Article

Sequence-Specific Resonance Assignment of Soluble Nonglobular Proteins by 7D APSY-NMR Spectroscopy

Sebastian Hiller, Christian Wasmer, Gerhard Wider, and Kurt Wthrich

J. Am. Chem. Soc., **2007**, 129 (35), 10823-10828• DOI: 10.1021/ja072564+ • Publication Date (Web): 11 August 2007

J|A|C|S
ARTICLES
Published on Web 08/11/2007

Sequence-Specific Resonance Assignment of Soluble Nonglobular Proteins by 7D APSY-NMR Spectroscopy

Sebastian Hiller,[†] Christian Wasmer,[†,§] Gerhard Wider,*[,†] and Kurt Wüthrich[†,‡]

Contribution from the Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland, and Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037

Received April 12, 2007; E-mail: gsw@mol.biol.ethz.ch

Abstract: Based on sequence-specific resonance assignments, NMR is the method of choice for obtaining atomic-resolution experimental data on soluble nonglobular proteins. So far, however, NMR assignment of unfolded polypeptides in solution has been a time-consuming task, mainly due to the small chemical shift dispersion, which has limited practical applications of the NMR approach. This paper presents an efficient, fully automated method for sequence-specific backbone and β-carbon NMR assignment of soluble nonglobular proteins with sizes up to at least 150 residues. The procedure is based on new APSY (automated projection spectroscopy) experiments which benefit from the short effective rotational correlation times in soluble nonglobular polypeptides to record five- to seven-dimensional NMR data sets, which reliably resolves chemical shift degeneracies. Fully automated sequence-specific resonance assignments of the backbone nuclei and C^{β} are described for the uniformly ^{13}C,^{15}N-labeled urea-denatured 148-residue outer membrane protein X (OmpX) from *E. coli*. The method is generally applicable to systems with similar spectroscopic properties as unfolded OmpX, and we anticipate that this paper may open the door for extensive atomic-resolution studies of chemical denaturant-unfolded proteins, as well as some classes of functional nonglobular polypeptides in solution.

Introduction

Globular proteins are represented with about 30 000 deposits in the protein databank.[1] In stark contrast, only a handful of soluble nonglobular polypeptides have been characterized at atomic resolution, in spite of their keen interest for investigations of protein folding as well as for insight into the structural basis of functional nonglobular polypeptides.[2−9] The scarcity of data on soluble nonglobular proteins arose because they are not amenable to meaningful single-crystal studies, and solution NMR studies have been limited by the small dispersion of the chemical shifts.[10−12] In a few instances, spots of local hydrophobic collapse of "random-coil" polypeptide chains have been experimentally identified in solutions of denatured proteins,[13−16] and it has been hypothesized that such clusters might constitute nucleation sites for folding pathways involving either natively or non-natively folded intermediates.[3,13,14] Further increased interest in detailed structural characterization of soluble nonglobular polypeptides has more recently been generated by the discovery of a rapidly increasing number of proteins that are intrinsically unfolded in their functional state in solution.[4,5,8]

Polypeptides cannot be crystallized in flexibly "unstructured" states, so that NMR in solution is the only source of atomic-resolution structural data. However, due to the small chemical shift dispersion, the NMR assignment of soluble nonglobular polypeptides has so far been a laborious process, which has been pursued to completion only for a small number of systems.[10−16] This paper describes a new, fully automated approach to solving this problem with the use of very high-dimensional NMR. High-dimensional NMR spectra have reduced resonance overlap (Figure 1), and the direct access to correlations of a large number of nuclei (Figure 2) eliminates one or multiple steps of chemical shift matching that would be needed for obtaining corresponding information from two or

[†] ETH Zurich.
[‡] The Scripps Research Institute.
[§] Present address: Laboratory for Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland.

(1) Berman, H.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980−980.
(2) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(3) Shortle, D. *Curr. Opin. Struct. Biol.* **1993**, *3*, 66−74.
(4) Plaxco, K. W.; Gross, M. *Nature* **1997**, *386*, 657−659.
(5) Dunker, A. K., et al. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.
(6) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18−25.
(7) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863−867.
(8) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 197−208.
(9) Lindorff-Larsen, K.; Rogen, P.; Paci, E.; Vendruscolo, M.; Dobson, C. M. *Trends Biochem. Sci.* **2005**, *30*, 13−19.
(10) Wüthrich, K. *Curr. Opin. Struct. Biol.* **1994**, *4*, 93−99.
(11) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977−8991.
(12) Dyson, H. J.; Wright, P. E. *Methods Enzymol.* **2001**, *339*, 258−270.
(13) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559−1563.
(14) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722.
(15) Crowhurst, K. A.; Forman-Kay, J. D. *Biochemistry* **2003**, *42*, 8687−8695.
(16) Tafer, H.; Hiller, S.; Hilty, C.; Fernández, C.; Wüthrich, K. *Biochemistry* **2004**, *43*, 860−869.

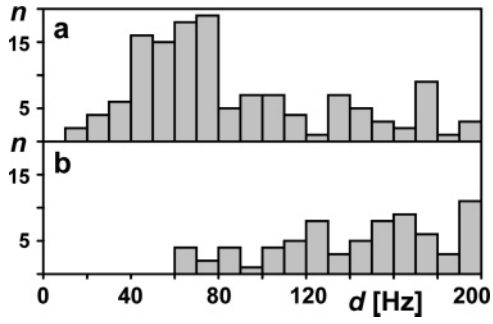

Figure 1. Spectral resolution for the urea-denatured 148-residue protein OmpX in a three-dimensional (3D) and a 7D NMR experiment. The histograms show the number, *n*, of frequency differences smaller than 200 Hz, *d*, between pairs of resonances in the 3D and 7D frequency space, respectively, as calculated from the chemical shifts at a ^{1}H frequency of 750 MHz.[16] (a) 3D seq-(H)N(COCA)NH.[35] (b) 7D APSY-seq-HNCO(CA)-CBCANH.

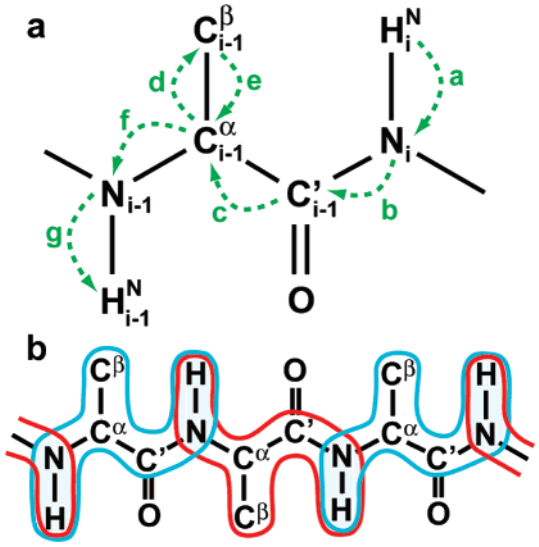

Figure 2. (a) Magnetization transfer pathway of the 7D APSY-seq-HNCO-(CA)CBCANH NMR experiment with polypeptide chains. The dashed green arrows indicate the scalar couplings used for the seven magnetization transfer steps a to g (back-transfer from C^{α}_{i-1} to N_i is suppressed by the experimental scheme used).[31] (b) Scheme for connecting sequentially adjacent seven-atom fragments (alternating cyan and red shapes), by matching of the chemical shifts of the overlapping amide moieties (gray areas). Each of these seven-atom fragments gives rise to one peak in the 7D spectrum.

several lower-dimensional data sets, which dramatically reduces ambiguities in the assignment process.

High-dimensional NMR experiments have become practical with recently introduced new sampling and processing techniques.[17−22] Here, we use automated projection spectroscopy (APSY), where two-dimensional projections of the higher-dimensional experiment are recorded, from which a precise peak list in the high-dimensional frequency space is calculated by a dedicated algorithm, GAPRO.[22] It is then demonstrated that the use of this peak list as input for an automatic algorithm for obtaining sequence-specific polypeptide backbone assignments, such as GARANT,[23] enables fully automated resonance assignments, without further human intervention after the start of the APSY-NMR experiment. It is a special advantage that the long transverse relaxation times in soluble nonglobular polypeptides enable very high-dimensional APSY-NMR experiments, as is illustrated below with the NMR assignment of the urea-denatured 148-residue outer membrane protein X (OmpX).[16] We have chosen this protein as a representative system to develop and demonstrate the method, since the assignments have been determined independently using conventional methods (BMRB entry 15201, ref 16). The method should be transferable to solutions of other unfolded proteins of comparable or even somewhat larger size, since similar nuclear spin relaxation properties can be expected for all chemical denaturant-unfolded proteins in the size range of OmpX.

Results

There are two common key elements in all presently used strategies for sequence-specific backbone resonance assignment of polypeptides.[24] First, sequential NMR connectivities lead to spectroscopic identification of discrete peptide fragments of different lengths. Second, these fragments are mapped onto the sequence, based on matching patterns of the sequence distribution of amino acid types. Thereby the C^{β} chemical shifts have a key role, since they allow the unambiguous identification of some of the 20 proteinogenic amino acid types in the NMR-connected polypeptide fragments, which results in a high local stability of the assignment result.[25,26]

Using APSY-NMR, sequential connectivities of neighboring backbone amide moieties and the C^{α}, C^{β}, and C′ chemical shifts between them can be obtained with a single, presently introduced experiment, 7D APSY-seq-HNCO(CA)CBCANH (Figures 2 and 3). In the pulse sequence of the 7D APSY-seq-HNCO(CA)-CBCANH experiment (Figure 3), steady-state magnetization of the amide proton *i* (Figure 2a) is transferred with subsequent INEPT[27] steps via the scalar couplings $^1J_{N,HN}$, $^1J_{N,C'}$, $^1J_{C',C\alpha}$, $^1J_{C\alpha,C\beta}$, $^1J_{C\alpha,N}$, and $^1J_{N,HN}$, with acquisition on the amide proton *i*-1. Along this pathway, six evolution periods for the indirect dimensions, t_1-t_6, are placed on the nuclei $^1H^N_i$, $^{15}N_i$, $^{13}C'_{i-1}$, $^{13}C^{\beta}_{i-1}$, $^{13}C^{\alpha}_{i-1}$, and $^{15}N_{i-1}$, respectively. In this experiment, each seven-atom fragment of residues *i* and *i*-1 gives rise to one peak, except if a proline residue or a chain end is located at either of the positions *i* or *i*-1. Correlations involving a non-glycine residue at position *i*-1 contain the C^{β} chemical shift in ω_4 and the C^{α} chemical shift in ω_5. Correlations involving a glycine residue at position *i*-1 contain the Gly C^{α} chemical shift in both ω_4 and ω_5. In future implementations, the pulse sequence of Figure 3 could readily be supplemented with water flip-back pulses[28] and sensitivity enhancement schemes.[29,30]

For the application with urea-unfolded OmpX, this experiment was recorded with 100 2D projections in a total measuring time

(17) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393.
(18) Orekhov, V. Y.; Ibraghimov, I.; Billeter, M. *J. Biomol. NMR* **2003**, *27*, 165−173.
(19) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126*, 6429−6440.
(20) Atreya, H. S.; Eletsky, A.; Szyperski, T. *J. Am. Chem. Soc.* **2005**, *127*, 4554−4555.
(21) Eghbalnia, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12528−12536.
(22) Hiller, S.; Fiorito, F.; Wüthrich, K.; Wider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10876−10881.
(23) Bartels, C.; Güntert, P.; Billeter, M.; Wüthrich, K. *J. Comput. Chem.* **1997**, *18*, 139−149.
(24) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(25) Grzesiek, S.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 185−204.
(26) Güntert, P.; Salzmann, M.; Braun, D.; Wüthrich, K. *J. Biomol. NMR* **2000**, *18*, 129−137.
(27) Morris, G. A.; Freeman, R. *J. Am. Chem. Soc.* **1979**, *101*, 760−762.
(28) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 12593−12594.
(29) Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151−170.
(30) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.

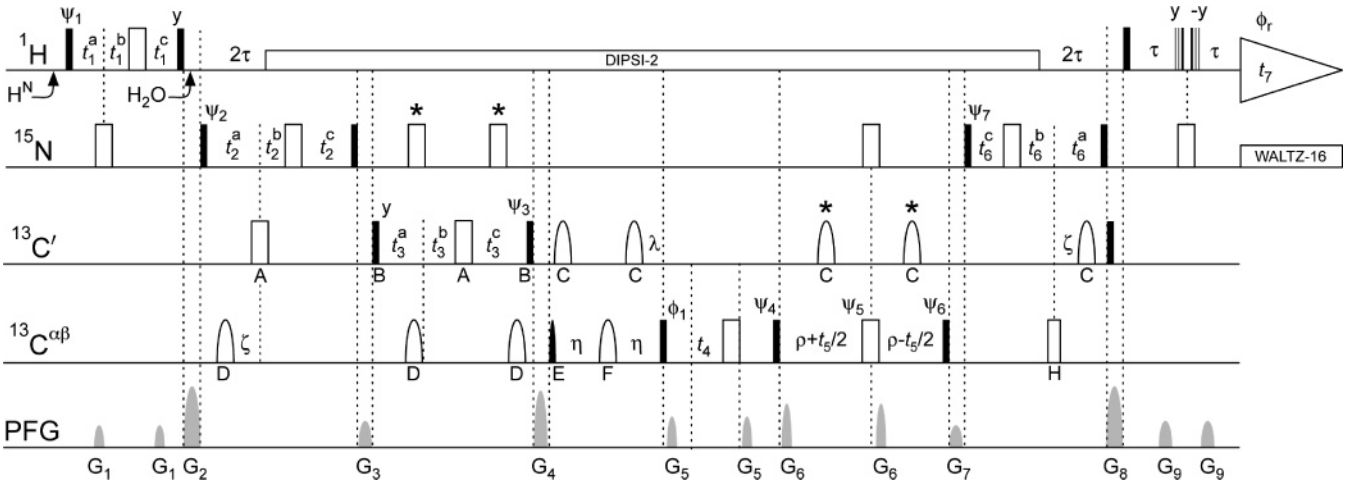

Figure 3. Pulse sequence used for the 7D APSY-HNCO(CA)CBCANH experiment with urea-denatured OmpX. Radio frequency pulses were applied at 118.0 ppm for ^{15}N, 173.0 ppm for ^{13}C′, and, unless otherwise stated, 42.0 ppm for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$. At the start of each transient, the ^{1}H carrier frequency was set at 8.24 ppm, indicated by "H^N", and at the time point "H_2O", the carrier was changed to 4.7 ppm. Black and white symbols represent 90°- and 180°-pulses, respectively. Unlabeled bars stand for rectangular pulses applied at maximum power. Pulses marked with capital letters have individually adjusted lengths and shapes, depending on their purpose. All pulse lengths are given for a ^{1}H frequency of 750 MHz: ^{13}C′-pulses: A, 180°, rectangular shape, 38.3 μs; B, 90°, rectangular shape, 42.8 μs; C, 180°, I-Burp,[36] 220 μs. $^{13}C^{\alpha\beta}$-pulses: D, 180°, I-Burp (applied at 51.0 ppm), 220 μs; E, 90°, Gaussian cascade Q5,[37] 280 μs; F, 180°, Gaussian cascade Q3,[37] 185 μs; H, 180°, rectangular shape, 38.3 μs. The ^{15}N-pulses labeled with an asterisk are centered with respect to $t_3^a + t_3^b$, and to t_3^c, respectively. The ^{13}C′-pulses labeled with an asterisk are centered with respect to $\rho + t_5/2$ and to $\rho - t_5/2$, respectively. The last six pulses on the ^{1}H line represent a 3−9−19 Watergate pulse train.[38] Decoupling using DIPSI-2[39] on ^{1}H and WALTZ-16[40] on ^{15}N is indicated by white rectangles. The triangle with t_7 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, 600 μs, 13 G/cm; G_2, 1000 μs, 37 G/cm; G_3, 800 μs, 16 G/cm; G_4, 800 μs, 34 G/cm; G_5, 600 μs, 19 G/cm; G_6, 600 μs, 27 G/cm; G_7, 800 μs, 13 G/cm; G_8, 1000 μs, 37 G/cm; G_9, 800 μs, 16 G/cm. Pulse phases different from *x* are indicated above the pulses. Phase cycling: $\phi_1 = \psi_2 = \phi_r = \{x, -x\}$, $\psi_4 = \{x, x, -x, -x\}$, $\psi_6 = y$. The initial delays were $t_1^a = t_1^c = 2.7$ ms, $t_2^a = t_2^c = 14.0$ ms, $t_3^a = t_3^c = 4.7$ ms, $t_6^a = t_6^c = 14.0$ ms, and $t_1^b = t_2^b = t_3^b = t_4 = t_5 = t_6^b = 0$ ms. Further delays were $\tau = 2.7$ ms, $\zeta = 14.0$ ms, $\eta = 6.8$ ms, $\lambda = 4.7$ ms, and $\rho = 20.75$ ms. Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem[41,19] with the phases ψ_1, ψ_2, ψ_3, $\psi_4-\psi_6$, ψ_6, and ψ_7 for t_1, t_2, t_3, t_4, t_5, and t_6, respectively, which were simultaneously incremented in 90° steps for consecutive FIDs. Evolution periods were implemented as direct evolution for t_4 and in constant-time fashion for t_1, t_2, t_3, t_5, and t_6. For t_1, t_2, t_3, and t_6, semiconstant time evolution was used for those maximal evolution periods that are too long to be accommodated in constant-time periods.

of 50 h. Based on the chemical structures in the presently used OmpX sample,[16] 142 peaks are predicted, of which 139 peaks were actually observed. The spectral analysis revealed that the three missing peaks connect residues 98−101, and the likely reason for their absence is unfavorable backbone dynamics in this segment.

Sequential assignments from 7D APSY-seq-HNCO(CA)-CBCANH data rely on matching of the combined amide ^{15}N and ^{1}H chemical shifts of pairs of sequentially neighboring amide moieties on the background of 7D signals from all the dipeptide segments in the protein. High precision of the chemical shift measurements is therefore crucial. In the present work, a precision of 0.46 and 0.44 Hz, respectively, for the ^{1}H and ^{15}N chemical shifts was achieved (Figure 4). Figure 4 also illustrates the significance of "peak separation" (Figure 1), which determines the overlap of nearby peak contours, as compared to the precision of the chemical shift measurements. This high precision enabled automated NMR assignment with the program GARANT[23] in spite of residual chemical shift degeneracy in individual ones of the seven dimensions.

In soluble nonglobular polypetides, glycine, alanine, serine, and threonine residues can be unambiguously identified based on their unique C^{α} and C^{β} chemical shifts, whereas the other 16 amino acids have more closely similar C^{α} and C^{β} shifts and in the present context jointly represent a fifth type of NMR-recognizable residue. Polypeptide fragments assembled from sequential NMR connectivities (Figure 2) that contain unique patterns of these five amino acid types can unambiguously be assigned to the corresponding segments of the amino acid sequence. Statistical analysis of the protein sequence database shows that sequence-specific NMR assignment based on this principle is a highly overdetermined problem (ref 23; see also Figure 5). The high precision of chemical shift measurement achieved with 7D APSY-NMR allowed sequential connection of long fragments bounded by the chain ends and the proline residues (black lines in Figure 5). With these fragments, correct sequence-specific backbone resonance assignments were obtained for all detected peaks. To indicate the additional high local stability in large parts of the polypeptide chain and the intrinsic overdetermination of the assignment problem, we identify in Figure 5 the locations of some of the peptide segments of length up to five residues that have unique patterns of the five amino acid types and contain at least two of the unique residues Ala, Gly, Ser, and Thr. An additional, independent check of the assignments is provided by the fact that no sequential connectivities are expected either with prolines or with the N-terminal residue (Figures 2 and 5). In the present analysis, the assignment was based only on five types of amino acids, with all non-Ala/Gly/Ser/Thr residues forming a single type. This strategy could be further refined by forming subgroups of the remaining 16 amino acids.[24] However, such a refinement was not further pursued here, since already the use of five types lead to a highly overdetermined assignment problem with a robust solution. Assignment algorithms with a probabilistic approach, such as GARANT, automatically exploit the full information.

As an alternative to the 7D APSY experiment we make reference to a combination of a 5D with a 6D APSY-NMR

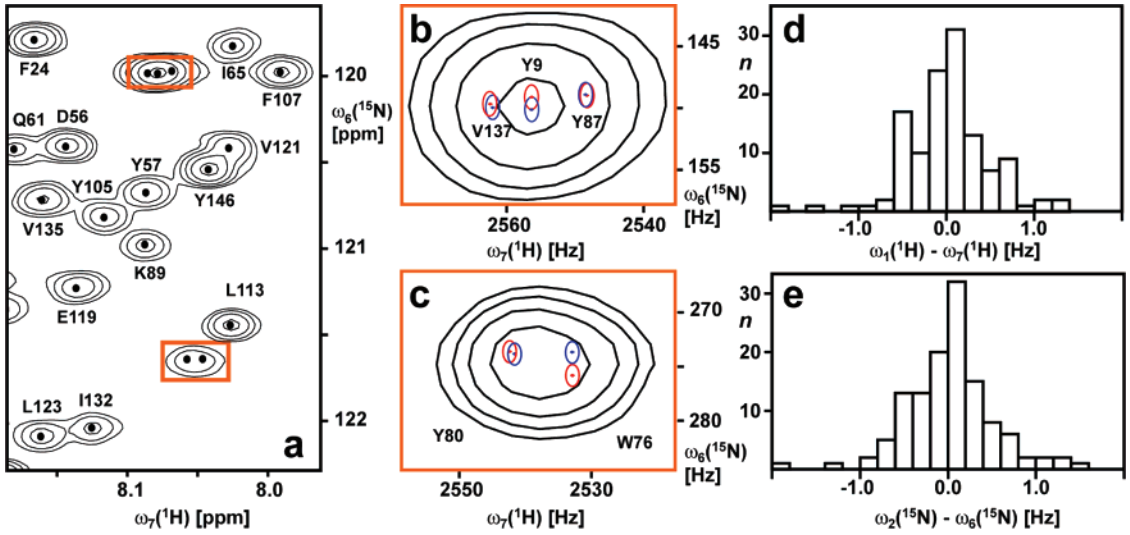

Figure 4. Precision of chemical shift measurements by the 7D APSY-seq-HNCO(CA)CBCANH experiment, illustrated with data of the urea-denatured protein OmpX. (a) Spectral region from the (0°,0°,0°,0°,0°)-projection, which corresponds to a 2D [^{15}N,^{1}H]-correlation spectrum. The black dots are projections of the 7D peak positions, as represented by the [ω_6,ω_7]-correlations onto the experimental 2D projection spectrum. Orange squares indicate two clusters of overlapped signals which are displayed on an expanded scale in (b) and (c). (b) The two different [^{15}N,^{1}H]-pairs contained in each 7D signal (Figure 2) are indicated in red ([ω_2, ω_1]-correlation) and blue ([ω_6,ω_7]-correlation). Contours are drawn at a distance of 1.0 Hz around the peak positions projected from the 7D data set. (c) Same presentation as that in (b). In (a)−(c), resonance assignments are given using one-letter amino acid symbols and the sequence positions. (d) Histogram of the variance between the measurements of the same amide proton chemical shift from the two 7D signals correlating two sequentially neighbouring groups of seven atoms (Figure 2). (e) Same as (d) for amide nitrogen-15 shifts.

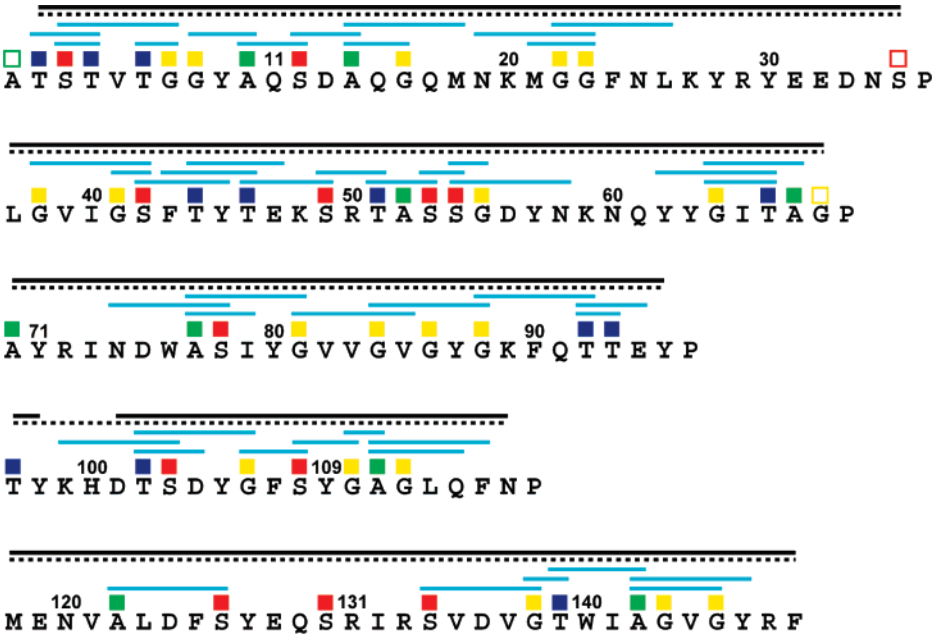

Figure 5. Illustration of automated sequence-specific NMR assignment using the approach introduced in this paper. Above the sequence of the 148-residue protein OmpX, colored squares mark the amino acid types with unique $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ chemical shifts, where the filled squares indicate that these shifts can be measured with the 7D APSY experiment (Figure 2): Ala (green), Gly (yellow), Ser (red), Thr (blue). The cyan lines show a selection of polypeptide segments of length two to five residues that have unique patterns of amino acid types containing at least two residues of Ala, Gly, Ser, or Thr (see text). The solid and dotted black lines above the sequence represent the result of the automated assignment with the software GARANT,[23] using input data from 7D APSY-seq-HNCO(CA)CBCANH or from combined use of 6D APSY-seq-HNCOCANH and 5D APSY-HNCOCACB, respectively.

experiment. The 6D APSY-seq-HNCOCANH scheme[31] is used to connect sequentially neighboring amide groups (similar to the case shown in Figure 2), and the newly developed 5D APSY-HNCOCACB scheme (Figure 6) is used to measure the C^{β} chemical shifts. This out-and-back experiment is based on a classical counterpart, the 3D HN(CO)CACB experiment.[32] Steady-state magnetization of the amide proton *i* (Figure 2a) is transferred with subsequent INEPT[27] steps via the scalar couplings $^1J_{N,HN}$, $^1J_{N,C'}$, $^1J_{C',C\alpha}$, and $^1J_{C\alpha,C\beta}$, with acquisition on the amide proton *i*. Along this pathway, four evolution periods for the indirect dimensions, t_1-t_4, are placed on the nuclei $^{15}N_i$, $^{13}C'_{i-1}$, $^{13}C^{\beta}_{i-1}$, and $^{13}C^{\alpha}_{i-1}$, respectively. In this experiment, each five-atom fragment of the amino acid residues *i* and *i*-1 gives rise to one peak, except if a proline residue is located at

(31) Fiorito, F.; Hiller, S.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2006**, *35*, 27−37.

(32) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655−11666.

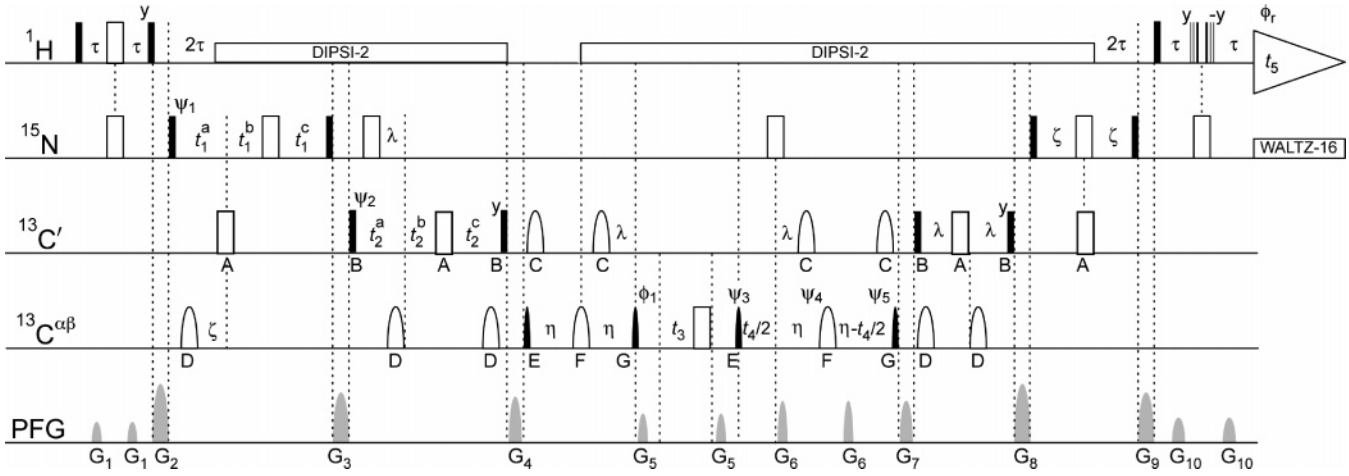

Figure 6. Pulse sequence used for the 5D APSY-HNCOCACB experiment with urea-denatured OmpX. Radio frequency pulses were applied at 118.0 ppm for ^{15}N, 173.0 ppm for ^{13}C′, and, unless otherwise stated, 42.0 ppm for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$. The ^{1}H carrier frequency was set at 4.7 ppm. Narrow and wide black and white symbols represent 90°- and 180°-pulses, respectively. Unlabeled bars are rectangular pulses applied at maximum power. Pulses marked with capital letters have individually adjusted lengths and shapes, depending on their purpose. All pulse lengths are given for a ^{1}H frequency of 750 MHz: ^{13}C′-pulses: A, 180°, rectangular shape, 38.3 μs; B, 90°, rectangular shape, 42.8 μs; C, 180°, I-Burp,[36] 220 μs. $^{13}C^{\alpha\beta}$-pulses: D, 180°, I-Burp (applied at 51.0 ppm), 220 μs; E, 90°, Gaussian cascade Q5,[37] 280 μs; F, 180°, Gaussian cascade Q3,[37] 185 μs; G, 90°, Gaussian cascade Q5 time-reversed, 280 μs. The six last pulses on the ^{1}H line represent a 3−9−19 Watergate pulse train.[38] Decoupling using DIPSI-2[39] on ^{1}H and WALTZ-16[40] on ^{15}N is indicated by white rectangles. 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, 600 μs, 13 G/cm; G_2, 1000 μs, 37 G/cm; G_3, 1000 μs, 32 G/cm; G_4, 800 μs, 29 G/cm; G_5, 600 μs, 19 G/cm; G_6, 600 μs, 27 G/cm; G_7, 800 μs, 27 G/cm; G_8, 1000 μs, 37 G/cm; G_9, 1000 μs, 32 G/cm; G_{10}, 800 μs, 16 G/cm. Pulse phases different from *x* are indicated above the pulses. Phase cycling: $\phi_1 = \phi_r = \psi_5 = \{x, -x\}$. The initial delays were $t_1^a = t_1^c = 14.0$ ms, $t_2^a = t_2^c = 4.7$ ms, and $t_1^b = t_2^b = t_3 = t_4 = 0$ ms. Further delays were $\tau = 2.7$ ms, $\lambda = 4.7$ ms, $\eta = 6.8$ ms, and $\zeta = 14.0$ ms. Quadrature detection for the indirect dimensions was achieved using the trigonometric addition theorem,[41,19] with the phases ψ_1, ψ_2, $\psi_3-\psi_5$, and ψ_5 for t_1, t_2, t_3, and t_4, respectively, which were simultaneously incremented in 90°-steps for consecutive FIDs. Evolution periods were implemented as direct evolution for t_3 and in constant-time fashion for t_1, t_2, and t_4. For t_1 and t_2, semiconstant time evolution was used for those maximal evolution periods that are too long to be accommodated in constant-time periods.

position *i*. Correlations involving a non-glycine residue at position *i*-1 contain the C^{β} chemical shift in ω_3 and the C^{α} chemical shift in ω_4. Correlations involving a glycine residue at position *i*-1 contain the Gly C^{α} chemical shift in both ω_3 and ω_4. During the fourth INEPT step of the pulse sequence, proton decoupling is interrupted for 6.8 ms, in order to achieve equal signs for glycine and non-glycine peaks.

All 142 predicted sequential connectivities were contained in the 6D APSY-seq-HNCOCANH peak list. The 5D APSY-HNCOCACB experiment yielded all expected peaks except for those of residues 29, 30, 33, and 72 (Figure 5), which were not identified as individually resolved peaks due to close proximity to other signals in the 5D spectral space. Identification of corresponding 6D APSY-seq-HNCOCANH and 5D APSY-HNCOCACB correlations relied on matching of the combined chemical shifts of the four nuclei H^N, N, C′, and C^{α}, which had been measured with a higher precision than 1.5 Hz for each spin type. Using the 5D and 6D peak lists as the only input besides the protein sequence, the algorithm GARANT[23] generated automatically the correct sequence-specific backbone resonance assignments for all the detected peaks (Figure 5).

Discussion

This paper makes use of the ability of APSY-NMR to exploit five-, six-, and seven-dimensional experiments, which yield peak lists that can directly be used as input for automated assignment algorithms. In the present experiments, sensitivity was not a limiting factor, and we operated in the sampling-limited regime where the duration of the experiments is determined by the desired resolution and the minimal acceptable phase cycling of the experiment. In future implementations the pulse schemes presented here could readily be supplemented with water flip-back pulses[28] and sensitivity enhancement schemes[29,30] for adaptation to more stringent requirements.

In our measurements, we find that the relative merits of the combination of a 5D APSY-HNCOCACB with a 6D APSY-seq-HNCOCANH when compared to the 7D APSY-seq-HNCO-(CA)CBCANH experiment are very similar, with either alternative having advantages and disadvantages. The average relative sensitivities in the 5D-, 6D-, and 7D-spectra of urea-unfolded OmpX were 2.5, 2.0, and 1.0, respectively. The combined 5D and 6D experiments have a higher sensitivity for signal detection than the 7D experiment, but the spectral resolution for the C^{β} signals is reduced and an additional step of chemical shift matching between the two experiments is required. For larger soluble nonglobular proteins than OmpX, the number of peak pairs that cannot be resolved in the 5D experiment is expected to increase relative to OmpX, where four peak pairs could not be resolved. With sufficient experimental sensitivity, the 7D experiment is clearly the better choice. If the experimental sensitivity is the limiting factor, the choice of the 5D plus 6D APSY-experiments is preferable.

Since the chemical shift values in soluble nonglobular proteins depend only on the amino acid types and their closest neighbors in the sequence,[24,33] it can be expected that similar results to those achieved here with OmpX will be obtained for many other unfolded proteins of sizes up to about 150 residues. Less complete assignments may be obtained when exchange phenomena lead to line broadening, or for systems with repetitive sequences where other approaches would, on principal grounds, also be limited. Work with proteins larger than 150 residues should also be possible, but the increased peak density even in

(33) Braun, D.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1994**, *116*, 8466−8469.

7D data sets may require the recording of a larger number of projections and the use of especially adapted experimental parameters. We expect this paper to open the door for efficient studies of large sets of flexibly disordered, nonglobular proteins in solution. This will enable systematic investigations of hitherto elusive phenomena of fundamental interest, such as the structural basis of nonglobular polypeptide functions,[5,8] the occurrence of residual structured clusters in proteins unfolded by chemical denaturants,[13,15,16] or possible cooperativity of the formation of multiple clusters in a single unfolded polypeptide chain.[14] With the use of the presently introduced resonance assignment techniques, solution NMR will have a further expanded potential for key contributions to the protein folding problem, as well as to new insights into structure−function correlations in the rapidly expanding class of natively nonglobular proteins.

Experimental Section

Calculation of distances in *N*-dimensional frequency space. The distance d between two peaks A and B with coordinates ν_1, ..., ν_N was calculated as

$$d = \sqrt{\sum_{i=1}^{N} (\nu_i^{\mathrm{A}} - \nu_i^{\mathrm{B}})^2}$$

OmpX Sample Preparation. [*U*-^{13}C,^{15}N]-labeled OmpX was expressed as inclusion bodies in *E. coli* and purified using anion-exchange chromatography.[34,16] An NMR sample with a protein concentration of 3 mM in H_2O solution at pH 6.5 containing 8 M urea, 20 mM phosphate, 0.1 mM NaN_3, and 5% D_2O was prepared.

NMR Experiments. All experiments were done at 15 °C on a Bruker DRX 750 spectrometer equipped with a triple-resonance probehead with a *z*-gradient coil. For all experiments the interscan delay was 1.0 s. 768 complex points were recorded in the acquisition dimension, with a sweep width of 7000 Hz centered about the water resonance. In the indirect dimension, 192 complex points were measured. The sweep widths for 5D APSY-HNCOCACB and 7D APSY-seq-HNCO(CA)-CBCANH were 1850 Hz for ^{15}N, 1100 Hz for ^{13}C′, 12000 Hz for $^{13}C^{\beta}$, 800 Hz for $^{1}H^{N}$ and 8000 Hz for $^{13}C^{\alpha}$. The same sweep widths were used for 6D APSY-seq-HNCOCANH, except that the sweep width for $^{13}C^{\alpha}$ was 3600 Hz. For each of the 5D-, 6D- and 7D-APSY experiments, all projections were acquired with a single experimental setup, using a single pulse sequence.

GAPRO. An improved version of the software GAPRO, which will be described in detail elsewhere, was used following a previously established protocol.[22]

5D APSY-HNCOCACB. In the pulse sequence of the 5D-APSY-HNCOCACB experiment (Figure 6), two transients were accumulated for each time increment. Each 2D projection was recorded in 16 min. A set of 60 projections (Table S1) was analyzed with GAPRO, using the parameters $R_{\mathrm{min}} = 4.0$, $\Delta\nu_{\mathrm{min}} = 3.0$ Hz, $r_{\mathrm{min}} = 15.0$ Hz, and $S_{\mathrm{min,1/2}} = 13$.[22] The calculations were run in 20 superloops of 30 iterations each. The calculation time on a standard LINUX PC with a 2.8 GHz Pentium-4 processor was approximately 10 min.

6D APSY-seq-HNCOCANH. This experiment has been described previously for use with folded globular proteins.[31] Four transients were accumulated for each time increment. Each 2D projection was recorded in 30 min. A set of 40 projections (Table S2) was analyzed with GAPRO, using the parameters $R_{\mathrm{min}} = 4.0$, $\Delta\nu_{\mathrm{min}} = 3.0$ Hz, $r_{\mathrm{min}} = 8.0$ Hz, and $S_{\mathrm{min,1/2}} = 25$.[22] The calculations were run in 20 superloops of 30 iterations each. The calculation time on a standard LINUX PC with a 2.8 GHz Pentium-4 processor was approximately 1 h.

7D APSY-seq-HNCO(CA)CBCANH. In the pulse sequence of the 7D-APSY-seq-HNCO(CA)CBCANH experiment (Figure 3), four transients were accumulated per time increment. Each 2D projection was recorded in 30 min. A set of 100 projections (Table S3) was analyzed with GAPRO, using the parameters $R_{\mathrm{min}} = 4.0$, $\Delta\nu_{\mathrm{min}} = 3.0$ Hz, $r_{\mathrm{min}} = 30.0$ Hz, and $S_{\mathrm{min,1/2}} = 22$.[22] The calculations were run in 20 superloops of 50 iterations each. The calculation time on a standard LINUX PC with a 2.8 GHz Pentium-4 processor was approximately 15 h.

Automated Sequence-Specific Resonance Assignment with GARANT. The experimental APSY peak lists were used as input for the algorithm GARANT,[23] which was applied using a standard annealing protocol.[23] Either the 7D peak list or the combination of the 5D and 6D peak lists was used as input for GARANT.

Acknowledgment. This work was supported by the Schweizerischer Nationalfonds and the ETH Zürich through the National Center for Competence in Research (NCCR) Structural Biology.

Supporting Information Available: The supporting material contains the sets of projection angles used for recording the 5D, 6D, and 7D experiments with OmpX in 8 M urea solution. This material is available free of charge via the Internet at http://pubs.acs.org.

JA072564+

(34) Fernández, C.; Adeishvili, K.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2358−2363.
(35) Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V. *J. Biomol. NMR* **2001**, *20*, 135−147.
(36) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93−141.
(37) Emsley, L.; Bodenhausen, G. *Chem. Phys. Lett.* **1990**, *165*, 469−476.
(38) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661−665.
(39) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274−293.
(40) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *J. Magn. Reson.* **1983**, *52*, 335−338.
(41) Brutscher, B.; Morelle, N.; Cordier, F.; Marion, D. *J. Magn. Reson., Ser. B* **1995**, *109*, 238−242.