

Automated NMR Assignment of Protein Side Chain Resonances Using Automated Projection Spectroscopy (APSY)

Sebastian Hiller,[†] Rosmarie Joss,[‡] and Gerhard Wider*

Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland

Received April 29, 2008; E-mail: gsw@mol.biol.ethz.ch

Abstract: This paper describes an automated method for sequence-specific NMR assignment of the aliphatic resonances of protein side chains in small- and medium-sized globular proteins in aqueous solution. The method requires the recording of a five-dimensional (5D) automated projection spectroscopy (APSY-) NMR experiment and the subsequent analysis of the APSY peak list with the algorithm ALASCA (Algorithm for local and linear assignment of side chains from APSY data). The 5D APSY-HC(CC-TOCSY)CONH experiment yields 5D chemical shift correlations of aliphatic side chain C–H moieties with the backbone atoms H^N, N, and C'. A simultaneous variation of the TOCSY mixing times and the projection angles in this APSY-type TOCSY experiment gives access to all aliphatic C–H moieties in the 20 proteinogenic amino acids. The correlation peak list resulting from the 5D APSY-HC(CC-TOCSY)CONH experiment together with the backbone assignment of the protein under study is the sole input for the algorithm ALASCA that assigns carbon and proton resonances of protein side chains. The algorithm is described, and it is shown that the aliphatic parts of 17 of the 20 common amino acid side chains are assigned unambiguously, whereas the remaining three amino acids are assigned with a certainty of above 95%. The overall feasibility of the approach is demonstrated with the globular 116-residue protein TM1290, for which reference assignments are known. For this protein, 97% of the expected side chain carbon atoms and 87% of the expected side chain protons were detected with the 5D APSY-HC(CC-TOCSY)CONH experiment in 24 h of spectrometer time, and all these resonances were correctly assigned by ALASCA. Based on the experience with TM1290, we expect that the approach presented in this work is routinely applicable to globular proteins with sizes up to at least 120 amino acids.

Introduction

Studies of protein structure and dynamics at atomic resolution by NMR require reliable sequence-specific assignment of NMR resonances.¹ Resonance assignment procedures for proteins usually establish the backbone assignment first and subsequently attach the side chain resonance assignment to this framework. Whereas several efficient automated and interactive methods for the backbone resonance assignment exist, side chain assignment still presents a substantial bottleneck. Common experiments used for side chain assignment are HCCH-COSY and HCCH-TOCSY experiments^{2–4} and the HC(CC-TOCSY)-CONH experiment.^{5–9} The HC(CC-TOCSY)CONH correlates side chains with backbone nuclei and in principle would be ideal to resolve chemical shift degeneracies between different identical side chains. However, a major drawback arises from the fact that the isotropic mixing time in TOCSY experiments can not be chosen such that all C–H moieties have sufficiently strong magnetization transfer amplitudes to be represented simultaneously in the same spectrum.⁴ Different solutions to this problem have been proposed, e.g., by selecting an intermediate

time at which a majority of the peaks have sufficient transfer, or by recording and possibly coadding two or more data sets measured with different mixing times.^{10,11}

In recent years, interactive programs have sped up the side chain assignment process, but it still takes considerable time and effort for acquisition and analysis of the data and to establish a fairly complete sequence-specific resonance assignment of a protein. Automation of this process would thus be highly

[†] Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

[‡] Present address: Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(2) Bax, A.; Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Ikura, M.; Kay, L. E. *J. Magn. Reson.* **1990**, *87*, 620–627.

(3) Bax, A.; Clore, G. M.; Gronenborn, A. M. *J. Magn. Reson.* **1990**, *88*, 425–431.

(4) Clore, G. M.; Bax, A.; Driscoll, P. C.; Wingfield, P. T.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 8172–8184.

(5) Montelione, G. T.; Lyons, B. A.; Emerson, S. D.; Tashiro, M. *J. Am. Chem. Soc.* **1992**, *114*, 10974–10975.

(6) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *FEBS Lett.* **1992**, *314*, 413–418.

(7) Grzesiek, S.; Anglister, J.; Bax, A. *J. Magn. Reson. B* **1993**, *101*, 114–119.

(8) Jiang, L.; Coggins, B. E.; Zhou, P. *J. Magn. Reson.* **2005**, *175*, 170–176.

(9) Sun, Z. Y.; Hyberts, S. G.; Rovnyak, D.; Park, S.; Stern, A. S.; Hoch, J. C.; Wagner, G. *J. Biomol. NMR* **2005**, *32*, 55–60.

(10) Celda, B.; Montelione, G. T. *J. Magn. Reson. B* **1993**, *101*, 189–193.

(11) Montelione, G. T.; Winkler, M. E.; Burton, L. E.; Rinderknecht, E.; Sporn, M. B.; Wagner, G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1519–1523.

desirable.¹² The combination of the assignment algorithm GARANT¹³ with automated peak picking of conventional three-dimensional COSY and TOCSY spectra has been shown to be a possible way to address this problem,¹⁴ but to achieve reasonably good side chain assignments, up to seven different conventional 3D experiments are necessary,¹⁵ requiring substantial spectrometer and calculation time.

To overcome the shortcomings of present interactive and automated procedures, we address the problem of side chain assignment using a fully automated method, automated projection spectroscopy (APSY). APSY is one of several recently introduced NMR techniques that give access to high-dimensional experiments not amenable with conventional techniques.^{16–21} APSY makes use of the acquisition technique to record projection spectra^{18,22} and then employs an algorithm, GAPRO, to identify local maxima in the projections and to calculate chemical shift correlations in the high-dimensional spectral space using geometrical considerations. The output of an APSY experiment is a correlation peak list of high quality which allows efficient and reliable subsequent use by computer algorithms. APSY has so far been successfully applied for automated backbone assignment of folded and unfolded proteins.^{23,24}

In the present work we describe the 5D APSY-HC(CC-TOCSY)-CONH experiment. The APSY technique enables the recording of five frequencies on the magnetization pathway with high robustness against artifacts and resulting in a very precise 5D chemical shift correlation list. Additionally, the problem of differential C–C magnetization transfers encountered in conventional TOCSY experiments is eliminated by varying the TOCSY mixing time together with the APSY projection angles. In this way, all aliphatic C–H moieties in the 20 amino acid types can be analyzed simultaneously. The result is a five-dimensional, artifact-free, and precise chemical shift correlation list, which allows straightforward sequence-specific resonance assignment by the new assignment algorithm ALASCA. We show the approach to be very time efficient and the linear algorithm to provide reliable and robust automated sequence-specific resonance assignments.

Materials and Methods

Description of the Algorithm ALASCA. The purpose of the ALASCA algorithm is the sequence-specific resonance assignment of aliphatic resonances using as input only the 5D APSY-HC(CC-

TOCSY)CONH correlation peak list and the known backbone assignment of the protein including the amino acid sequence. ALASCA obtains the assignment by matching the observed chemical shift correlations to statistical average values for the 20 amino acids in a linear procedure that consists of five steps.

Step 1. Each 5D APSY-HC(CC-TOCSY)CONH correlation is attributed to the residue, which has the nearest backbone chemical shifts in the 3D ($\omega(^{13}\text{C})$, $\omega(^{15}\text{N})$, $\omega(^1\text{H}^{\text{N}})$) space, corresponding to the dimensions ω_3 , ω_4 , and ω_5 of the 5D correlation list. The distances are calculated geometrically in the 3D coordinate system. Any correlation for which the nearest residue is farther away than a user-defined threshold $\Delta\omega_A$ is excluded from the assignment process. The value of $\Delta\omega_A$ reflects the precision of the backbone resonance assignment relative to the chemical shift measurement in the 5D APSY-HC(CC-TOCSY)CONH experiment. The result does not critically depend on $\Delta\omega_A$, and a standard value of $\Delta\omega_A = 150$ Hz is proposed for the calculation.

Step 2. All 5D correlations that have been attributed during step 1 to the same amino acid residue in the protein are put together into the “TOCSY peak group” of this residue, of which the amino acid type is known from the backbone assignment.

Step 3. Any two 5D correlations within one TOCSY peak group, which have the $\omega_2(^{13}\text{C})$ shift differing by 0.2 ppm or less, are assigned to a single CH_2 moiety. The limit of 0.2 ppm is large enough (>10 times the precision of the chemical shift measurement) to ensure that all detected pairs are identified.

Step 4. This step is applied only for amino acid residues, for which the C^α and/or C^β are known from the backbone assignment. For these residues, the corresponding 5D APSY-HC(CC-TOCSY)-CONH chemical shift correlations are identified from the TOCSY peak group and definitively assigned to the respective atoms.

Step 5. For each amino acid, the remaining correlations of the TOCSY peak group are assigned to the remaining side chain atoms by matching the chemical shifts of the 5D correlations to statistical values from the BMRB database.²⁵ Matching is achieved by finding the permutation σ_{ij} that minimizes the geometrical distance between the observed correlations and statistical average values described by the function $f(\sigma)$:

$$f(\sigma) = \sqrt{\sum_{k=1}^2 \sum_{i=1}^N \sum_{j=1}^M \sigma_{ij} (\omega_{ik}^{\text{P}} - \omega_{jk}^{\text{B}})^2} \quad (1)$$

where N is the number of remaining assignable C–H moieties in the amino acid, M is the number of remaining observed correlations and correlation pairs (see step 4), ω^{P} are the observed chemical shifts measured in Hz, ω^{B} are the database average chemical shifts in Hz, and the index $k = 1$ is for carbons and $k = 2$ for protons. For CH_2 pairs, the average chemical shift of the two protons is considered. A permutation σ is an $N \times M$ matrix with $\sigma_{ij} \in \{0;1\}$, with the restriction that in each row and in each column there can be at most one 1. Only permutations with $\sum_{i=1}^N \sum_{j=1}^M \sigma_{ij} = \min(N,M)$ are considered. $f(\sigma)$ is calculated for all possible permutations σ , and the permutation with the minimal $f(\sigma)$ value describes the sequence-specific resonance assignment.

We have termed the algorithm performing steps 1 to 5 above ALASCA, for “Algorithm for local and linear assignment of side chains from APSY data). It is implemented in the APSY environment which is available from www.apsy.ch.

Sample Preparation. A sample of [$U\text{-}^{13}\text{C}$, ^{15}N]-labeled *Thermotoga maritima* protein TM1290 in 20 mM phosphate buffer at pH 6.0, containing 5% D_2O and 0.1% NaN_3 , was produced as

- (12) Huang, Y. P. J.; Moseley, H. N. B.; Baran, M. C.; Arrowsmith, C.; Powers, R.; Tejero, R.; Szyperski, T.; Montelione, G. T. *Methods Enzymol.* **2005**, *394*, 111–141.
- (13) Bartels, C.; Güntert, P.; Billeter, M.; Wüthrich, K. *J. Comput. Chem.* **1997**, *18*, 139–149.
- (14) Malmödin, D.; Papavoine, C. H. M.; Billeter, M. *J. Biomol. NMR* **2003**, *27*, 69–79.
- (15) López-Méndez, B.; Güntert, P. *J. Am. Chem. Soc.* **2006**, *128*, 13112–13122.
- (16) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J. Biomol. NMR* **2003**, *27*, 165–173.
- (17) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385–1393.
- (18) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126*, 6429–6440.
- (19) Eghbalnia, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12528–12536.
- (20) Hiller, S.; Fiorito, F.; Wüthrich, K.; Wider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10876–10881.
- (21) Malmödin, D.; Billeter, M. *J. Am. Chem. Soc.* **2005**, *127*, 13486–13487.
- (22) Brutscher, B.; Morelle, N.; Cordier, F.; Marion, D. *J. Magn. Reson. B* **1995**, *109*, 238–242.
- (23) Fiorito, F.; Hiller, S.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2006**, *35*, 27–37.
- (24) Hiller, S.; Wasmer, C.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **2007**, *129*, 10823–10828.

- (25) Seavey, B. R.; Farr, E. A.; Westler, W. M.; Markley, J. L. *J. Biomol. NMR* **1991**, *1*, 217–236.

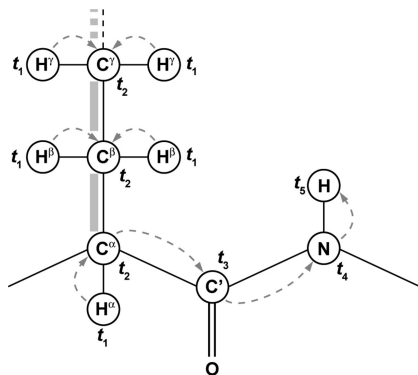


Figure 1. Magnetization transfer pathway in the 5D APSY-HC(CC-TOCSY)CONH experiment. Dashed gray arrows indicate INEPT magnetization transfer steps.³⁹ Thick gray lines represent isotropic mixing. The pathway starts simultaneously on all protons which are bound to carbon atoms. After labeling of the frequencies of the protons, the magnetization is transferred by INEPT to the attached carbon. After evolution at the carbon frequency the magnetization is distributed along the aliphatic carbon spin system during an isotropic mixing period. The fraction of magnetization ending up at the C^α position is transferred by three subsequent INEPT transfers to ¹³C, ¹⁵N, and ¹H^N, with evolution periods on ¹³C and ¹⁵N and signal acquisition on ¹H^N. The four evolution periods are indicated with t_1 – t_4 , and the signal acquisition is indicated by t_5 .

described.^{26,27} The protein concentration was 1.0 ± 0.05 mM as determined by PULCON.²⁸

NMR Measurements. The 5D APSY-HC(CC-TOCSY)CONH experiment (Figure 1) was recorded on a Bruker DRX-750 spectrometer equipped with a room temperature triple-resonance probe with a z -gradient coil. The measurements were done at 40 °C where published reference assignments are available.²⁶ Using the pulse sequence shown in Figure 2, a set of 36 2D projections (Table 1) was recorded. The projection angles were chosen following general considerations on sensitivity and on optimization of dispersion, which are valid for all APSY experiments.^{20,23} Further, at least one of the two evolution times t_1 (¹H) and t_2 (¹³C) were evolved in each projection. Carrier frequencies and sweep widths were 2.7 ppm and 4000 Hz for ω_1 (¹H), 38 ppm and 11 400 Hz for ω_2 (¹³C), 174 ppm and 1700 Hz for ω_3 (¹³C'), 118 ppm and 1800 Hz for ω_4 (¹⁵N), and 4.7 ppm and 12 020 Hz for ω_5 (¹H), respectively. The interscan delay was 1 s, 8 transients were added, and 512 complex points were recorded in the acquisition dimension. The acquired raw data were processed using PROSA.²⁹ Prior to Fourier transformation the FID was multiplied with a 75°-shifted sine bell and zero-filled to 1024 complex points. In the indirect dimension, 128 complex points were measured and the data were multiplied with a 75°-shifted sine bell and zero-filled to 256 complex points before Fourier transformation. The baseline was corrected using the IFLAT method in the ω_5 (¹H)-dimension.³⁰ The algorithm GAPRO²⁰ was used for the calculation of high-dimensional correlation peak lists with the parameter values $\Delta\nu_{\min} = 5$ Hz, $S_{\min} = 8$, $r_{\min} = 1.0$ pt, $R_{\min} = 5$.

Calculation of TOCSY Transfer Amplitudes. The magnetization transfer amplitudes from aliphatic side chain moieties to C^α for all 20 amino acids were calculated with the software AN-TIOPE.³¹ For the calculation an established protocol⁴ was used that assumes a transverse relaxation rate of all carbon atoms of R_2

= 33 Hz and accounts for the nonideal excitation profile of the mixing sequence by reduced carbon–carbon scalar coupling constants depending on the chemical shift difference of the nuclei with a maximal value of ${}^1J = 35$ Hz.³

Results and Discussion

The 5D APSY-HC(CC-TOCSY)CONH experiment exploits a magnetization transfer pathway (Figure 1) commonly used for ¹³C,¹⁵N-labeled proteins.^{5–7} The pathway starts simultaneously on all aliphatic side chain protons including H^α and uses an INEPT element for magnetization transfer to the covalently bound carbon. Subsequently, the magnetization is transferred among the aliphatic carbon nuclei by isotropic mixing, and at the end of the mixing time the part on the C^α nucleus is transferred via the carbonyl carbon to the amide nitrogen of the successive amino acid residue and finally to the attached amide proton, from which the signal is acquired. In the APSY-version of this experiment, four evolution periods are placed along the pathway, resulting in four indirect dimensions (Figure 2). For a given amino acid, the resulting 5D APSY correlation peak list thus contains a group of correlations with C–H moieties which have identical chemical shifts in the three backbone dimensions ω_3 (¹³C'), ω_4 (¹⁵N), and ω_5 (¹H).

The peak intensities of the correlation peaks in CC-TOCSY experiments depend strongly on the amplitude of magnetization transfer during the isotropic mixing period and hence on the length of this period.³² The mixing-time dependence of the transfer amplitudes for all 20 common amino acids were calculated as described previously,⁴ and the result corresponds well to the experimentally observed magnetization transfer in the 5D APSY-HC(CC-TOCSY)CONH experiment (Figure 3 and Figures S1, S2 in the Supporting Information). The calculated transfer amplitudes can thus serve as a guideline to select suitable mixing times. It is evident that there is no single mixing time for which all C–H moieties have sufficiently large transfer amplitudes. This problem can be circumvented in the APSY-version of a TOCSY experiment, since the analysis of the set of projection spectra with GAPRO does not require that a given 5D peak is present in all projections. Thus, the TOCSY mixing time can be varied along with the projection angles. By using a set of mixing times that enables sufficiently high transfer for all aliphatic side chain carbon moieties in some of the projections, resonance frequencies of all C–H moieties from all 20 amino acids can be expected in the resulting APSY correlation peak list. However, peaks can only be part of the final result if the number of projections that support the transfer of a given C–H moiety is above the minimal support threshold S_{\min} .

Based on the calculations of the transfer amplitudes to be expected in CC-TOCSY experiments we have selected three different values for the mixing times: 12, 18, and 28 ms (Table 1). A mixing time of 18 ms in the CC-TOCSY element results in the transfer of magnetization from a majority of carbons in the side chains to the α-carbon nuclei and is also commonly used in classical experiments. The mixing time of 12 ms is favorable for signals which have a small transfer at 18 ms. An example is the H^βC^β peak of Glu, which does not appear with a mixing time of 18 ms. The long mixing time of 28 ms favors signals of long side chains, such as the Lys H^εC^ε, but also signals of short side chains, which are very weak or not present at the two other mixing times, such as the Ser H^αC^α (Figures S1 and S2).

(26) Etezady-Esfarjani, T.; Peti, W.; Wüthrich, K. *J. Biomol. NMR* **2003**, *25*, 167–168.

(27) Etezady-Esfarjani, T.; Herrmann, T.; Peti, W.; Klock, H. E.; Lesley, S. A.; Wüthrich, K. *J. Biomol. NMR* **2004**, *29*, 403–406.

(28) Wider, G.; Dreier, L. *J. Am. Chem. Soc.* **2006**, *128*, 2571–2576.

(29) Güntert, P.; Dötsch, V.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **1992**, *2*, 619–629.

(30) Bartels, C.; Güntert, P.; Wüthrich, K. *J. Magn. Reson. A* **1995**, *117*, 330–333.

(31) Debouregas, F. S.; Waugh, J. S. *J. Magn. Reson.* **1992**, *96*, 280–289.

(32) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521–528.

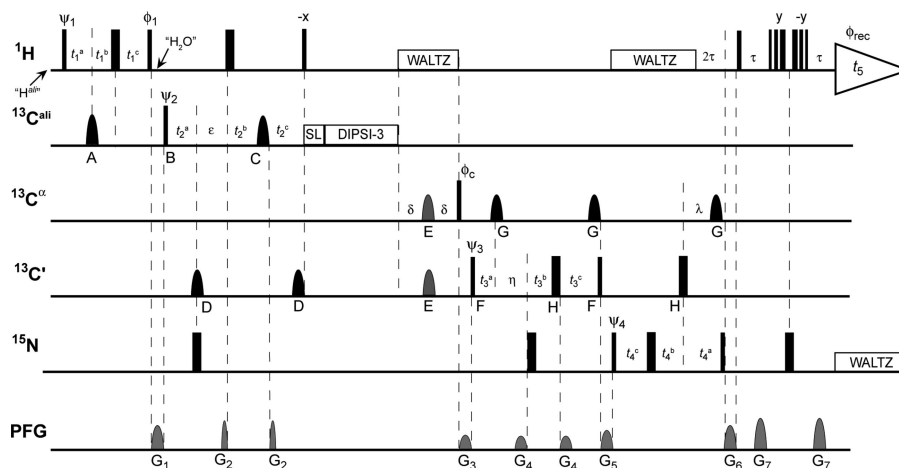


Figure 2. Pulse sequence of the 5D APSY-HC(CC-TOCSY)CONH experiment. Radio-frequency pulses were applied at 38 ppm for aliphatic carbon nuclei, $^{13}\text{C}^{\text{all}}$, at 55 ppm for α -carbons, $^{13}\text{C}^{\alpha}$, at 174 ppm for carbonyls, $^{13}\text{C}^{\beta}$ and 118 ppm for ^{15}N . The carrier frequency for protons was set in the aliphatic region at 2.5 ppm at the beginning of the experiment, indicated on the line ^1H by “ H^{all} ”; at the position “ H_2O ” the carrier was set to the water frequency (4.7 ppm). The thin bars represent 90° pulses. The wide bars and the bell shapes show 180° pulses. Pulses marked with capital letters have special lengths and shapes, depending on their purpose; unmarked bars stand for rectangular pulses applied at maximum power. For the pulses A–H the following types, rotation angles, and durations were used on a 750 MHz spectrometer: $^{13}\text{C}^{\text{all}}$ -pulses: A: Gauss Cascade G3, 180° , 180 μs ; B: Block, 90° , 37.7 μs ; C: Gauss Cascade Q3, 180° , 180 μs ; $^{13}\text{C}^{\alpha}$ -pulses: E: dualband $^{13}\text{C}^{\alpha}/^{13}\text{C}^{\beta}$ pulse⁴⁰ consisting of an reBURP on $^{13}\text{C}^{\alpha}$, 180° , 220 μs and a Gauss (5% truncation) on $^{13}\text{C}^{\beta}$, 180° , 1500 μs ; G: iBurp, 180° , 220 μs ; $^{13}\text{C}^{\beta}$ -pulses: D: Gauss (5% truncation), 180° , 80 μs ; F: Block, 90° , 43.1 μs ; H: Block, 180° , 220 μs . The six last pulses on the line ^1H represent a 3–9–19 Watergate element.⁴¹ The TOCSY mixing on $^{13}\text{C}^{\text{all}}$ (DIPSI-3)⁴² and the decoupling on ^1H and ^{15}N (WALTZ-16)⁴³ are indicated by white rectangles. Before the TOCSY mixing, a 1000 μs spin-lock pulse, SL, is applied. The triangle with t_5 represents the acquisition period. On the line marked PFG, curved shapes indicate sine bell-shaped, pulsed magnetic field gradients along the z -axis, with the following durations and strengths: G_1 : 800 μs , 55%; G_2 : 400 μs , 60%; G_3 : 800 μs , 30%; G_4 : 800 μs , 28%; G_5 : 800 μs , 40%; G_6 : 800 μs , 50%; G_7 : 800 μs , 65%. The initial delays in the evolution periods were $t_1^a = t_1^c = 1.7$ ms, $t_2^c = 1.1$ ms, $t_3^a = 4.5$ ms, $t_3^c = 11.4$ ms, $t_4^a = t_4^c = 12$ ms and $t_1^b = t_2^a = t_2^b = t_3^b = t_4^b = 10$ μs . Further delays were $\epsilon = t_2^c = 1.1$ ms, $\delta = 3.7$ ms, $\eta = t_3^c - t_3^a = 6.9$ ms, $\lambda = 12$ ms, and $\tau = 2.7$ ms. All pulses were applied with phase x unless indicated otherwise above the pulse bar. ϕ_c was set to -10° to compensate for the off-resonance effects of the preceding pulse E. This phase correction was determined by maximizing the signal intensity of 1D traces of the experiment. The following phase cycling was used: $\phi_1 = \{y, y, -y, -y\}$, $\phi_2 = \{x, -x\}$ and $\phi_{\text{rec}} = \{x, -x, -x, x\}$. 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.^{18,22} The pulse phases ϕ_1, ϕ_2, ϕ_3 , and ϕ_4 for t_1, t_2, t_3 , and t_4 , respectively, were used for this purpose, where ϕ_1 – ϕ_3 were incremented and ϕ_4 was decremented in 90° steps for consecutive FIDs; only the pulse phases of the evolution periods which are part of the given projection are incremented.

Table 1. Experimental Parameters of the 2D projections in 5D APSY-HC(CC-TOCSY)CONH

| α | β | γ | spectra ^a | τ_m^b [ms] | exp. time ^c [min] |
|------------------|-----------------|------------------|----------------------|--------------------|---------------------------------|
| 0° | 0° | 0° | 1 | 17.75 | 40 |
| 0° | 0° | 90° | 1 | 17.75 | 40 |
| 0° | 90° | 0° | 1 | 17.75 | 40 |
| 90° | 0° | 0° | 1 | 17.75 | 40 |
| 0° | 90° | $\pm 70.7^\circ$ | 2 | 17.75 | 80 |
| 90° | $\pm 8.5^\circ$ | 0° | 2 | 12.45 | 80 |
| 0° | 0° | $\pm 24.2^\circ$ | 2 | 27.52 | 80 |
| 90° | 0° | $\pm 23^\circ$ | 2 | 12.45 | 80 |
| 0° | $\pm 9^\circ$ | 0° | 2 | 17.75 | 80 |
| 0° | 90° | $\pm 45^\circ$ | 2 | 12.45 | 80 |
| 90° | $\pm 8.5^\circ$ | $\pm 22.8^\circ$ | 4 | 27.52 | 160 |
| $\pm 46.6^\circ$ | $\pm 6.2^\circ$ | 0° | 4 | 17.75 | 160 |
| $\pm 46.6^\circ$ | 0° | $\pm 17.2^\circ$ | 4 | 12.45 | 160 |
| 0° | $\pm 9^\circ$ | $\pm 24^\circ$ | 4 | 17.75 | 160 |
| 0° | $\pm 9^\circ$ | $\pm 40^\circ$ | 4 | 27.52 | 160 |

^a Number of projections for the respective set of projection angles α, β, γ . ^b Duration of the CC-TOCSY mixing time. ^c Acquisition time for each set of projection angles α, β, γ , adding up to a total of 24 h.

With these three mixing times, the 5D APSY-HC(CC-TOCSY)CONH experiment was recorded with the globular 116-residue protein TM1290 in 24 h of spectrometer time using 36 projections (Figure 4, Table 1). Based on the reference assignment of this protein, 424 cross peaks are expected in the resulting 5D APSY correlation peak list.²⁶ 368 thereof, or 87%, were actually found in the present experiment, and in addition three correlations were contained in the APSY correlation peak list that could not be accounted for based on the reference

assignment. These three correlations did not match any backbone chemical shifts triplet ($\text{H}^{\text{N}}, \text{N}, \text{C}'$) of the protein and were thus eliminated during the first step of the automated analysis of the data (see below).

For the analysis and evaluation of the results the 424 expected 5D peaks were grouped according to their spin system type. The group “C–H (1)” contains all CH moieties, for which one peak is expected. These are all CH and CH_3 moieties except the methyls of Val and Leu and all CH_2 moieties with degenerate proton chemical shifts. The group “C–H (2)” consists of all side chain carbons in CH_2 groups with nondegenerate protons. This group thus consists of pairs of peaks with identical chemical shifts in four out of five dimensions. The stereo methyl groups of valine and leucine form the third group “ CH_3 stereo pairs (Val, Leu)”.

Classification of the peaks represented in the 5D APSY peak list to these three groups showed a different degree of completeness for individual groups (Table 2). From the group “C–H (1)” 98% of the expected peaks were found. The same ratio was also obtained for the observation of at least one of the expected two peaks of CH_2 moieties in the “C–H (2)” group, whereas both peaks were found only in 51% of the cases. This lack of completeness originates from the chemical shift degeneracy of these peak pairs in four of the five dimensions. In projections, where the chemical shift difference in the $\omega_1(^1\text{H})$ dimension is close to the digital resolution, or when the evolution period $t_1(^1\text{H})$ is not part of the projection, only one of the peaks in the pair obtains support in the GAPRO calculation and the two peaks are thus not always resolved. In the third group “ CH_3

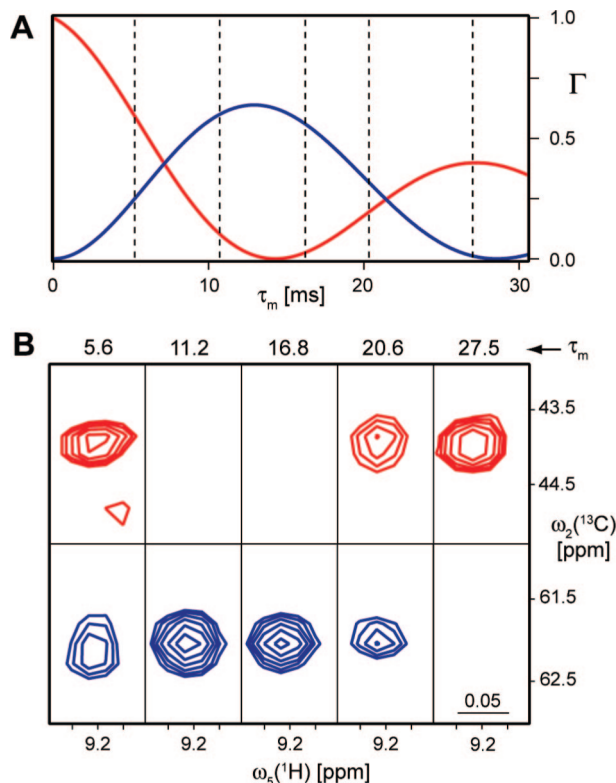


Figure 3. (A) Simulation of the magnetization transfer function $\Gamma(\tau_m)$ during a CC-TOCSY mixing sequence for serine residues; the red line shows the C^α to C^α transfer, and the blue line, the transfer from C^β to C^α . (B) Strips from 5D APSY-HC(CC-TOCSY)CONH experiments recorded with different TOCSY mixing times τ_m . The strips were taken for the residue Ser1 of the protein 434-repressor(1–63); the C^α peak is colored red, and the C^β peak, blue.

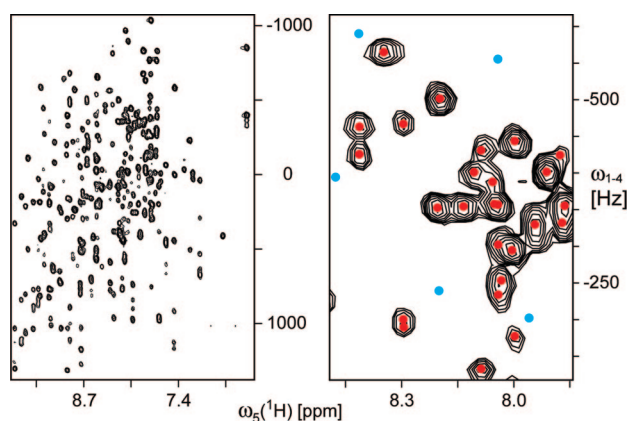


Figure 4. 2D projection ($\alpha = -46.6^\circ$, $\beta = 0^\circ$, $\gamma = -17.2^\circ$) and an expanded region of a 5D APSY-HC(CC-TOCSY)CONH experiment with TM1290 recorded on a 750 MHz spectrometer using a TOCSY mixing time of $\tau_m = 17.75$ ms. In the expansion on the right-hand side, the colored dots represent the projection of the 5D APSY peak: red dots indicate peaks present at the mixing time $\tau_m = 17.75$ ms; blue dots indicate peaks only present in spectra with other mixing times.

stereo pairs (Val, Leu)” a similar situation can occur, when the two methyl moieties of Val or Leu have similar chemical shifts. In this situation only one of the two methyls may be detected; in our application this happened in 4 out of 20 cases. Overall, 368 correlations in the 5D APSY peak list contained the chemical shifts of 97% of the aliphatic carbons and 87% of the aliphatic protons in the protein.

Table 2. Completeness of the 5D APSY-HC(CC-TOCSY)CONH Peak List of TM1290^a

| | expected number of signals | detected number of signals |
|----------------------------|----------------------------|----------------------------|
| C–H (1) | 194 | 191 (98%) |
| C–H (2) | 95 | 93 (98%) |
| | 95 | 48 (51%) |
| CH ₃ (Leu, Val) | 20 | 20 (100%) |
| | 20 | 16 (80%) |
| protons | 424 | 368 (87%) |
| carbon atoms | 329 | 320 (97%) |

^aC–H (1) and C–H (2) are CH_n moieties, for which 1 and 2 resonance signals are expected, respectively (see text).

A key feature of the APSY method is the precision of the resulting peak list. The precision of the chemical shifts obtained for $\omega_5(^1\text{H})$, $\omega_4(^{15}\text{N})$, and $\omega_3(^{13}\text{C}')$ from the 5D APSY-HC(CC-TOCSY)CONH experiment can be assessed since for these values several independent measures are available from the individual “TOCSY peak group” of each residue. For TM1290, 111 amino acid residues were covered by the 368 peaks. For each TOCSY peak group the average values of the chemical shifts $\omega_5(^1\text{H})$, $\omega_4(^{15}\text{N})$, and $\omega_3(^{13}\text{C}')$ were calculated. The overall precision of the experiment is given by the standard deviation of the measurements to their respective average value. The resulting precision was 0.5 Hz for $\omega_5(^1\text{H})$, 2.3 Hz for $\omega_4(^{15}\text{N})$, and 3.6 Hz for $\omega_3(^{13}\text{C}')$. This high precision, which is substantially below the digital resolution of the individual projection spectra, results from the averaging of the independent measurements of the peak positions in the set of projections, and it forms the basis for the subsequent assignment procedure with ALASCA. For $\omega_1(^1\text{H})$ and $\omega_2(^{13}\text{C}')$ the calculation of the standard deviation is not possible, since each of these nuclei is measured at most twice.

The high quality of the GAPRO peak list of the 5D APSY-HC(CC-TOCSY)CONH experiment in terms of dimensionality, completeness, precision, and very low number of artifacts provides an excellent basis for a reliable automated assignment of aliphatic side-chain atoms. Although the TOCSY mixing does not provide information on the direct covalent connectivities among the carbon nuclei, the 5D peaks can be used for sequence-specific resonance assignment of aliphatic resonances by matching their chemical shifts to statistical average values for the 20 amino acids. The assignment can be obtained by a linear procedure that consists of five steps. The input of the algorithm consists of the peak list of the 5D APSY-HC(CC-TOCSY)CONH experiment and the known backbone assignment including the amino acid sequence.

The algorithm ALASCA was applied with the experimental APSY peak list of the 5D APSY-HC(CC-TOCSY)CONH data of TM1290, together with the sequence-specific resonance assignment of the $^{13}\text{C}'$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, ^{15}N , and $^1\text{H}^\text{N}$ nuclei obtained from APSY backbone experiments.²³ With ALASCA all peaks contained in the 5D peak list of TM1290 were correctly assigned (Figure S3).

Two issues that may affect the quality of the ALASCA results will be discussed in the following: (i) the precision of the chemical shift list and (ii) overlap of statistical chemical shift distributions for two CH moieties in a given amino acid.

(i) The precision of the experimental chemical shifts in the input data is critical for the identification of the correct residue for a given peak in step 1. The precision of the 3D ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^\text{N}$)-correlation in the 5D APSY-HC(CC-TOCSY)CONH

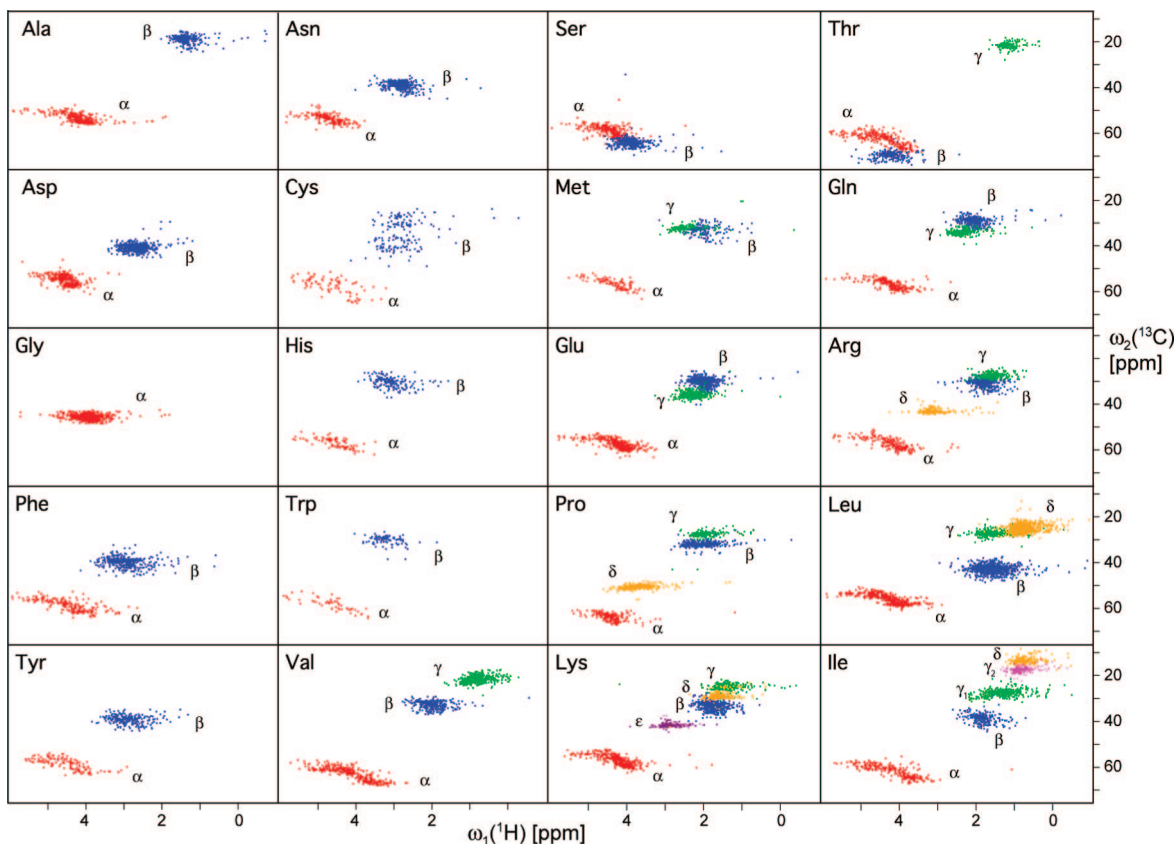


Figure 5. Distribution of chemical shifts of aliphatic protons, ω_1 , and the attached carbon, ω_2 , in a set of 42 representative proteins taken from the BMRB database²⁵ (see Supporting Information Table T1). The data points are colored and labeled with Greek letters according to their position in the amino acid.

experiment must be smaller than the distances between any two ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$) chemical shift triplets of individual residues in the protein to avoid ambiguities. In our application with the protein TM1290 the worst precision for any 3D ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$) subpeak was 3.6 Hz, but the closest distance between two ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$) chemical shift triplets for TM1290 is 37 Hz and 95% of these triplets were separated by more than 80 Hz. The precision of chemical shift measurement in the 5D APSY-HC(CC-TOCSY)CONH experiment is thus approximately an order of magnitude better than the closest distance between ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$) chemical shift triplets for any two residues, allowing a reliable and unambiguous formation of the TOCSY peak groups. Similar values for the distance distributions between 3D ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$)-correlations were found for other proteins with similar or larger sizes. For example, for the 141 residue protein Core Binding Factor b (BMRB 4092),³³ the closest distance between two ($^{13}\text{C}'$, ^{15}N , $^1\text{H}^{\text{N}}$) chemical shift triplets is 45 Hz and 92% of the closest pair distances are above 80 Hz. These considerations indicate that only in exceptional cases an individual error of this kind is to be expected.

(ii) In Step 5 of ALASCA nonoverlapping statistical chemical shift distributions are required for two different C–H moieties in a given amino acid for an unambiguous assignment (Figure 5). Otherwise, the assignments of two C–H groups can be interchanged, but only if the chemical shifts of both C–H moieties are closer to the average value of the respective other moiety than to their own average value. The probability of such an interchanged

assignment was analyzed using the chemical shift data of 42 proteins from the BMRB database²⁵ (see Supporting Information Table T1). Assuming that the α - and β -carbon chemical shifts are known from the backbone assignment, we found that an interchange can occur only in three of the 20 proteinogenic amino acids. In isoleucine, leucine, and lysine, the chemical shift distributions of, respectively, $\text{H}^{\gamma}\text{C}^{\gamma 2}$ and $\text{H}^{\delta}\text{C}^{\delta}$, $\text{H}^{\gamma}\text{C}^{\gamma}$ and $\text{H}^{\delta}\text{C}^{\delta}$, and $\text{H}^{\gamma}\text{C}^{\gamma}$ and $\text{H}^{\delta}\text{C}^{\delta}$ moieties overlap partially (Figure 5). Since these chemical shift distributions overlap however only slightly, an interchanged assignment of these groups occurs only with probabilities of 2%, 4%, and 4% for Ile, Leu, and Lys, respectively, as determined from the set of 42 representative proteins.

In practical applications, the assignments for C^{β} , C^{α} , or both may be missing in the backbone assignment of some of the residues in a protein. It is thus of interest how ALASCA performs for residues with missing C^{α} and C^{β} assignment; i.e., step 4 of the ALASCA procedure can not be applied. We have addressed this question statistically and experimentally. First, we did a statistical analysis of the set of 42 proteins that we used above and investigated misassignments when C^{α} and C^{β} are not contained in the backbone assignment. Due to the distances between the chemical shift distributions for the different CH moieties (Figure 5), the aliphatic resonances of the amino acids Ala, Asn, Asp, Cys, Gly, His, Phe, Trp, Tyr, and Val are always correctly assigned, independent of whether the C^{α} and C^{β} assignment is available or not. For Ser and Thr, the $\text{H}^{\alpha}\text{C}^{\alpha}$ and $\text{H}^{\beta}\text{C}^{\beta}$ assignment can be interchanged, with a 2% (Ser) and 3% (Thr) probability. For Arg, Gln, Glu, Met, and Pro, $\text{H}^{\beta}\text{C}^{\beta}$ and $\text{H}^{\gamma}\text{C}^{\gamma}$ can be interchanged with probabilities of 0.5%, 0.5%, 2%, 3%, and <0.5%, respectively. For lysine

(33) Huang, X.; Speck, N. A.; Bushweller, J. H. *J. Biomol. NMR* **1998**, *12*, 459–460.

residues, in addition to the interchange of $H^{\gamma}C^{\gamma}$ and $H^{\delta}C^{\delta}$ (discussed above), upon missing C^{β} information the assignments of $H^{\beta}C^{\beta}$ and $H^{\gamma}C^{\gamma}$ are interchanged with 1.5% probability. The presence or absence of the C^{α} assignments of lysines does not have an impact on the assignment result. In a second approach, we tested the effect of missing C^{α} and C^{β} assignments with the experimental data for TM1290. In this case, the correct assignment for all residues was obtained when all C^{α} and C^{β} assignments were deleted; i.e., the backbone assignment consisted only of the $^1H^N$, ^{15}N , and $^{13}C'$ assignments.

For an evaluation of the usefulness of the APSY side chain assignment, we used the aliphatic side chain assignments achieved here together with manually obtained assignments of the aromatic residues for a structure calculation of the protein TM1290 (Figure S4). The 87% complete aliphatic assignments achieved in the present work result in structures of TM1290 of virtually identical quality as the manually established assignments.²⁷ This result is in line with theoretical analyses showing that assignments with completions of as small as 40% can enable the correct fold of a protein.^{1,34,35} With the help of such an initial structure, an improvement of the APSY side chain assignment can even be obtained without recording additional experiments based on the initial structure and the NOESY spectra used for the structure determination. Several interactive and automated approaches for this purpose are available.^{13,15,36,37} Such an analysis of the NOESY spectra will be strongly facilitated by the almost complete assignment of the aliphatic side chain carbon atoms obtained with ALASCA (97% for TM1290). Alternatively, the overall completeness of the assignments could be increased or the assignment errors could be reduced by the use of a COSY-type side chain experiment. For an initial study, the 4D APSY-HCCH-COSY experiment was recorded as described in the Supporting Information (Figure S5). A preliminary analysis showed that the resulting 4D correlation peak list is in principle well suitable to address these issues. However, since the results obtained with a single 5D APSY-HC(CC-TOCSY)CONH experiment alone were already of very high quality, also on the level of a structure calculation, we did not further investigate these possibilities for improvement of the assignments.

In summary, the 5D APSY-HC(CC-TOCSY)CONH experiment applied to the 116-residue protein TM1290 provided a peak list that contained 97%/87% of the expected aliphatic carbon/proton resonances. With the independently established

sequence-specific backbone assignment, containing only the $^1H^N$, ^{15}N , and $^{13}C'$ assignments, unambiguous and correct assignments of the aliphatic C–H moieties in the side chain spin systems were obtained. The assignment procedure relies on simple rules, implemented in the algorithm ALASCA. Statistical and practical considerations showed that ALASCA assigns the detected peaks with an average error rate of below 1% for typical globular proteins, also in the case of partially incomplete backbone assignments. The 5D APSY-HC(CC-TOCSY)CONH can be applied with any globular protein of the size of TM1290 and possibly larger, and the analysis is fully automated. Thus, we expect that a fully automated aliphatic side chain assignment based on TOCSY-type transfer is feasible for globular proteins up to at least 120 amino acids. Combined with automated approaches for the backbone assignment,²³ the presently suggested strategy provides automated assignments for most parts of a protein. Aromatic residues cannot be assigned using the 5D APSY-HC(CC-TOCSY)CONH experiment with standard technology, and other magnetization pathways have to be exploited for their assignment using standard NMR equipment. However, using microcoil probes³⁸ simultaneous TOCSY magnetization transfer between aromatic and aliphatic carbons becomes possible, and complete side chain assignments by 5D APSY-HC(CC-TOCSY)CONH seem feasible. The automated backbone and side chain assignment using APSY-type experiments together with automated algorithms for the NOESY spectrum analysis and structure calculation can be set up for automated structure determinations of small- and medium-sized proteins.

Acknowledgment. Financial Support for this project was provided by grants to Kurt Wüthrich from the Schweizerischer Nationalfonds and the ETH Zürich through the NCCR “Structural Biology”. We thank Prof. Wüthrich for his interest in this work and for his continuous support. Dr. Touraj Etezady-Esfarjani and Olivier Duss are acknowledged for providing the TM1290 sample.

Supporting Information Available: Figures S1–S5 and Table T1. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA803161D

- (34) Nabuurs, S. B.; Spronk, C. A. E. M.; Krieger, E.; Maassen, H.; Vriend, G.; Vuister, G. W. *J. Am. Chem. Soc.* **2003**, *125*, 12026–12034.
(35) Kuszewski, J.; Schwieters, C. D.; Garrett, D. S.; Byrd, R. A.; Tjandra, N.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 6258–6273.
(36) Grishaev, A.; Llinas, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6707–6712.
(37) Ab, E.; Pugh, D. J.; Kaptein, R.; Boelens, R.; Bonvin, A. M. *J. Am. Chem. Soc.* **2006**, *128*, 7566–7571.

- (38) Peti, W.; Norcross, J.; Eldridge, G.; O'Neil-Johnson, M. *J. Am. Chem. Soc.* **2004**, *126*, 5873–5878.
(39) Morris, G. A.; Freeman, R. *J. Am. Chem. Soc.* **1979**, *101*, 760–762.
(40) Hilty, C.; Fernández, C.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2002**, *23*, 289–301.
(41) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson. A* **1993**, *102*, 241–245.
(42) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274–293.
(43) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *J. Magn. Reson.* **1983**, *52*, 335–338.