varphi})$, $n(\vec{\varphi})$, $h_m(\vec{\varphi}, t, t_N)$ and $t_{\max}(\vec{\varphi})$ can be individually adapted for each projection angle $\vec{\varphi}$. Since the signal function in the 2D projection spectrum is the product of the direct and indirect dimensions, it is common practice to apply a filter consisting of the product of two different window functions for the indirect dimension and the direct dimension, $h_m(\vec{\varphi}, t)$ and $h_A(t_N)$, with $h_m(\vec{\varphi}, t, t_N) = h_m(\vec{\varphi}, t) \cdot h_A(t_N)$. The acquisition and processing parameters for the direct dimension can then be optimized independently of the projection angles.

We express the sensitivity of an NMR experiment by the signal-to-noise ratio, as defined by the ratio of the signal amplitude at the peak maximum and the root mean square amplitude of the thermal noise, $\sigma(\vec{\varphi})$. For 2D projection spectra,

$$\sigma(\vec{\varphi}) = \sqrt{2^q n(\vec{\varphi}) M(\vec{\varphi}) \overline{h^2(\vec{\varphi})} \overline{h_A^2} M_A \rho} \frac{1}{\sqrt{t_{\max,A}}}, \tag{3}$$

where $\overline{h^2(\vec{\varphi})}$ and $\overline{h_A^2}$ describe the weighting of thermal noise by the window functions (Ernst et al. 1987), ρ is the square root of the frequency-independent power spectral density (Ernst et al. 1987), q is the number of angles in the vector $\vec{\varphi}$ that differ from 0° or 90° , 2^q is the number of subspectra that need to be combined in order to obtain a pure projection spectrum (Brutscher et al. 1995; Kupce and Freeman 2004), and $n(\vec{\varphi})$ is the number of scans recorded for each of the subspectra. We then have for the sensitivity of a given projection, $[S/\sigma]_m(\vec{\varphi})$, as obtained as the ratio of the signal intensity (Eq. 2) and the noise (Eq. 3),

$$[S/\sigma]_m(\vec{\varphi}) = K_A \cdot s_m(0) \cdot f_m(\vec{\varphi}), \tag{4}$$

where the three terms K_A , $s_m(0)$ and $f_m(\vec{\varphi})$ represent, respectively, the impact of the detected spin type, the signal intensity at time zero, and the selection of the projection angles. These three factors are individually discussed in the following.

K_A accounts for the dependence of the experiment on the properties of the detected nuclear species, A , with

$$K_A = \frac{1}{\rho \sqrt{\overline{h_A^2} t_{\max,A}}} \int_0^{t_{\max,A}} dt_N s_A^e(t_N) h_A(t_N). \tag{5}$$

In Eq. 5, ρ accounts for the sensitivity of the probehead and the main polarizing magnetic field strength. For each

nucleus A and the NMR instrument used, the optimal acquisition time, $t_{\max,A}$, and the optimal window function, $h_A(t_N)$, can be found by maximizing the value of K_A .

Knowledge of the signal intensity of the experiment m at the time domain origin, $s_m(0)$, enables a comparison of the relative intrinsic sensitivities of different APSY-NMR experiments, and can thus help to identify high-sensitivity experiments. Values for $s_m(0)$ can be estimated either experimentally, for example, from comparison of the 1D NMR spectra of the nucleus A that correspond to the time domain origin of the different experiments, or from model calculations (Sattler et al. 1999). In echo-antiecho detected experiments (Palmer et al. 1991; Kay et al. 1992; Pervushin et al. 1997), the pure amplitude-modulated signals are obtained from a linear combination of the acquired real and imaginary signals prior to Fourier transformation. This combination leads to a signal enhancement by a factor of 2 (Eqs. 1 and 2) and an increase of the noise by a factor $\sqrt{2}$ (Eq. 3), resulting in an increase of the overall sensitivity by $\sqrt{2}$ (Eq. 4). This gain is in part offset because the prolonged pulse sequence leads to a reduced value of $s_m(0)$ (Eqs. 1, 2, and 4).

The dependence of the sensitivity of an experiment m on the projection angles and the acquisition and processing parameters is accounted for by $f_m(\vec{\varphi})$, with

$$f_m(\vec{\varphi}) = \frac{1}{(\sqrt{2})^q} \frac{\sqrt{n(\vec{\varphi}) \cdot M(\vec{\varphi})}}{\sqrt{\overline{h^2(\vec{\varphi})}}} \cdot \frac{1}{t_{\max}(\vec{\varphi})} \int_0^{t_{\max}(\vec{\varphi})} dt s_m^e(\vec{\varphi}, t) \cdot h(\vec{\varphi}, t). \tag{6}$$

Equation 6 provides a basis for optimizing the acquisition and processing parameters. Thereby, $f_m(\vec{\varphi})$ is largely governed by the signal envelope in the indirect dimension, $s_m^e(\vec{\varphi}, t)$. It can be shown that monoexponential relaxation in all the individual indirect dimensions results in a combined envelope that is also monoexponential, $s_m^e(\vec{\varphi}, t) = \exp\{-R_{2,m}^*(\vec{\varphi})t\}$, with an effective relaxation rate

$$R_{2,m}^*(\vec{\varphi}) = \vec{p}_1(\vec{\varphi}) \cdot \vec{R}_{2,m}. \tag{7}$$

In Eq. 7, $\vec{R}_{2,m}$ is an N -dimensional vector containing the transverse relaxation rates along all indirect dimensions, with $R_{2,m}^i = 0$ for constant-time evolution elements in the dimension i . Since the standard GAPRO analysis attaches equal weight to each projection spectrum, it is desirable to have similar sensitivities for all individual projection experiments. If the projection angle-dependence of $s_m^e(\vec{\varphi}, t)$ is known, Eq. 6 provides a basis for producing similar sensitivities for all the projections used in a given APSY experiment, since the user-defined parameters $n(\vec{\varphi})$, $M(\vec{\varphi})$, $h_m(\vec{\varphi}, t)$ and $t_{\max}(\vec{\varphi})$ can be individually adjusted for each projection experiment. The FID is characterized by $R_{2,m}^*(\vec{\varphi})$ for each set of projection angles, and the user-

defined parameters can be optimized based on similar considerations as for conventional NMR spectroscopy (Ernst et al. 1987; Wider 1998).

Overall, the formalism of Eqs. 1–7 enables the evaluation of the sensitivity of the individual projection measurements in a given APSY experiment, and thus provides a basis for selecting APSY-NMR experiments with intrinsic high sensitivity (illustrations are provided below in the section on APSY-NMR experiments for polypeptide backbone assignments), and for optimizing the overall sensitivity of a given APSY-NMR experimental scheme.

Selection of projection angles

In principle, GAPRO does not impose any restrictions on the number of projections and the combinations of projection angles, except that, on fundamental grounds, the total number of projections must be at least $N-1$, where N is the dimensionality of the parent experiment, and each indirect dimension needs to be evolved at least once in each set of projections. On this basis, the selection of projection angles for APSY-NMR experiments should be guided by two main considerations. Firstly, the projection angles sampled by the set of projections used in a given experiment should be distributed about evenly in the time domain. Secondly, projections with large q -values are to be disfavored, since the sensitivity for the recording of the 2D projection spectra is proportional to $2^{-q/2}$ (Eq. 6), where q is the number of projection angles that differ from 0° or 90° . It is further recommended that the decrease in sensitivity in projection experiments that must be recorded with higher q -values is compensated by acquiring a larger number of transients ($n(\vec{\varphi})$ in Eq. 6) for these measurements.

An additional improvement may be achieved with the use of dispersion-optimized projection angles, in particular if the sweep widths of the different indirect dimensions are significantly different. Dispersion-optimized projection angles adjust the sweep widths of the different indirect dimensions to the same size, and can thus contribute to eliminating chemical shift overlap. The dispersion-optimized “matching” projection angle α^* for two dimensions, i and j , with sweep widths SW_i and SW_j (Fig. 6) is given by

$$\tan \alpha^* = \frac{SW_i}{SW_j}. \quad (8)$$

For example, if the sweep widths of two dimensions differ 11-fold, then $\alpha^* = 84^\circ$ (Fig. 7), so that a set of three projection angles with values of 60° , 84° and 87° would be a good choice, whereas a seemingly more symmetric selection, e.g. with angles of 22.5° , 45° and 67.5° can lead to two basically identical projections. Similar expressions to Eq. 8 can be derived for combinations of three or more dimensions.

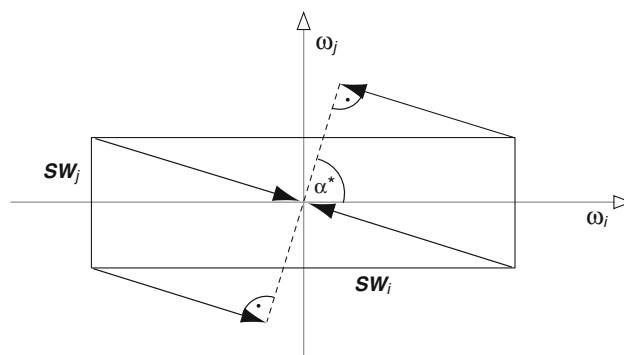


Fig. 6 Definition of the matching projection angle α^* by the sweep-widths of two indirect dimensions ω_i and ω_j (Eq. 8). The spectrum (black rectangle) is spanned by the sweep widths SW_i and SW_j . The angle α^* generates a projection where the two different sweep widths cover the same frequency range (dashed line)

Selection of the number of projections

Choosing a minimal number of projections for a given experiment is important whenever it is of overriding interest to collect an informative data set with minimal measurement time. In general practice, based on our experience so far, we tend to handle this parameter rather liberally. A good decision on the number of projections must be based on consideration of the type of APSY-NMR experiment used, the expected number of correlation peaks per amino acid residue, the size and type of protein under study, the choice of the projection angles, and the problem to be solved with the chosen experiment. For example, in the experiment of Figs. 3 and 4, 13 projections were sufficient to yield a good separation of correlation signals and artifacts, so that GAPRO could generate a complete peak list with 110 peaks. The number of projections could be reduced if a less complete result were sufficient for the intended purpose. The fact that a high-quality result could be obtained with a small number of projections in this experiment is also due to the large ^1H chemical shift dispersion (Wüthrich 1986) in the protein TM1290. For polypeptides with smaller chemical shift dispersion, such as denatured or otherwise non-globular proteins, a twofold to threefold higher number of projections may be needed for obtaining comparable results. In our laboratory we use such rather general considerations as guidelines for deciding on the number of projections in a given APSY-NMR experiment. In doubt, we record rather too many than too few projections, since ensuing redundancies will positively affect the robustness of the GAPRO output, and lead to improved precision of the chemical shift measurements for the N -dimensional correlation peaks (Hiller et al. 2005). Representative illustrations of the choice of the number of projections used in particular experiments are given below in the section on resonance assignments.

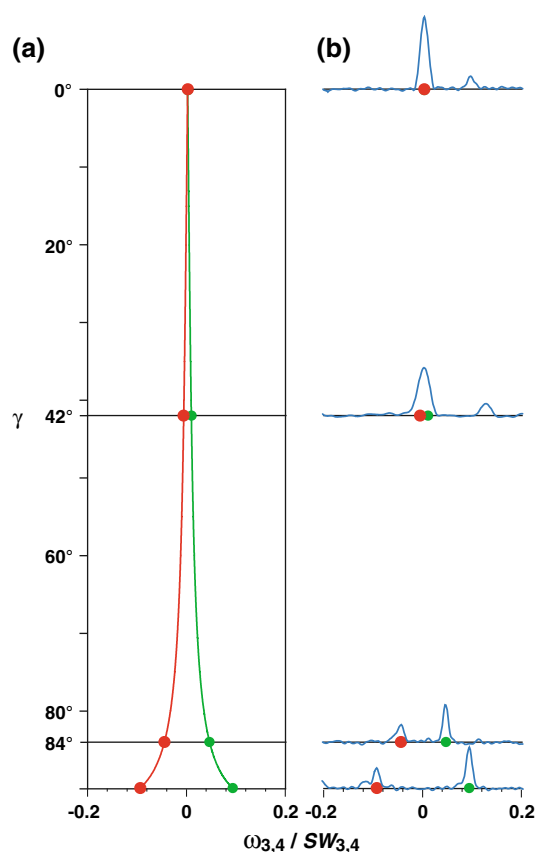


Fig. 7 Simulated and experimental peak positions in (0° , 90° , γ , 0° , 0°)-projections of a 7D APSY-seq-HNCO(CA)CBCANH experiment with the urea-unfolded protein OmpX (outer membrane protein X from *E. coli*; data from Hiller et al. 2007). **(a)** Simulation for two peaks (red and green lines) with degenerate C^β chemical shifts, as seen at $\gamma = 0^\circ$, but different CO chemical shifts, as seen at $\gamma = 90^\circ$. At $\gamma = 0^\circ$, only the C^β chemical shifts contribute to the $\omega_{3/4}$ projection, and at $\gamma = 90^\circ$ only the CO chemical shifts contribute. The red and green dots indicate the calculated peak positions at $\gamma = 0^\circ$, 90° , $84^\circ = \alpha^*$ (Eq. 8), and $42^\circ = 0.5 \cdot \alpha^*$. The spectra are normalized with the sweep widths ($\cos\gamma \cdot 11,000 \text{ Hz} + \sin\gamma \cdot 1,000 \text{ Hz}$), as expressed in the drawing by $\omega_{3,4}/SW_{3,4}$. **(b)** Experimental cross-sections from projection spectra at $\gamma = 0^\circ$, 42° , 84° and 90° . As a reference, the red and green dots indicate the peak positions predicted from **(a)**

Finally it should be mentioned that an automatic convergence scheme can be installed that stops the data recording as soon as a preset number of peaks have been resolved.

Selection of the sweep widths for individual projections

In a conventional NMR experiment, the spectrum is spanned by the user-defined sweep widths for each dimension, which can be selected with the use of a default value evaluated for each nucleus. For example, for proteins these default values may be based on the statistical chemical shift distributions evaluated in the BMRB data base (Seavey et al. 1991). For the recording of projection spectra, an appropriate sweep width for a projection with

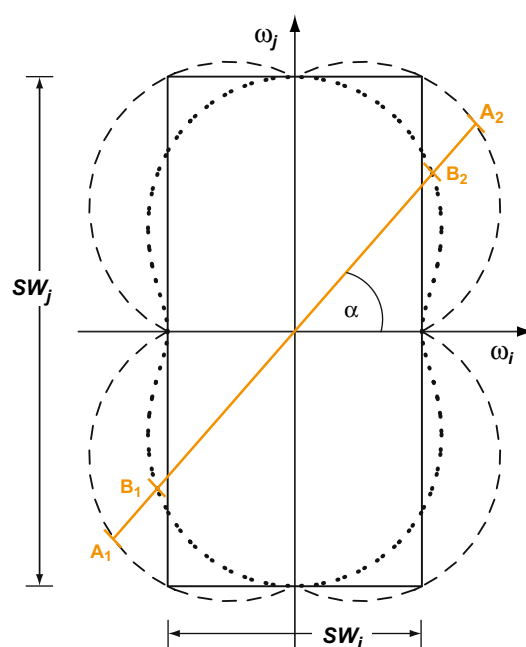


Fig. 8 Calculation of the projection sweep width. The black rectangle outlines a conventional 2D sub-space of the N -dimensional spectrum, which is spanned by default values of the sweep widths SW_i and SW_j for the nuclei in the indirect ω_i and ω_j dimensions. For a given projection with projection angle α (orange line) the value SW for the sweep width can be calculated using the conventional approach of Eq. 9, resulting in the segment from point A_1 to point A_2 . Alternatively, when using the refined approach of Eq. 11, one obtains the shorter segment from B_1 to B_2 . The dashed and dotted lines indicate the trajectories for all possible values of the projection angle α of the locations of the points $A_{1,2}$ (Eq. 9) and the points $B_{1,2}$ (Eq. 11), respectively

given projection angles, SW , can be calculated from these default values (SW_i and SW_j in Fig. 8) for the different nuclei in the individual indirect dimensions. A straightforward approach is the “projected rectangle” (Fig. 8), which is guided by the consideration that all expected correlation peaks should be contained in the projection spectrum without aliasing,

$$SW = \sum_i^{N-1} SW_i \cdot p_i \quad (9)$$

In Eq. 9, the SW_i 's are default values for the different nuclei in the individual indirect dimensions, and p_i represents the coordinates of the projection vector. In practice, this approach may lead to unnecessarily large sweep widths and correspondingly reduced digital resolution, since the chemical shift distributions in any two different dimensions are only weakly correlated.

A more refined sweep width selection can be obtained by calculating the standard deviation for the distribution of chemical shifts in a given projection (Fig. 8). Assuming that the chemical shift distributions in the individual dimensions i are uncorrelated, with standard deviations σ_i ,

the chemical shift distribution in a given projection spectrum has an overall standard deviation, σ , given by

$$\sigma = \sqrt{\sum_i^{N-1} (\sigma_i \cdot p_i)^2}. \quad (10)$$

With the additional assumption that the chemical shifts are uniformly distributed in the statistically defined range, the fraction of peaks can be evaluated which is contained in any frequency interval that is proportional to the standard deviation. For example, 90% of the chemical shifts are within the interval $\pm 1.64 \sigma$, or 95% of the shifts are contained within $\pm 1.96 \sigma$. Thus, if the carrier frequency and the sweep widths in each dimension are adjusted to the chemical shift distribution in such a way that a given fraction of peaks, ϕ , is located within the spectral range covered, then the same value of ϕ is maintained in all projections, with the sweep widths determined by Eq. 11

$$SW = \sqrt{\sum_i^{N-1} (SW_i \cdot p_i)^2}. \quad (11)$$

Equation 11 leads to projection spectra with optimal use of the spectral space for high digital resolution, provided that the individual sweep widths SW_i correspond closely to the actual chemical shift dispersion in the protein under study.

Optimizing the GAPRO parameters for a given experiment

Among the input variables of the geometric algorithm GAPRO, three parameters have a dominant effect on the result of the spectral analysis (Table 1). The selection of the minimal support is most important, since all candidate signals with a support $S \geq S_{\min}$ will be included in the final peak list. Figure 9 shows the variation of the peak list resulting from the 4D APSY-HNCOCA experiment of Figs. 3 and 4 when different values of S_{\min} are used. For S_{\min} between 3 and 8, the final result contains all the 110 expected peaks. However, with $S_{\min} = 3$, GAPRO also selected 27 artifacts, and the result obtained with $S_{\min} = 4$ includes two artifacts, whereas with S_{\min} values from 5 to 8, a complete and artifact-free peak list was obtained. If the number of projections was increased, the range of S_{\min} values that yielded a complete and artifact-free peak list was increased (data not shown), which can be readily rationalized by the fact that additional projections contribute more support to NMR signals than to artifacts. As a general guideline, it is advisable to set S_{\min} to about one-third of the number of input projections, and to keep $S_{\min} > (N + 2)$ for an N -dimensional experiment with 2D projections.

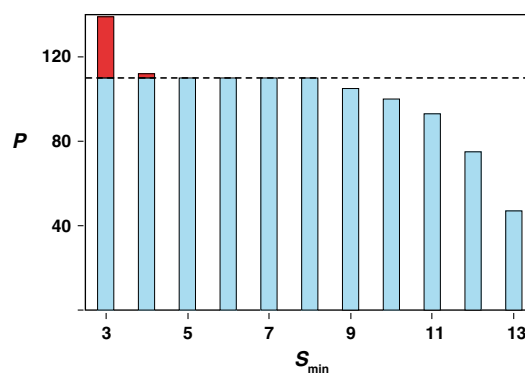


Fig. 9 Dependence of the number of 4D APSY-HNCOCA peaks of TM1290 calculated by GAPRO, P , on the value chosen for the GAPRO parameter S_{\min} . Blue bars represent the number of identified correlation peaks, red bars the number of artifacts maintained in the final peak list and the dotted line is a reference indicating the expected 110 peaks

The two additional key parameters are the intersection tolerances for the direct and indirect dimensions, Δv_{\min} and r_{\min} . The choice of the values for these parameters, which have to be adjusted to the digital resolution, the signal line widths and the sensitivity of a given experiment, affects the completeness of the output peak list from GAPRO. As a general guideline, we recommend to use one to two times the respective digital resolution in the 2D projection spectrum as values for Δv_{\min} and r_{\min} . For example, the values $r_{\min} \geq 5.0$ Hz and $\Delta v_{\min} \geq 2.4$ Hz yielded a complete peak list for the experiment of Figs. 3 and 4 (Fig. 10), which was recorded with a digital resolution of 4.7 Hz in the direct dimension, and digital resolutions between 5.6 and 15 Hz in the indirect dimensions, depending on the projection angles used for the individual 2D recordings.

Application of the newly introduced GAPRO secondary peak filter had no impact on the final result obtained with S_{\min} values in the range 5–8, which yielded a complete, artifact-free peak list from the 4D APSY-HNCOCA experiment of Figs. 3 and 4 (Fig. 9). The potential impact is illustrated by its use, with $n = 0$ and $R_{\text{single}} = 0$ (Table 1), to validate the results obtained with S_{\min} values of 3 and 4 (Fig. 9). In these data sets, all the artifacts were removed, and identical results were obtained as with the S_{\min} values of 5–8 without use of the secondary peak filter.

APSY-NMR experiments for protein backbone assignments

As a practical application of the general guidelines provided so far in this paper, we describe here experimental details of APSY-NMR measurements that we recommend for collecting the input data needed to obtain polypeptide backbone assignments in globular proteins. The selection of the APSY-NMR experiments is based on general

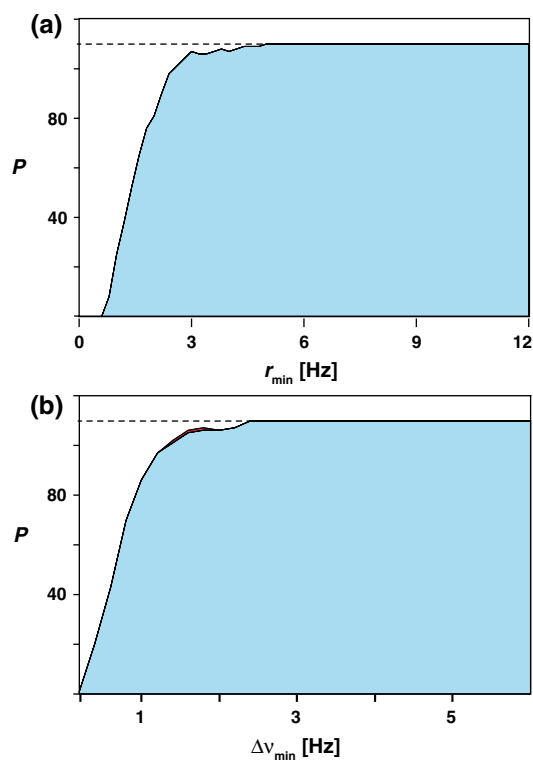


Fig. 10 Dependence of the number of 4D APSY-HNCOCA peaks of TM1290 calculated by GAPRO, P , on two operator-selected GAPRO parameters. **(a)** Impact of the intersection tolerance of subspaces in the indirect dimension, r_{\min} . **(b)** Impact of the intersection tolerance of subspaces in the direct dimension, Δv_{\min} . Blue areas represent correctly identified correlation peaks, red areas represent artifacts maintained in the final peak list, and the dotted line is a reference indicating the expected 110 peaks

considerations on resonance assignment procedures, and on an evaluation of the relative sensitivities for signal detection of different experiments that could provide the desired information. These selection criteria are described in the Appendix 1.

The discussion on polypeptide backbone resonance assignments in the Appendix 1 shows that the desired data can be obtained from suitable combinations of two or three 4- and 5-dimensional APSY-NMR experiments (for a survey see Fig. A1). The Table 2 lists a selection of experiments which can thus be combined, and it further compares estimates of the relative sensitivities for signal detection of the different individual experiments. The Table 3 provides recommended carrier frequencies to be used in these experiments. The table further lists default values for the sweep widths of the nuclei in the different relevant atom positions, from which optimized sweep widths for the individual 2D projection experiments can be evaluated using either Eq. 9 or, preferably, Eq. 11.

From among the different recommended combinations of APSY-NMR experiments, the three measurements

Table 2 Relative sensitivities, $s_m(0)$ (Eq. 4), of correlations in APSY-NMR experiments for polypeptide backbone assignments

Correlated atoms	Experiment	$s_m(0)$
H^N-N	2D [$^{15}N, ^1H$]-HSQC	100
$H^N-N-C^\alpha-C^\beta$ seq	5D-APSY-HNCOCACB	11
$H^N-N-C^\alpha-C^\beta$ intra	5D-APSY-CBCACONH ^a	10–4
	4D-APSY-HNCACB	7
	4D-APSY-CBCANH ^a	5–2
$H^N-N-C^\alpha-H^\alpha$ seq	5D-APSY-HNCOCAHA	17
	5D-APSY-HACACONH	15
$H^N-N-C^\alpha-H^\alpha$ intra	4D-APSY-HNCAHA	11
	4D-APSY-HACANH	8
$H^N-N-C^\alpha-C'$ seq	4D-APSY-HNCOCA	32
$H^N-N-C^\alpha-C'$ intra	4D-APSY-HNCACO	7

The data were calculated with the formalism described in the Appendix 1 for a protein with a rotational correlation time of 10 ns. As a reference, the sensitivity of the 2D [$^{15}N, ^1H$]-HSQC experiment was set to 100

seq, sequential connectivity and intra, intraresidual connectivity

^a For these experiments, the intensity depends on the amino acid side chain. The ranges given indicate the minimal and maximal values expected for the 20 common amino acid residues

Table 3 Recommended carrier frequencies and sweep widths for APSY-NMR experiments designed for polypeptide backbone resonance assignments in proteins

Atom position ^a	Carrier frequency (ppm) ^b	SW_i (ppm) ^c	
		$\phi = 90\%$ ^d	$\phi = 95\%$ ^d
$^1H^N$	8.29	2.63	3.14
^{15}N	116.1	26.1	28.5
$^{13}C'$	176.1	10.5	11.9
$^1H^\alpha$	4.41	2.37	2.73
$^{13}C^\alpha$	55.3	24.1	25.4
$^1H^\beta$	2.83	3.80	4.00
$^{13}C^\beta$	44.3	56.7	57.9

^a As a starting platform for the evaluation of the data presented here, the chemical shift range for each atom position was determined from the statistical data in the Biological Magnetic Resonance Data Bank (Seavey et al. 1991)

^b The carrier frequencies are relative to the chemical shift of DSS

^c SW_i represents the default value for the sweep width which are used in Eqs. 9 and 11. It is centered about the carrier frequency

^d ϕ is the fraction of all resonances predicted to be contained within the spectral range SW_i without aliasing; the values for SW_i have been chosen so as to satisfy this criterion

providing the HA–CA(CB) information (Fig. A1) were used to obtain backbone resonance assignments for TM1290, which is the same protein as was studied in the experiment of Figs. 3 and 4. The Table 4 presents key parameters used for the recording of these three experiments, i.e., 4D APSY-HACANH, 5D APSY-HACACONH

Table 4 Key experimental parameters used when recording the three APSY-NMR experiments for the collection of the HA-CA(CB) data affording the polypeptide backbone assignment of the protein TM1290

Parameters survey of results	4D APSY-HACANH	5D APSY-HACACONH	5D APSY-CBCACONH
2D Projections ^a	19	18	20
Recording time ^b	20 min	30 min	25 min
¹ H frequency	750 MHz	750 MHz	750 MHz
S_{\min} ^c	8	8	8
Correlations expected ^d	109 (109)	109	110
Correlations observed ^d	88 (109)	109	110

The lower part presents a survey of the results obtained

The measurements were performed with a 1 mM TM1290 solution at 25°C

^a Number of independent 2D projection recordings using the projection angles of Table 5

^b The same recording time was used for each 2D projection experiment

^c The GAPRO parameter S_{\min} is defined in Table 1

^d For all three experiments the number of sequential connectivities (Fig. A1) is given, and for 4D APSY-HACANH the intraresidual connectivities, which are the desired data from this measurement, are listed in parentheses. The important correlations in the input for the automated assignment are indicated in bold

and 5D APSY-CBCACONH. This includes the number of projections recorded, the instrument time used, and the value of the GAPRO parameter S_{\min} . The projection angles and the projection angle-dependent sweep widths used for each of the three experiments are listed in Table 5. For 4D APSY-HACANH and 5D APSY-HACACONH the projection-dependent sweep widths were calculated from the default values in Table 3 with the use of Eq. 11, whereas Eq. 9 was used for 5D APSY-CBCACONH, since the carrier frequencies of $\omega_1(C^\beta)$ and $\omega_2(C^\alpha)$ have to be identical and the assumptions leading to Eq. 11 are thus not satisfied in this experiment. Additional details on the experimental parameters and the data processing: In the direct dimension, 1,024 complex points were recorded with an acquisition time of 97 ms, and prior to Fourier transformation the FID was multiplied with a 75°-shifted sine bell (De Marco and Wüthrich 1976) and zero-filled to 2,048 complex points. In the indirect dimension, 64, 64 and 90 complex points were recorded for the 4D APSY-HACANH, 5D APSY-HACACONH and 5D APSY-CBCACONH experiments, the data was multiplied with a 75°-shifted sine bell, and the FIDs were zero-filled to 128, 128 and 256 complex points, respectively. The baseline was corrected using the IFLAT method in the direct dimension (Bartels et al. 1995), and with 5th order polynomials in the indirect dimension. The other GAPRO parameters (Table 1) were $R_{\min} = 5.0$, $\Delta v_{\min} = 5.0$ Hz, $r_{\min} = 15.0$ Hz and $k = 400$. For each data set, the GAPRO calculation time was shorter than 1 min when using a standard PC running Linux with a 2.8 GHz Pentium 4 processor.

Some special data handling was used for particular amino acid residues. In the data from the 5D APSY-CBCACONH experiment, a standard GAPRO filter routine removed peaks

arising from glutamine and asparagine, i.e., peaks located within 1 ppm from the BMRB values (29.2, 33.7, 179.7, 111.89 and 7.03/7.21 ppm for Gln C^β , C^γ , C^δ , N^ϵ and $H^{\epsilon 1/2}$; 53.5, 38.7, 176.7, 112.8 and 7.34/7.14 ppm for Asn C^α , C^β , C^γ , N^δ and $H^{\delta 1/2}$) (Seavey et al. 1991). In the same data set, peaks from glycine residues, which contain the C^α chemical shift in the first and the second dimension, were identified by GAPRO and the apparent C^β chemical shift was removed.

In the 4D APSY-HACANH and 5D APSY-HACACONH experiments, glycine residues with non-degenerate H^α chemical shifts give rise to two peaks which are degenerate in all dimensions except for the HA dimension. The GAPRO analysis of the projection spectra with the commonly used parameters (Tables 1, 4) often reports only one of these two peaks, since all projections with projection angles that do not resolve the HA dimension do not contribute to the support, S , for the second glycine peak. For this peak, S may then be below the minimal support, S_{\min} (Table 4), so that it will not appear in the final peak list (Hiller et al. 2005). Since the absence of the second peak does not interfere with the backbone assignment, no special handling of this situation was implemented.

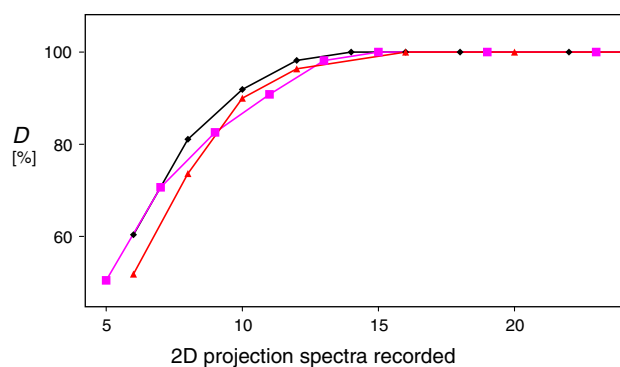
The lower part of Table 4 lists the numbers of expected and observed correlations for the three experiments used in the HA-CA(CB) approach. It should be noted that the 4D APSY-HACANH experiment was used here to obtain the intraresidual C^α -NH connectivity, although it yields also interresidual C^α -NH correlations. However, since the $^2J_{N,CA}$ -coupling can be quite small (Table A1), the intensity of some sequential peaks tends to be below the noise level in the projection spectra. As a consequence, in the experiment with TM1290 only about 80% of the sequential peaks were detected, which are redundant with sequential

Table 5 Values of the projection angles (α , β , γ) and the projection angle-dependent sweep widths (SW) used when collecting the HA–CA(CB) data set of Table 4. (a) 4D APSY-HACANH. (b) 5D APSY-HACACONH. (c) 5D APSY-CBCACONH

α (deg)	β (deg)	γ (deg)	SW (Hz)
(a)			
90	0		5,500
0	0		2,400
0	90		2,100
0	± 49		2,234
90	± 70		2,726
± 24	0		3,132
90	± 30		4,877
± 60	0		4,912
0	± 25		2,350
± 24	± 46		2,649
(b)			
0	0	0	2,400
0	0	90	2,100
90	± 22	± 42	2,572
0	0	± 49	2,234
0	± 24	0	3,132
± 48	0	0	2,292
90	0	± 46	2,149
90	0	0	2,200
0	90	0	5,500
0	90	± 69	2,780
(c)			
0	0	0	2,200
0	0	90	11,500
0	90	0	9,000
90	0	0	2,400
0	0	± 11	4,316
0	± 14	0	4,269
90	0	± 12	4,701
0	90	± 38	14,172
90	± 15	± 11	6,815
± 43	0	± 8	4,812

C^α –NH connectivities obtained from the 5D APSY-HACACONH experiment. Using the boldly printed correlations from the HA–CA(CB) combination of APSY-NMR experiments (Table 4) as input for the automatic assignment program GARANT (Bartels et al. 1997), all the detectable TM1290 backbone resonances (Etezady-Esfarjani et al. 2003) were correctly assigned.

The resonance assignment of TM1290 provides also an illustration of the impact on the completeness of the data sets from the number of 2D projection experiments recorded. In generating the Fig. 11 we made use of the fact that when

**Fig. 11** Impact of the number of 2D projections recorded when using the three experiments 4D APSY-HACANH (magenta), 5D APSY-HACACONH (black) and 5D APSY-CBCACONH (red) for collecting the data needed to obtain sequence-specific polypeptide backbone resonance assignments for the protein TM1290 (see text). For each of the three experiments, the plot shows the percentage of the expected number of correlations, D , observed with a given number of 2D projection spectra, which were recorded in the order of the listing in Table 5

using a combination of multiple experiments, such as the HA–CA(CB) group, the individual measurements can be recorded in an “interleaved fashion”, so that the number of 2D projections is augmented by one at a time across the group of experiments. In the study of TM1290, about 95% of the expected correlations for each of the three experiments were obtained with the first twelve 2D projections (Table 5). All expected correlations were obtained with sixteen or more projections, and therefore the numbers of projections listed in Table 4 were used. This type of consideration was used in developing the aforementioned convergence criteria, which automatically prevent the recording of additional projection spectra when the observed number of correlations reaches a user-imposed minimal value (unpublished results).

Conclusions

We hope that this paper may serve as a practical guide for a growing number of spectroscopists using APSY-NMR with proteins and possibly other biological macromolecules. The text is fully focused on technical aspects that apply specifically to APSY-NMR spectroscopy, including the selection of particular experiments, the optimization of the set-up of given experiments, and recommendations for complete and reliable spectral analysis with a new version of the GAPRO algorithm. Clearly, all considerations on optimization of conventional heteronuclear multidimensional NMR experiments with proteins must also be taken into account for obtaining high-quality results.

The experience from our work so far indicates for the experiments used to obtain scalar coupling-based sequential backbone assignments in ^{13}C , ^{15}N -labeled proteins, that

the sensitivity for signal detection rather than overcrowding the 2D projection spectra will be a limiting factor in practical applications of APSY-NMR. Considering the coherence transfer pathways of the key experiments, it appears that the improvements of the sensitivity should be achievable primarily with optimization of the instrumentation at field strengths corresponding to ^1H frequencies around 600 MHz, including the use of cryogenic probeheads, rather than with the use of higher magnetic field strengths.

Acknowledgments We thank Olivier Duss and Dr Touraj Etezady-Esfarjani for the protein sample of TM1290, and Markus Basan and Christian Wasmer for discussions on optimizing the projection sweep widths and on matching of the projection angles. Financial support from the Schweizerischer Nationalfonds and the ETH Zürich through the NCCR “Structural Biology” is gratefully acknowledged.

Appendix 1

Here, we describe the strategy followed for the selection of the APSY-NMR experiments to be used for sequence-specific polypeptide backbone resonance assignments in proteins (Tables 2, 4). This includes the criterion that the experiment must yield the heteronuclear correlations needed for the resonance assignments, and among the different possible experiments the one with the highest sensitivity is then selected.

APSY-NMR and polypeptide backbone resonance assignments

The process of sequence-specific resonance assignment for the polypeptide backbone includes the generation of molecular fragments from experimental scalar coupling correlations and chemical shift comparisons, and the mapping of these fragments onto the known amino acid sequence (Wüthrich 1986; Güntert et al. 2000; Moseley et al. 2001; Baran et al. 2004). In APSY-NMR experiments the magnetization transfer pathways cover short stretches of the polypeptide backbone and combine nuclei contained in these fragments into a single correlation (Fig. A1). Mutual overlap of the small molecular fragments defined by such correlations from different experiments is exploited for generating longer fragments, which can then be matched with the amino acid sequence.

When using exclusively signal acquisition on the amide proton, the amide proton and amide nitrogen-15 chemical shifts are contained in each correlation, and can therefore serve to match correlations from different experiments for the generation of longer fragments. When using experiments with dimensions below 6 (Fiorito et al. 2006; Hiller et al. 2007), additional chemical shift matching is needed.

As illustrated in Fig. A1, the C^α atom is always available for this additional matching, and a second matching atom can be either HA, CB or CO. Thereby, the C^β chemical shift allows the unambiguous distinction between different amino acid types, thus increasing the reliability of sequence-specific assignments and allowing unambiguous assignments of smaller fragments (Güntert et al. 2000; Hiller et al. 2007). With the requirement that there must be two overlapping chemical shifts between adjoining correlations, groups of two or three four- and five-dimensional correlations can be devised (Fig. A1). To identify suitable pulse sequences to be used for each assignment strategy outlined in Fig. A1, we compared the relative sensitivities of different experiments that could, in principle, provide the desired correlations.

Sensitivity of APSY-NMR experiments

Sensitivity comparisons of different APSY-NMR experiments that are detected on the same nucleus can be based on a comparison of the signal intensities at the time domain origin, $s_m(0)$ Eq. 4. Estimates of $s_m(0)$ for different experiments were obtained with the following assumptions (Sattler et al. 1999): (1) All 90° and 180° pulses are considered to be ideal pulses. (2) Magnetization losses occur only by transverse relaxation and from incomplete evolution under J -couplings. (3) The relaxation rate of a product operator is written as the sum of the relaxation rates of the individual nuclei, and relaxation of longitudinal operators is neglected. This yields, for example, that $R_2(\text{H}^{\text{N}}_x\text{N}_y) = R_2(\text{H}^{\text{N}}) + R_2(\text{N})$ and $R_2(\text{H}^{\text{N}}_x\text{N}_z) = R_2(\text{H}^{\text{N}})$. (4) The steady-state polarization of single protons was assumed to

Table A1 J -couplings and relaxation rates R_2 used for the calculation of the relative sensitivities for selected APSY-NMR experiments

Coupling constant ^a	Value (Hz)	Nucleus	R_2 (in s^{-1})
$^1J_{\text{HN,N}}$	93	$^1\text{H}^{\text{N}}$	47 ^b
$^1J_{\text{N,CA}}$	10 (7–11)	$^1\text{H}^\alpha$, $^1\text{H}^\beta$	82 ^b
$^2J_{\text{N,CA}}$	7 (4–9)	$^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$	44 ^b
$^1J_{\text{N,CO}}$	14	$^{13}\text{C}'$	20 ^c
$^1J_{\text{CO,CA}}$	53	^{15}N	12 ^b
$^1J_{\text{HA,CA}}$	140		
$^1J_{\text{HB,CB}}$	132		
$^1J_{\text{CA,CB}}$	37		

For the $J_{\text{N,CA}}$ couplings, numbers in parentheses represent typical ranges in proteins. The transverse relaxation rates R_2 are for a protein with a rotational correlation time τ_c of 10 ns at a ^1H frequency of 600 MHz

^a From Wider (1998)

^b From Cavanagh et al. (1996)

^c From Allard and Härd (1997)

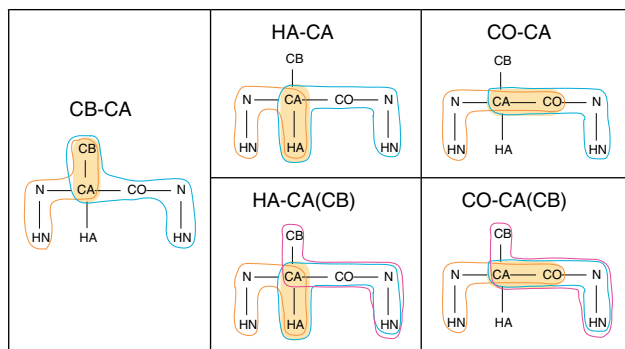


Fig. A1 Combinations of intraresidual and sequential correlations to be recorded with H^N -detected APSY-NMR experiments used to collect the data needed for polypeptide backbone assignment of ^{13}C , ^{15}N -labeled proteins. Each colored shape contains the nuclei correlated by a 4D or 5D experiment. The yellow areas contain the nuclei for which the individual correlations overlap. Chemical shift matching for these nuclei results in the generation of larger molecular fragments from the correlations obtained by the individual experiments. The notations used for the different groups of experiments reflect these matching chemical shifts, and indicate in parentheses additional important information for obtaining sequence-specific assignments (see text)

be independent of the atom position, so that $P(H^N) = P(H^\alpha) = P(H^\beta)$. We further used standard values for the J -coupling constants and the transverse relaxation rates R_2 (Table A1) and based the calculations on ideal pathways with transfer delays that optimize the transfer function.

As an illustration we describe the sensitivity evaluation for the 4D APSY-HACANH experiment (Fig. A2). The magnetization pathway of four INEPT steps, A–D, can be described as

$$H_z^\alpha \xrightarrow{A} 2H_z^\alpha C_z^\alpha \xrightarrow{B} 2C_z^\alpha N_z \xrightarrow{C} 2N_z H_z^N \xrightarrow{D} H_{x/y}^N$$

The step A transfers H^α polarization to $H^\alpha C^\alpha$ longitudinal 2-spin order, with the transfer function $\Gamma(A) = \sin(\pi^1 J_{HA,CA} \delta) \exp(-\delta R_2(HA))$, where δ is the INEPT transfer delay. If the rotational diffusion of the protein is characterized by a correlation time, τ_c , of 10 ns, the optimal value for δ is 3.3 ms and the resulting optimal transfer amplitude is $\Gamma(A) = 0.75$. The step B has the function $\Gamma(B) = 0.86 \cdot \sin(\pi^1 J_{N,CA} \rho) \cos(\pi^2 J_{N,CA} \rho) \cos$

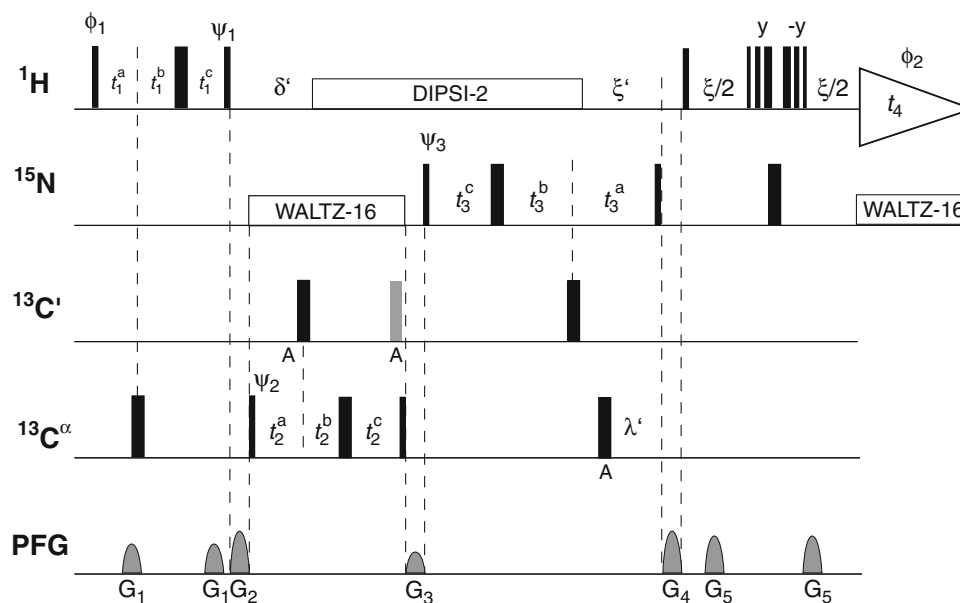


Fig. A2 Pulse sequence used for the 4D APSY-HACANH experiment. Radio-frequency pulses were applied at 4.7 ppm for 1H , 118.0 ppm for ^{15}N , 173.0 ppm for $^{13}C'$ and 54 ppm for $^{13}C^\alpha$. Narrow and wide bars represent 90° and 180° pulses, respectively. Pulses marked “A” were applied as Gaussian shapes, with 120 μs duration on a spectrometer with a 1H frequency of 500 MHz. All other pulses on $^{13}C'$ and $^{13}C^\alpha$ had rectangular shape, with a duration of $\sqrt{3}/(\Delta\omega(C^z, C') \cdot 2)$ and $\sqrt{15}/(\Delta\omega(C^z, C') \cdot 4)$ for 90° and 180° pulses, respectively. Pulses on 1H and ^{15}N were applied with rectangular shape and high power. The last six 1H pulses represent a 3-9-19 WATERGATE element (Sklernar et al. 1993). The grey pulse on $^{13}C'$ was applied to compensate for off-resonance effects of selective pulses (McCoy and Mueller 1992). Decoupling using DIPSI-2 (Shaka et al. 1988) on 1H and WALTZ-16 (Shaka et al. 1983) on ^{15}N is indicated by rectangles. t_4 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 , 800 μs , 18 G/cm; G_2 , 800 μs , 26 G/cm; G_3 , 800 μs , 13 G/cm; G_4 , 800 μs , 26 G/cm; G_5 , 800 μs , 23 G/cm. Phase cycling: $\phi_1 = \{x, x, -x, -x\}$, $\psi_2 = \{x, -x\}$, $\phi_2 = \{x, -x, -x, x\}$, $\psi_1 = y$, all other pulses = x . Fixed delays were $\delta' = 4.75$ ms and $\xi' = \xi = 5.4$ ms. Initial delays were $t_1^a = t_1^c = 1.8$ ms, $t_2^a = t_2^c = 11.5$ ms, $t_3^a = t_3^c = 11$ ms, $t_1^b = t_2^b = t_3^b = 0$ ms. The switched constant-time/semi-constant time elements were used as described (Fiorito et al. 2006), with $\lambda' = (t_3^a + t_3^c - t_3^b)/2$. Quadrature detection for the indirect dimension was achieved using the hypercomplex Fourier transformation method for projections (Brutscher et al. 1995; Kupce and Freeman 2004) with the angles ψ_1 , ψ_2 and ψ_3 (ψ_1 and ψ_2 were incremented and ψ_3 was decremented in 90° steps)

$(\pi^1 J_{CA,CB}\rho) \exp(-\rho R_2(CA))$, with $^1 J_{CA,CB} = 0$ Hz for glycine residues. This element has an optimal value of $\rho = 25$ ms for $\tau_c = 10$ ns, and the resulting transfer amplitude is $\Gamma(B) = 0.20$. The step C has the transfer function $\Gamma(C) = \sin(\pi^1 J_{N,CA}\chi) \cos(\pi^2 J_{N,CA}\chi) \exp(-\chi R_2(N))$, with an optimal value of $\chi = 27$ ms, yielding a transfer amplitude of $\Gamma(C) = 0.45$. For the step D, the transfer function is $\Gamma(D) = \sin(\pi^1 J_{HN,N}\xi) \exp(-\xi R_2(HN))$, with a theoretical optimum for ξ of 4.8 ms, yielding a transfer amplitude of $\Gamma(D) = 0.79$. From the four individual transfer functions, the overall sensitivity was calculated as $s_m(0) = P(HA) \cdot \Gamma(A) \cdot \Gamma(B) \cdot \Gamma(C) \cdot \Gamma(D)$, where $P(HA)$ is the equilibrium polarization of the α -proton.

The relative overall sensitivities calculated with this approach for a selection of APSY-NMR experiments that can provide the correlations of Fig. A1 are listed in Table 2.

References

- Allard P, Härd T (1997) NMR relaxation mechanisms for backbone carbonyl carbons in a ^{13}C , ^{15}N -labeled protein. *J Magn Reson* 126:48–57
- Baran MC, Huang YJ, Moseley HNB, Montelione GT (2004) Automated analysis of protein NMR assignments and structures. *Chem Rev* 104:3541–3555
- Bartels C, Güntert P, Wüthrich K (1995) IFLAT—a new automatic baseline-correction method for multidimensional NMR spectra with strong solvent signals. *J Magn Reson A* 117:330–333
- Bartels C, Güntert P, Billeter M, Wüthrich K (1997) GARANT—a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J Comput Chem* 18:139–149
- Brutscher B, Morelle N, Cordier F, Marion D (1995) Determination of an initial set of NOE-derived distance constraints for the structure determination of $^{15}\text{N}/^{13}\text{C}$ -labeled proteins. *J Magn Reson B* 109:238–242
- Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1996) *Protein NMR spectroscopy: principles and practice*. Academic, San Diego
- Coggins BE, Zhou P (2006) PR-CALC: a program for the reconstruction of NMR spectra from projections. *J Biomol NMR* 34:179–195
- Cornilescu G, Bahrami A, Tonelli M, Markley JL, Eghbalnia HR (2007) HIFI-C: a robust and fast method for determining NMR couplings from adaptive 3D to 2D projections. *J Biomol NMR* 38:341–351
- De Marco A, Wüthrich K (1976) Digital filtering with a sinusoidal window function—alternative technique for resolution enhancement in FT-NMR. *J Magn Reson* 24:201–204
- Eghbalnia HR, Bahrami A, Tonelli M, Hallenga K, Markley JL (2005) High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J Am Chem Soc* 127:12528–12536
- Ernst RR, Bodenhausen G, Wokaun A (1987) *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford University Press, Oxford
- Etezady-Esfarjani T, Peti W, Wüthrich K (2003) NMR assignment of the conserved hypothetical protein TM1290 of *Thermotoga maritima*. *J Biomol NMR* 25:167–168
- Etezady-Esfarjani T, Herrmann T, Peti W, Klock HE, Lesley SA, Wüthrich K (2004) Letter to the editor: NMR structure determination of the hypothetical protein TM1290 from *Thermotoga maritima* using automated NOESY analysis. *J Biomol NMR* 29:403–406
- Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. *J Biomol NMR* 35:27–37
- Güntert P, Salzmann M, Braun D, Wüthrich K (2000) Sequence-specific NMR assignment of proteins by global fragment mapping with the program MAPPER. *J Biomol NMR* 18:129–137
- Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24:171–189
- Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). *Proc Natl Acad Sci USA* 102:10876–10881
- Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. *J Am Chem Soc* 129:10823–10828
- Jaravine VA, Zhuravleva AV, Permi P, Ibragimov I, Orekhov VY (2008) Hyperdimensional NMR spectroscopy with nonlinear sampling. *J Am Chem Soc* 130:3927–3936
- Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663–10665
- Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc* 125:1385–1393
- Kupce E, Freeman R (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc* 126:6429–6440
- Kupce E, Freeman R (2006) Hyperdimensional NMR spectroscopy. *J Am Chem Soc* 128:6020–6021
- Lescop E, Brutscher B (2007) Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J Am Chem Soc* 129:11916–11917
- Luan T, Jaravine V, Yee A, Arrowsmith CH, Orekhov VY (2005) Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. *J Biomol NMR* 33:1–14
- Malmodin D, Billeter M (2005) Multiway decomposition of NMR spectra with coupled evolution periods. *J Am Chem Soc* 127:13486–13487
- McCoy MA, Mueller L (1992) Nonresonant effects of frequency-selective pulses. *J Magn Reson* 99:18–36
- Mishkovsky M, Kupce E, Frydman L (2007) Ultrafast-based projection-reconstruction three-dimensional nuclear magnetic resonance spectroscopy. *J Chem Phys* 127:034507
- Moseley HNB, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Methods Enzymol* 339:91–108
- Moseley HNB, Riaz N, Aramini JM, Szyperski T, Montelione GT (2004) A generalized approach to automated NMR peak list editing: application to reduced-dimensionality triple resonance spectra. *J Magn Reson* 170:263–277
- Nagayama K, Bachmann P, Wüthrich K, Ernst RR (1978) The use of cross-sections and of projections in two-dimensional NMR spectroscopy. *J Magn Reson* 31:133–148

- Orekhov VY, Ibraghimov I, Billeter M (2003) Optimizing resolution in multidimensional NMR by three-way decomposition. *J Biomol NMR* 27:165–173
- Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J Magn Reson* 93:151–170
- Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34:93–158
- Seavey BR, Farr EA, Westler WM, Markley JL (1991) A relational database for sequence-specific protein NMR data. *J Biomol NMR* 1:217–236
- Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broad-band decoupling—WALTZ-16. *J Magn Reson* 52:335–338
- Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators—application to spin decoupling. *J Magn Reson* 77:274–293
- Sklenar V, Piotto M, Leppik R, Saudek V (1993) Gradient-tailored water suppression for ^1H – ^{15}N HSQC experiments optimized to retain full sensitivity. *J Magn Reson A* 102:241–245
- Szyperski T, Atreya HS (2006) Principles and applications of GFT projection NMR spectroscopy. *Magn Reson Chem* 44:S51–S60
- Szyperski T, Wider G, Bushweller JH, Wüthrich K (1993a) 3D ^{13}C – ^{15}N -heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C – ^{15}N -double-labeled proteins. *J Biomol NMR* 3:127–132
- Szyperski T, Wider G, Bushweller JH, Wüthrich K (1993b) Reduced dimensionality in triple-resonance NMR experiments. *J Am Chem Soc* 115:9307–9308
- Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* 41:127–138
- Wider G (1998) Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Prog Nucl Magn Reson Spectrosc* 32:193–275
- Wider G, Dreier L (2006) Measuring protein concentrations by NMR spectroscopy. *J Am Chem Soc* 128:2571–2576
- Wüthrich K (1986) *NMR of proteins and nucleic acids*. Wiley, New York
- Zhang Q, Atreya HS, Kamen DE, Girvin ME, Szyperski T (2008) GFT projection NMR based resonance assignment of membrane proteins: application to subunit C of *E. coli* F_1F_0 ATP synthase in LPPG micelles. *J Biomol NMR* 40:157–163