Journal of Biomolecular NMR, **27:** 57–67, 2003.
KLUWER/ESCOM

Optimized set of two-dimensional experiments for fast sequential assignment, secondary structure determination, and backbone fold validation of $^{13}C/^{15}N$-labelled proteins

Beate Bersch[a], Emmanuel Rossy[b], Jacques Covès[a,b] & Bernhard Brutscher[a,*]
[a]*Institut de Biologie Structurale - Jean-Pierre Ebel, UMR5075 CNRS-CEA-UJF, 41, rue Jules Horowitz, 38027 Grenoble Cedex, France*
[b]*Laboratoire de Chimie et Biochimie des Centres Rédox Biologiques, CEA-Grenoble, UMR 5047 CNRS-CEA-UJF, 17 Avenue des Martyrs, 38054 Grenoble Cedex 9, France*

Received 24 February 2003; Accepted 16 April 2003

Key words: high throughput NMR, reduced dimensionality, resonance assignment, residual dipolar couplings

Abstract

NMR experiments are presented which allow backbone resonance assignment, secondary structure identification, and in favorable cases also molecular fold topology determination from a series of two-dimensional ^{1}H-^{15}N HSQC-like spectra. The ^{1}H-^{15}N correlation peaks are frequency shifted by an amount $\pm\ \omega_X$ along the ^{15}N dimension, where ω_X is the C^{α}, C^{β}, or H^{α} frequency of the same or the preceding residue. Because of the low dimensionality (2D) of the experiments, high-resolution spectra are obtained in a short overall experimental time. The whole series of seven experiments can be performed in typically less than one day. This approach significantly reduces experimental time when compared to the standard 3D-based methods. The here presented methodology is thus especially appealing in the context of high-throughput NMR studies of protein structure, dynamics or molecular interfaces.

Introduction

It is now generally acknowledged that liquid state NMR will play an important role in structural genomics, especially for the rapid screening of secondary and tertiary structures of small to medium sized polypeptides (Montelione et al., 2000). High throughput NMR studies require fast data acquisition, as well as partially or fully automated data analysis. As a first step of an NMR study of protein structure, sequential resonance assignment is usually obtained from triple-resonance experiments correlating the frequencies of at least three nuclei in a multidimensional spectrum. Generally, a set of three-dimensional (3D) experiments is acquired for the backbone assignment. For many protein samples, the minimum experimental time required for recording a 3D spectrum is not determined by sensitivity considerations, but by the necessity of reasonably high spectral resolution in all dimensions, required for reliable peak picking and automated assignment. Some years ago it has been shown (Simorre et al., 1994; Brutscher et al., 1994) that the same spectral information is obtained from a series of 2D spectra where two nuclei are frequency labeled in the same indirectly recorded dimension by employing the accordion principle (Bodenhausen and Ernst, 1982). Such *reduced dimensionality* approaches (Szyperski et al., 1993) allow a significant reduction in experimental time, data processing, and spectral analysis. Here we present a new set of triple-resonance experiments optimized in terms of resolution and sensitivity, where some of the problems related to the initially proposed pulse sequences are removed. The first six experiments allow sequential assignment of the H^{N}, N, C^{α}, H^{α}, and C^{β} nuclei. A seventh experiment adds the C' frequencies to the assignment

*To whom correspondence should be addressed. E-mail: Bernhard.Brutscher@ibs.fr

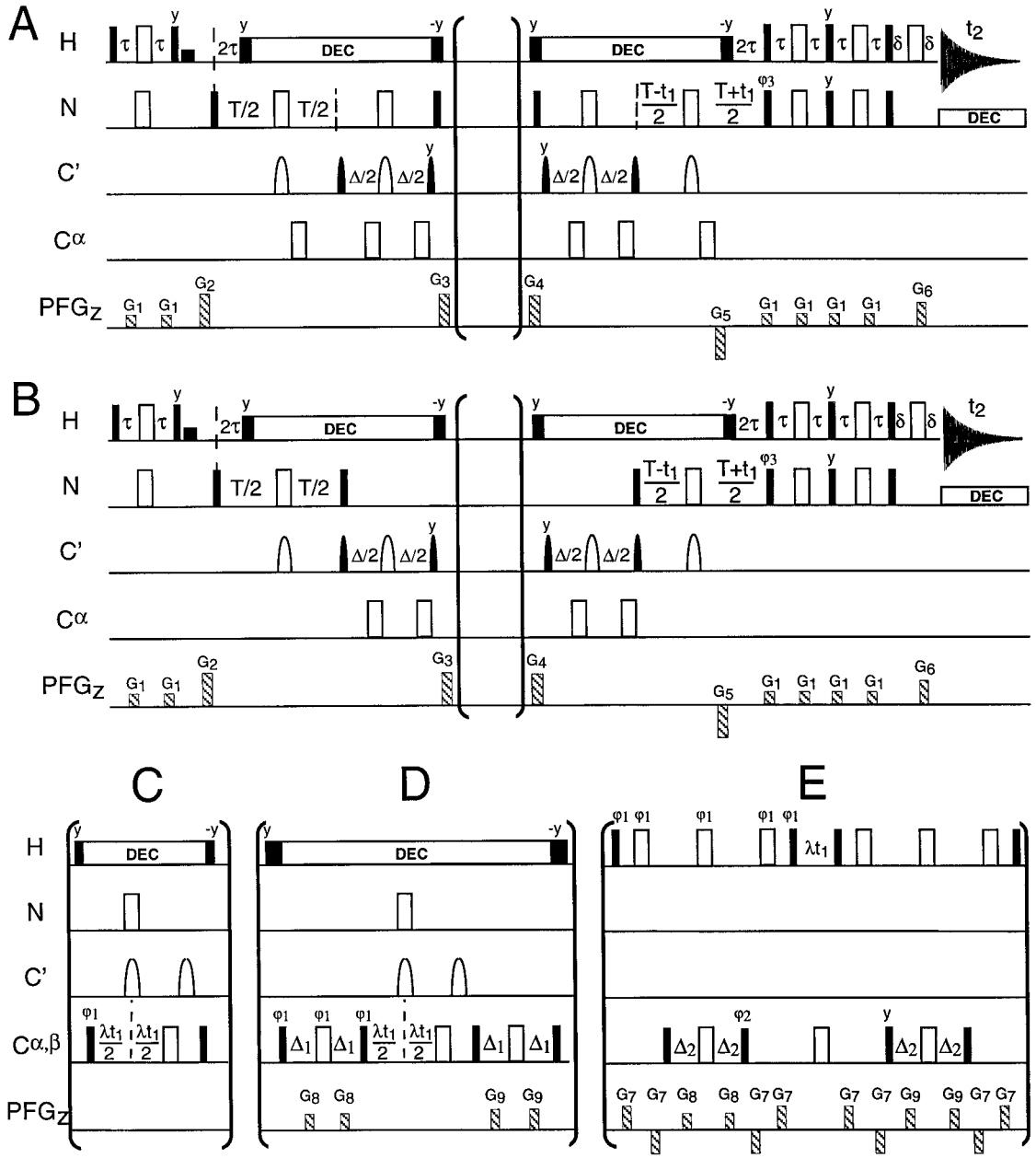

Figure 1. Pulse sequences for 2D H-N-X correlation experiments. A set of 6 experiments can be generated by inserting one of the blocks labeled (C), (D), or (E) into the common pulse sequence part for (A) *intra-residue* and (B) *sequential* correlation sequences: (A) & (C): HNCA; (A) & (D): HN(CA)CB, (A) & (E): HN(CA)HA; (B) & (C): HN(CO)CA; (B) & (D): HN(COCA)CB; (B) & (E): HN(COCA)HA. All radio-frequency (rf) pulses are applied along the x axis unless indicated. 90° and 180° rf pulses are represented by filled and open pulse symbols, respectively. The ^{13}C pulses applied to C′ have the shape of the center lobe of a sinx/x function, whereas the C$^{\alpha}$ pulses are applied with a rectangular shape and field strength of $\Delta/\sqrt{15}$ (90°) and $\Delta/\sqrt{3}$ (180°), where Δ is the separation in Hz between the centers of the C$^{\alpha}$ and C′ chemical shift regions. For all experiments the ^{13}C carrier is centered on the C$^{\alpha}$ region (57 ppm) during the whole pulse sequence. Excitation of the C′ and C$^{\alpha,\beta}$ frequency regions is obtained as usual by linear phase incrementation during the pulses. Water-selective $\pi/2$ pulses serve to avoid saturating the water resonance. The transfer delays are adjusted to $\tau = 1/(4J_{NH}) \cong 2.7$ ms, $T = 1/(2J_{NC'}) \cong 34$ ms, $\Delta = 1/(2J_{C^{\alpha}C'}) \cong 9$ ms, $\Delta_1 = 1/(4J_{C^{\alpha}C^{\beta}}) \cong 7$ ms, and $\Delta_2 = 1/(4J_{C^{\alpha}H^{\alpha}}) \cong 1.7$ ms. Pulsed field gradients, G_1, G_2, G_3, G_4, G_5 and G_6 are applied along the z-axis (PFG$_z$) with a gradient strength of approximately 20 G/cm and lengths ranging from 100 to 2000 μs, followed by a recovery delay of 100 μs. The relative durations of G_5 and G_6 are given by the gyromagnetic ratios of ^{1}H and ^{15}N as $G_5/G_6 = \gamma_H/\gamma_N$. A two-step phase cycle is used where φ_1 and the receiver phase are inverted simultaneously. Four experiments are recorded per t_1 increment with the following phase settings: (I) $\varphi_1 = x$, $\varphi_3 = x$; (II) $\varphi_1 = y$, $\varphi_3 = x$; (III) $\varphi_1 = x$, $\varphi_3 = -x$; (IV) $\varphi_1 = y$, $\varphi_3 = -x$. In addition the

sign of G_5 is inverted for (III) and (IV). In the case of the H^α correlation experiments an additional phase cycling of φ_2 allows separation of glycine from non-glycine signals. Two sets of experiments with $\varphi_2 = y$ and $\varphi_2 = -y$ are recorded and separately stored. The glycine spectrum is obtained from the sum of the two data sets, whereas the difference yields the signals from all non-glycine residues. Before Fourier transformation the four data sets have to be combined adequately. First, a set of purely amplitude-modulated data is reconstructed. The linear combination (I)+(III) yields the time evolution of $2N_xX_x$ coherence. The combination (I)−(III) corresponds to $2N_yX_x$, (II)+(IV) to $2N_xX_y$, and (II)−(IV) to $2N_yX_y$. In addition the real and imaginary parts along t_2 are inverted (90° phase shift) for the data corresponding to $2N_yX_x$ and $2N_yX_y$. In a second step, the real and imaginary parts of the time domain data for the ZQ and the DQ spectra are obtained as described by Equations 1–4 or Equations 5–8. Pulse sequences in Varian pulse sequence language and transformation protocols in Felix macro language can be obtained from the authors upon request.

and measures the two spin-spin coupling constants $^1J_{NH}$ and $^2J_{C'H}$ along the protein backbone. The latter is particularly interesting when the experiment is repeated on a partially aligned protein sample using a suitable alignment medium (Bax et al., 2001). All experiments are demonstrated on a small protein segment consisting of the 68 N-terminal residues of the cytosolic mercuric reductase merA from *Ralstonia metallidurans* CH34. The protein is part of the bacterial response machinery to mercury stress. The N-terminal fragment, which we will refer to as MerAa, contains two cysteine residues which can bind a Hg(II) ion. No three-dimensional structure of this protein fragment is available so far. Because of the high sequence homology, MerAa is suspected to belong to a family of proteins including metallo-chaperones and metal-transporting ATPases (Arnesano et al., 2002). Here we report backbone assignment of the metal-free and the Hg(II)-bound forms of MerAa obtained from a set of 2D spectra recorded in an overall experimental time of only 15 h per sample. The secondary structural elements identified from the C^α and C' chemical shifts provide a first useful information on the protein structure. In addition the measured residual dipolar couplings D_{NH} and $D_{C'H}$ can be used for experimental validation of a structural homology model.

Unidirectional 2D H-N-X experiments

Sequential resonance assignment of the protein backbone makes use of a set of experiments correlating the amide proton and nitrogen frequencies of one residue with the C^α, C^β, C', and/or H^α frequencies of the same (*intra-residue correlation*) or the preceding residue (*sequential correlation*). Several pairs of such intra-residue and sequential H-N-X experiments are necessary to resolve an eventual degeneracy in the frequency of the X nucleus, and to allow for unambiguous sequential assignment. In the ideal case, one cross peak per residue (except for proline residues) is detected in these spectra. The relative small number of correlation peaks in the spectrum makes these experiments particularly well suited for low-dimensionality implementations as demonstrated previously (Simorre et al., 1994; Brutscher et al., 1994; Ding and Gronenborn, 2002). The major drawback of the original 2D experiments was the presence of a sequential correlation peak in the intra-residue correlation spectra due to the similar size of the $^1J_{NC\alpha}$ and $^2J_{NC\alpha}$ scalar couplings. This doubles the number of correlation peaks and makes spectral interpretation more difficult due to increased accidental peak overlap and the distinction of intra-residue and sequential correlation peaks. Recently, a new pulse sequence element has been introduced which performs a selective intra-residue $N \rightarrow C^\alpha$ coherence transfer (Brutscher, 2002; Nietlispach et al., 2002; Permi, 2002; Ding and Gronenborn, 2002). As already recognized by Ding and Gronenborn (2002), an assignment strategy based exclusively on 2D experiments is significantly improved by the use of a set of intra-residue H-N-X experiments which include this new pulse sequence element.

Figure 1 shows a set of optimized pulse sequences to record pairs of intra-residue and sequential 2D H-N-X spectra: HNCA & HN(CO)CA, HN(CA)CB & HN(COCA)CB, and HN(CA)HA & HN(COCA)HA. All experiments are of the out and back transfer type. The common part of the intra-residue experiments is shown in (A). After the initial INEPT transfer step from the amide proton $H^N(i)$ to the nitrogen N(i), the concatenated pulse sequence element of Brutscher (2002) is used for the selective coherence transfer $N(i) \rightarrow C^\alpha(i)$. The common part of the sequential correlation experiments shown in (B) realizes a relayed transfer $N(i) \rightarrow C'(i-1) \rightarrow C^\alpha(i-1)$. For all experiments, ^{15}N frequency labeling is achieved in a sensitivity enhanced manner (Palmer et al., 1991). Insertion of the pulse sequence blocks labeled (C), (D) and (E) into the sequences of (A) and (B) allows additional frequency labeling of C^α, C^β, or H^α nuclei, respectively. The same incremented time delay t_1 is used

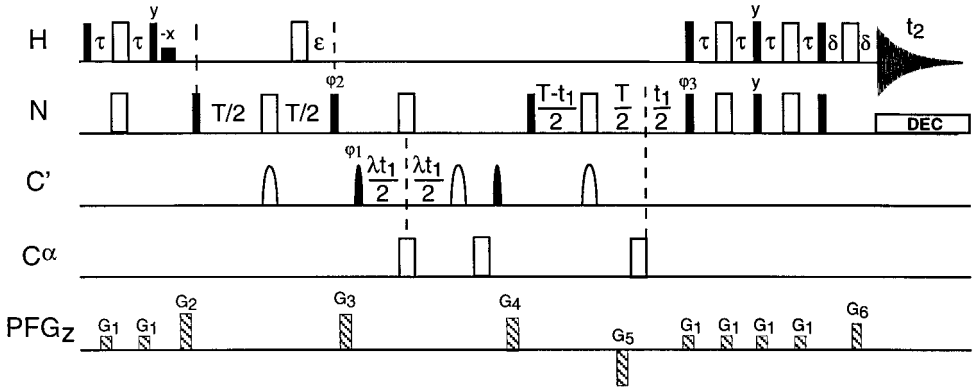

Figure 2. Pulse sequence for the 2D J-HNCO(H) experiment. Acquisition parameters are identical to the sequences in Figure 1. An additional DIPSAP-filter (Brutscher, 2001) is applied for α/β spin-state selection of the amide proton during t_1 evolution. For the DIPSAP filter the experiment is repeated three times with the following delay and phase settings: (1) $\varepsilon = 0$, $\varphi_2 = y$, (2) $\varepsilon = 1/(4J_{NH})$, $\varphi_2 = x$, and (3) $\varepsilon = 1/(2J_{NH})$, $\varphi_2 = y$. The α- and β-sub-spectra are obtained from the three data sets by linear combination $0.73 \times (1) \pm (2) - 0.27 \times (3)$. Otherwise, the same transformation rules apply as explained for the sequences in Figure 1.

to encode the time evolution of the two spins N and X (C^{α}, C^{β}, H^{α}). As for the corresponding 3D versions, four FIDs are recorded per t_1 increment with different phase settings of φ_1 and φ_2 as explained in the figure caption. This quadrature detection scheme allows reconstruction of two sub-spectra where the resonance frequencies along the ω_1 dimension correspond to the sum and the difference of the chemical shifts of the N and X nuclei (Brutscher et al., 1995). After the usual manipulation of the data accounting for the sensitivity enhanced ^{15}N quadrature detection (see figure caption), the four data sets correspond to the t_1 time modulation of $2N_xX_x$, $2N_xX_y$, $2N_yX_x$, and $2N_yX_y$ coherence. It is thus convenient to describe the unitary spin evolution during t_1 by the concept of zero-quantum (ZQ) and double-quantum (DQ) coherence spectroscopy, although no such coherences are present during the pulse sequence. The real and imaginary part of the ZQ and DQ spectra are obtained by linear combination of the four data sets in the following way:

$$ZQ_x^{NX} = \left(2N_xX_x + 2N_yX_y\right)/2, \quad (1)$$

$$ZQ_y^{NX} = \left(2N_xX_y - 2N_yX_x\right)/2, \quad (2)$$

$$DQ_x^{NX} = \left(2N_xX_x - 2N_yX_y\right)/2, \quad (3)$$

$$DQ_y^{NX} = \left(2N_xX_y + 2N_yX_x\right)/2. \quad (4)$$

The resulting 2D DQ and ZQ spectra are identical to a ^1H-^{15}N correlation map where the individual peaks are shifted along the ω_1 dimension by $+\lambda\omega_x$ and $-\lambda\omega_x$, respectively. The scaling factor λ is adjusted to account for the different spin relaxation and scalar coupling properties of the two nuclei and to optimize spectral resolution and sensitivity (Brutscher et al., 1994; Szyperski et al., 1994). The separation of ZQ and DQ peaks into different sub-spectra allows maintaining the carrier frequency of the X nucleus in the center of the chemical shift range. In previous implementations where the two peaks are detected in the same spectrum, the carrier had to be shifted to the edge of the spectrum by time-proportional phase incrementation (TPPI) to allow for unambiguous sign discrimination. Note that this procedure is conceptually similar to the recently introduced GFT NMR approach of Kim and Szyperski (2003). For large proteins it may become interesting to encode N nuclear frequencies ($N > 2$) in the same incremented time dimension. A series of 2D spectra is created by appropriate linear combination of the recorded 2^N quadrature data sets, where the peak positions along the ω_1 frequency dimension correspond to different linear combinations of the chemical shifts of the N nuclei.

The HN(CA)HA and HN(COCA)HA experiments (insert E) are further optimized for simultaneous detection of $C^{\alpha}H_2$ (glycine residues) and $C^{\alpha}H$ (all other residues) spin systems. Spin evolution during t_1 for the non-glycine residues is described by the single-quantum (SQ) product operator $2C_zH_y$. For glycine residues, the 90° ^1H pulse after the initial INEPT transfer from C^{α} to H^{α} creates a multiple-quantum (MQ) coherence of the form $4C_yH_y^1H_y^2$. ^{13}C chemical shift evolution is refocused by the 180° ^{13}C pulse applied in the middle of t_1, and only the chemical shifts of the two proton spins are encoded in the t_1 evolution. Addition or subtraction of two data sets recorded with phase settings $\varphi_2 = y$ and $\varphi_2 = -y$

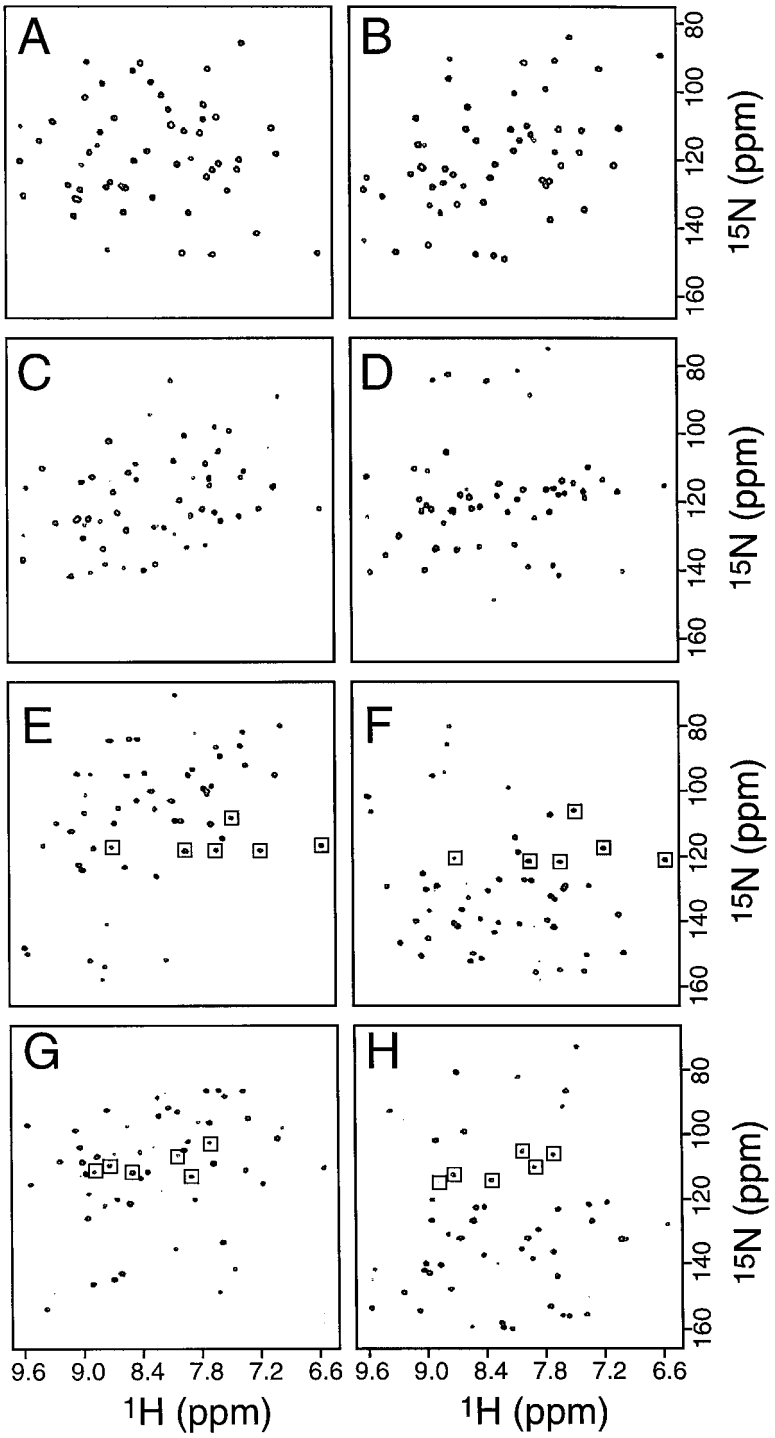

Figure 3. 2D correlation spectra recorded on a sample of metal-free MerAa at 600 MHz ^{1}H frequency: (A) ZQ-HN(CO)CA, (B) DQ-HN(CO)CA, (C) ZQ-HNCA, (D) DQ-HNCA, (E) ZQ-HN(COCA)CB, (F) DQ-HN(COCA)CB, (G) ZQ-HN(CA)CB, and (H) DQ-HN(CA)CB. Data sets of 512 × 120 complex points were recorded for spectral widths of 7000 Hz (^{1}H) × 6000 Hz (DQ/ZQ). For glycine residues without an attached C^{β} carbon, the correlation peaks in spectra (E-H) are negative (boxes), and the frequency shift in the ^{15}N dimension corresponds to $\pm\,\lambda\omega(C^{\alpha})$.

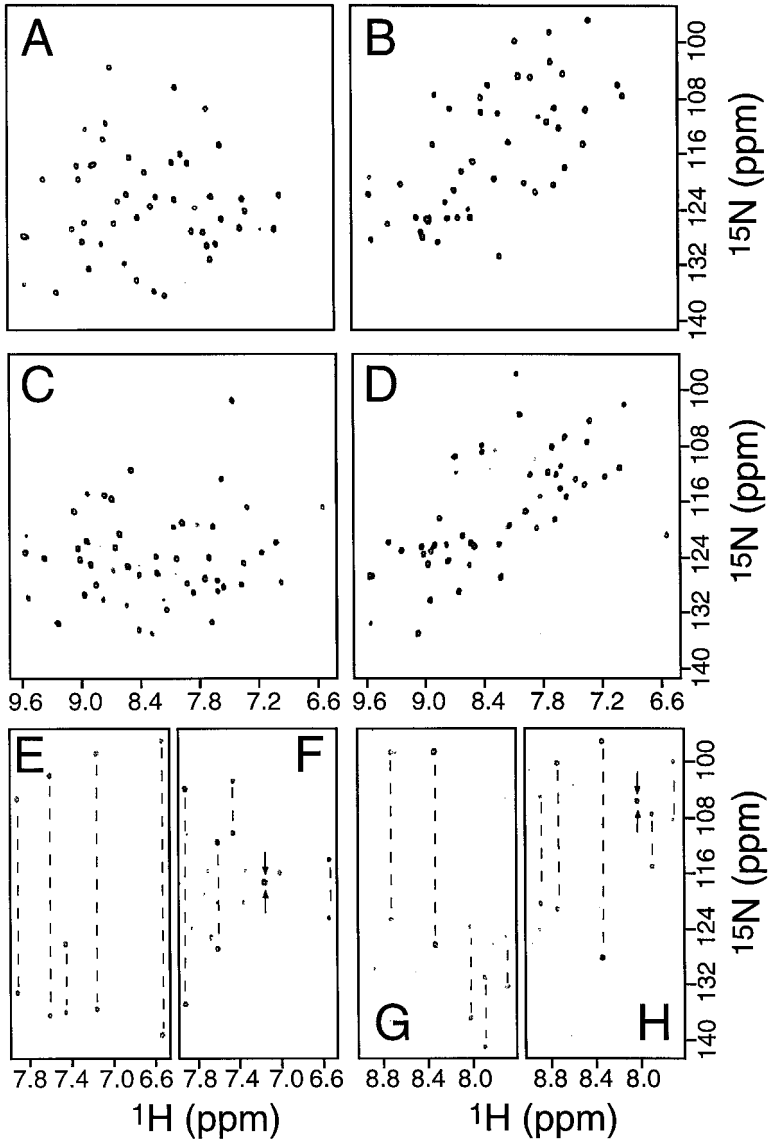

Figure 4. 2D H-N-HA correlation spectra recorded on a sample of metal-free MerAa at 600 MHz ^{1}H frequency for non-glycine (A–D) and glycine residues (E–H). (A) and (E): ZQ-HN(COCA)HA, (B) and (F): DQ-HN(COCA)HA, (C) and (G): ZQ-HN(CA)HA, (D) and (H): DQ-HN(CA)HA. Data sets of 512 × 83 complex points were recorded for spectral widths of 7000 Hz (^{1}H) × 3000 Hz (DQ/ZQ).

allows the separation of the glycine from the non-glycine signals. While the ZQ and DQ spectra for non-glycine residues are obtained using the procedure described above (Equations 1–4), a slightly different protocol has to be applied for the glycine data. The four quadrature data sets corresponding to the evolution of $2N_xH_xH_x$, $2N_xH_yH_y$, $2N_yH_xH_x$, and $2N_yH_yH_y$ coherence are combined as follows:

$$2N_xZQ_x^{HH} = (4N_xH_xH_x + 4N_xH_yH_y)/2, \quad (5)$$

$$2N_yZQ_x^{HH} = (4N_yH_xH_x + 4N_yH_yH_y)/2, \quad (6)$$

$$2N_xDQ_x^{HH} = (4N_xH_xH_x - 4N_xH_yH_y)/2, \quad (7)$$

$$2N_yZQ_x^{HH} = (4N_yH_xH_x - 4N_yH_yH_y)/2. \quad (8)$$

Again two ^{1}H-^{15}N HSQC-like spectra are generated from these data where the cross peaks are splitted along the ^{15}N dimension by either $\omega_{ZQ} = \omega_{H1} - \omega_{H2}$ (SQ^NZQ^{HH}-spectrum), or (SQ^NDQ^{HH}-spectrum). For glycine residues, the chemical shifts of four different nuclei (H^N, N, H_1^α, and H_2^α) are obtained from a pair of 2D spectra. Note that the intensity of the correlation peaks in the glycine spectra is only about half the

intensity obtained for the other residues. In addition, the water magnetization is manipulated in a controlled manner by inserting additional ^{1}H pulses in order to be close to equilibrium between scans. This is important for application to proteins studied at neutral or basic pH to reduce signal loss due to amide proton-water exchange.

The assignment approach introduced so far does not yield any information on the carbonyl chemical shifts, which are useful probes of regular secondary structure in the protein. An HNCO-type experiment is required to extend the backbone assignment to the carbonyl ^{13}C. Here we propose an implementation of the HNCO experiment, which not only provides the carbonyl chemical shift information but also measures spin-spin couplings between the amide proton and nitrogen $^1J_{NH}$, and between the amide proton and carbonyl $^2J_{C'H}$. If spectra can be recorded for samples in isotropic and anisotropic solution, the residual dipolar contributions D_{NH} and $D_{C'H}$ to these couplings provide precious information about the tertiary fold of the molecule (Bax et al., 2001). The pulse sequence of the 2D J-HNCO(H) experiment is shown in Figure 2. The sequence contains the common features of 2D H-N-X experiments introduced above. No ^{1}H decoupling is applied during ^{13}C′ and ^{15}N frequency editing, which yields a peak splitting of $\omega_{DQ} \pm \pi J_{DQ}$ in the DQ spectrum and $\omega_{ZQ} \pm J_{ZQ}$ in the ZQ spectrum. As for ZQ and DQ spectroscopy, the line splitting observed in the two sub-spectra, J_{DQ} and J_{ZQ}, is related to the sum and the difference of the individual spin-spin couplings as follows:

$$J_{DQ} = {}^1J_{NH} + \lambda^2 J_{C'H}, \quad (9)$$

$$J_{ZQ} = {}^1J_{NH} - \lambda^2 J_{C'H}. \quad (10)$$

In order to retain the same spectral simplicity as for the experiments described above (one correlation peak per residue per spectrum) an additional *J*-mismatch-compensated DIPSAP-type α/β-spin-state selection filter (Brutscher, 2001) is added during the first ^{15}N→^{13}C coherence transfer step to separate the doublet lines into different sub-spectra (Ottiger et al., 1998). Three data sets are recorded with different settings of the delay 'ε' and the phase φ_2 as explained in the figure caption.

Backbone assignment of MerAa

For sequential resonance assignment of MerAa six 2D triple-resonance experiments (Figures 1A–E) were performed on a Varian INOVA 600 spectrometer at 20 °C in an overall time of 14 h. A first set of spectra was recorded on metal-free MerAa in the presence of 10mM DTT. The experiments were then repeated on a second sample containing MerAa and $HgCl_2$ salt in a 1:1.2 molar ratio. Both protein samples were prepared at a concentration of about 1.5 mM in 50 mM TRIS-HCl buffer at pH = 7. The experimental times for the individual experiments were: 30 min for HN(CO)CA, 1 h for HNCA, 2 h 40 min for HN(COCA)CB, 5 h 20 for HN(CA)CB, 1 h 30 for HN(COCA)HA, and 3 h for HN(CA)HA. The scaling factor λ was set to λ = 1 for the C^α and H^α, and to λ = 0.5 for the C^β correlation experiments. Additional acquisition parameters are given in the figure captions. In addition a ^{1}H-^{15}N HSQC spectrum was acquired in an experimental time of 15 min. Off-resonance ^{1}H irradiation was applied between each transient in the ^{1}H-^{15}N HSQC to get similar sample heating as in the triple-resonance experiments (Wang and Bax, 1993).

The spectral resolution can be appreciated from Figures 3 and 4 showing 2D spectra recorded for metal-free MerAa. Each spectrum can be simply understood as a ^{1}H-^{15}N correlation map where the individual cross peaks are shifted by $+\lambda\omega_x$ (DQ) and $-\lambda\omega_x$ (ZQ) along the ^{15}N dimension. In the case of a degenerate amide ^{1}H frequency, the ZQ and DQ peaks detected at this H^N frequency are assigned by applying a symmetry criterion. For each peak detected in the ZQ spectrum another peak has to be present in the DQ spectrum at the inverse distance from the corresponding ^{1}H-^{15}N correlation peak, and vice versa. This procedure also allows the detection of accidental peak overlaps in the ZQ and DQ spectra. In practice, spectral analysis is easily performed either 'by hand' using a multiple-cursor graphical interface, or by a computer script working on a text file of the peak positions obtained by a peak picking routine. A total of 56 correlation peaks are observed in each of the spectra of Figures 3A–H and 4A-D. This corresponds to the number of observed cross peaks in the ^{1}H-^{15}N correlation spectrum of this protein, except for the signals arising from NH_2 groups in side chains (see Figure 5A). In addition, the H^α frequencies of all six glycine residues can be identified from the spectra shown in Figures 4E–H.

As long as the ^{1}H-^{15}N correlation peaks do not overlap, one obtains as the result of this spectral analysis a list of unambiguously assigned protein segments

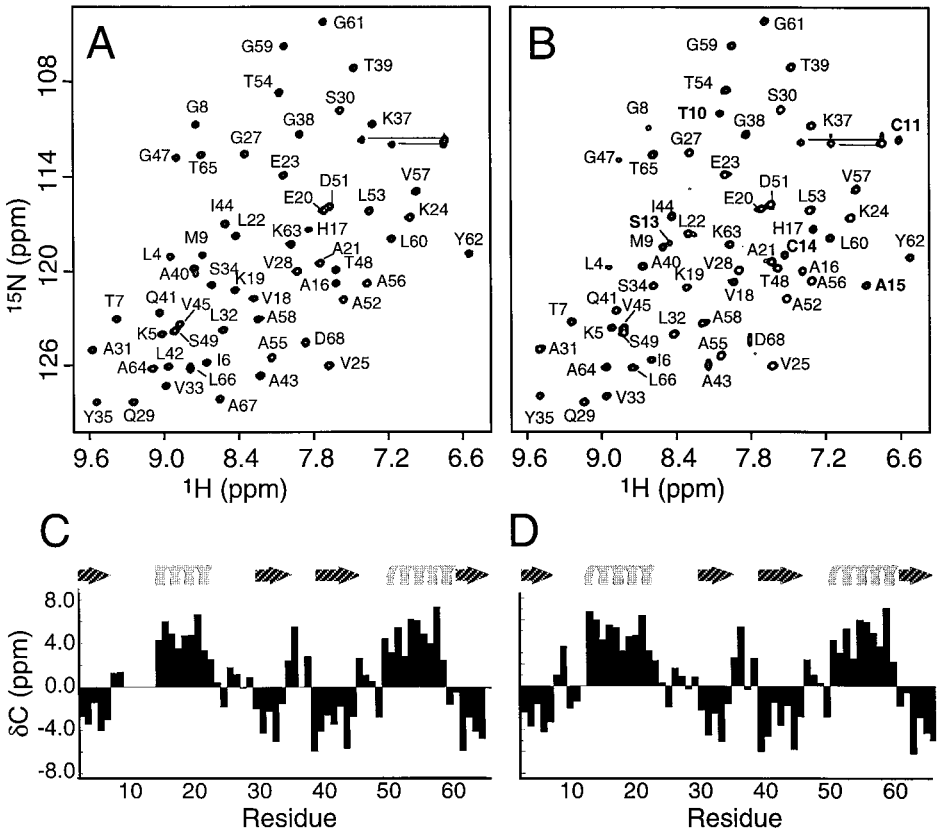

Figure 5. ^{1}H-^{15}N correlation spectra of metal-free (A) and Hg(II)-bound MerAa (B). The residue type and number assign the observed correlation peaks. Additional correlation peaks detected for residues T10, C11, S13, C14, and A15 after metal binding are highlighted by bold assignments. The secondary chemical shift deviations δC plotted as a function of the residue number for the metal-free form (C) and the Hg(II)-bound form (D) allow identification of the secondary structural elements in MerAa.

of the form:

$$C^{\beta}(i-1) - C^{\alpha}(i-1) - H^{\alpha}(i-1) - H^{N}(i) - N(i) - C^{\beta}(i) - C^{\alpha}(i) - H^{\alpha}(i). \quad (11)$$

For glycine residues at the (i) or (i−1) position, the corresponding C^{β} frequency is replaced by the chemical shift of the second H^{α}. Sequential assignment is thus based on the frequency match of three different nuclei, and the amino acid type information contained in the C^{α} and C^{β} chemical shift data (Richarz and Wüthrich, 1978). For MerAa the H^{α} sequential information proved to be crucial to solve C^{α} frequency degeneracy of glycine residues, and to correctly assign them to their position in the primary sequence. Because of the high spectral resolution of the 2D spectra, the precision of the extracted chemical shift values is high. Chemical shift tolerances between different experiments are found to be <0.1 ppm for C^{α} and C^{β}, and <0.02 ppm for H^{α}, which facilitates computer-assisted sequential assignment. Here we used the program ALPS (Morelle et al., 1995) to assign the individual protein segments to positions in the primary sequence. The assigned ^{1}H-^{15}N HSQC spectrum of metal-free MerAa is shown in Figure 5A. Complete backbone assignment is obtained except for the three N-terminal residues, and a region comprising residues T10 to A15, which contains the two cysteines involved in metal ligation. This indicates some conformational flexibility in the absence of a bound Hg(II) ion, which may be important for the biological function of the protein.

In the presence of a Hg(II) ion, this protein region becomes less flexible, as indicated by the additional cross peaks detected for T10, C11, S13, C14, and A15 in the ^{1}H-^{15}N HSQC spectrum of Hg-MerAa shown in Figure 5B. Residue D12 is still conformational exchange-broadened beyond detection.

Secondary structure and backbone fold topology

The 2D J-HNCO(H) experiment was first applied to the two samples of reduced metal-free and Hg(II)-bound MerAa in isotropic solution, and then repeated on a sample of Hg(II)-MerAa dissolved in a solution containing a 5% $C_{12}E_6$/hexanol mixture in a molar ratio of 0.64 (Rückert and Otting, 2000). All spectra were recorded at a ^{1}H frequency of 600 MHz and a sample temperature of 25 °C. Despite the additional line splitting, high signal to noise is obtained in a short experimental time, 45 min in the present case, because of the high intrinsic sensitivity of the HNCO experiment. A small part of the ZQ and DQ correlation spectra comprising residues C11, A15, K24, and Y62 is shown in Figure 6A for the alcohol sample. The scaling factor λ was set to $\lambda = 2.0$. This ensures high spectral resolution, and high precision for the measurement of the C′ chemical shifts and the relatively small $^2J_{C'H}$ coupling constants. The experiment allowed C′ frequency assignment for all residues visible in the ^{1}H-^{15}N HSQC spectrum. In addition D_{NH} and $D_{C'H}$ couplings could be measured accurately for 52 residues of Hg(II)-MerAa. The precision of the coupling constants as calculated from two different data sets was on average about $\sigma_{NH} = 1.0$ Hz for the D_{NH} and $\sigma_{C'H} = 0.5$ Hz for the $D_{C'H}$ couplings.

The regular secondary structural elements of MerAa in the metal-free and Hg(II)-bound form were identified from the ^{13}C secondary chemical shifts δC calculated as $\delta C = (\omega C'_{exp} - \omega C'_{rc}) + (\omega C^{\alpha}_{exp} - \omega C^{\alpha}_{rc})$, where ωC_{rc} are the amino-acid-type-specific random coil chemical shift values (Richarz and Wüthrich, 1978). The δC chemical shift data for metal-free MerAa (Figure 5C) and Hg(II)-bound MerAa (Figure 5D) are in accordance with the classical 'ferredoxin-like' β-α-β-β-α-β fold found in metallo-chaperones and metal-transporting ATPases (Arnesano et al., 2002). The major difference between the two forms is a 3-residue N-terminal extension of the first α-helix upon metal binding. This region of the protein, containing the two metal-binding cysteine residues, seems to be in conformational exchange in the metal-free protein.

In order to get more information about the spatial arrangement of the experimentally determined structural elements, we used the measured residual dipolar couplings in the Hg(II)-bound form of MerAa. Measurement of two couplings per peptide plane, D_{NH} and $D_{C'H}$, provides a reasonable sampling of orientational space for the different structural elements (α-helices and β-strands). This is not the case if only information on the ^{1}H-^{15}N bond vectors is available which, in regular secondary structural elements, all point more or less along the helix axis or orthogonal to the β-strand direction. The $D_{C'H}$ couplings are measured with two-times higher precision than the D_{NH} couplings (see above) which partially compensates for the different coupling strength of $D_{C'H}$ and D_{NH} ($D^{max}_{CH} = 0.3\ D^{max}_{NH}$). Residual dipolar couplings may be used in many different ways for molecular structure determination (Bax et al., 2001), ranging from molecular fold validation to structure refinement or even *de novo* structure calculation. In the present context, the program Module (Dosset et al., 2001) was used to evaluate the accordance between a structural homology model and the measured $D_{C'H}$ and D_{NH} couplings. The solution structure of the copper-bound first domain of the Ccc2 ATPase from yeast (Banci et al., 2001) was chosen as a template. The two sequences were aligned, taking into account sequence similarity and the experimentally derived secondary structural information, and a structural model was then obtained simply by projection of the MerAa sequence on the Ccc2 ATPase structure. The alignment tensor, characterized by the axial and rhombic components A_a and A_r, and three Euler angles was obtained by nonlinear least-squares minimization of the target function

$$\chi^2 = \sum_i \left\{ \left(D^{exp}_{NH_i} - D^{calc}_{NH_i} \right)^2 / \sigma^2_{NH} + \left(D^{exp}_{C'H_i} - D^{calc}_{C'H_i} \right)^2 / \sigma^2_{C'H} \right\}.$$

Only couplings measured for residues in the core of the secondary structural elements were used for the calculation. RDC data from the terminal residues of helices and strands, as identified from the ^{13}C CSI, as well as the first three residues of helix 1, only visible in the Hg(II)-bound form, were removed as they introduce an un-proportionally high increase in the χ^2-value. All residues used for the model validation are highlighted in black on the backbone ribbon representation shown in Figure 6B. The quality of the model structure can be evaluated by comparing measured dipolar couplings with couplings back-calculated from the best-fit alignment tensor, shown in Figure 6C. The correlation is reasonably good ($r = 0.92$ and $r = 0.96$ for $D_{C'H}$ and D_{NH}, respectively), which allows us to conclude that the spatial arrangement of the α-helices and β-sheets is similar to the backbone fold of the Ccc2 ATPase domain. Such a crude model represents a

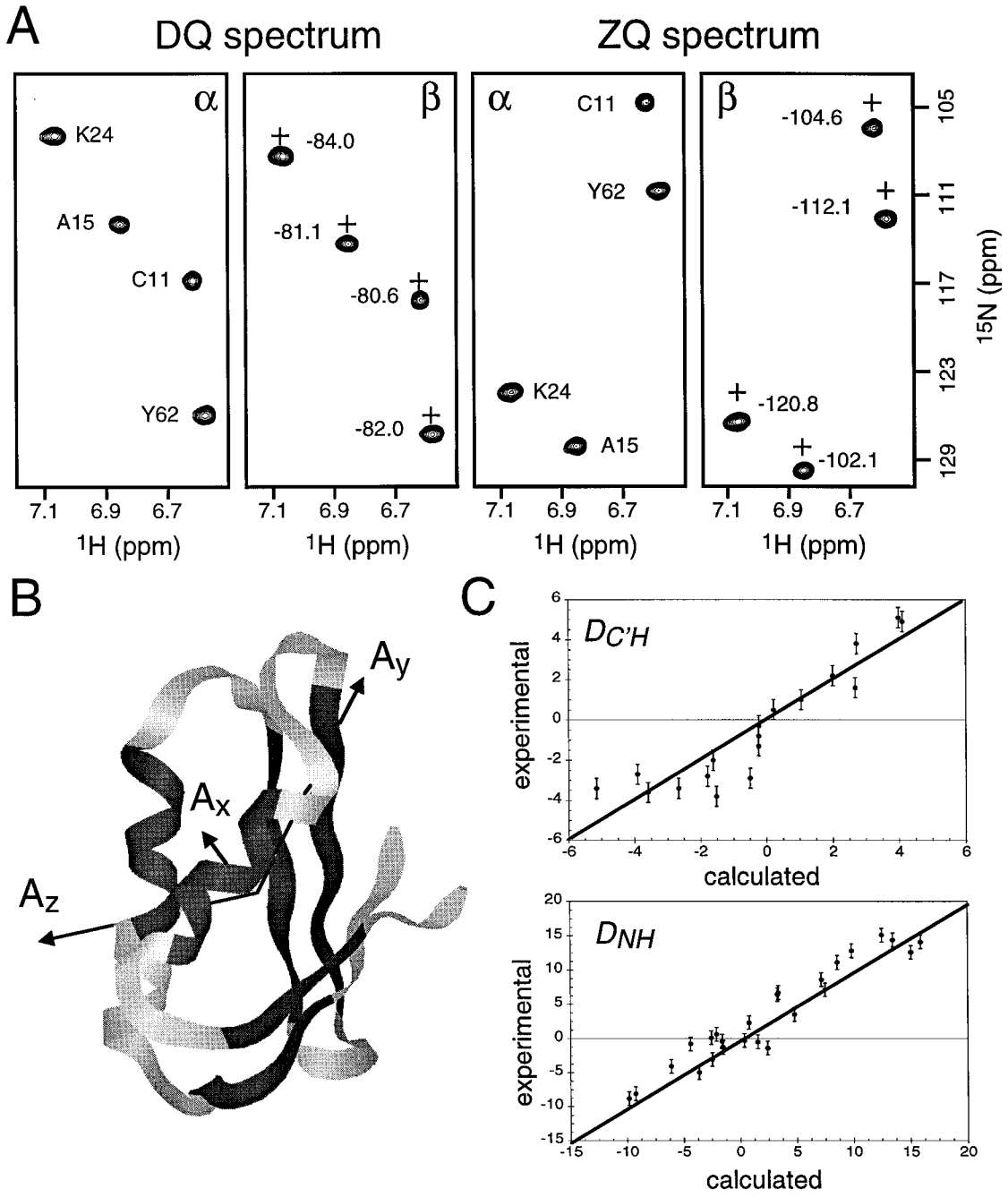

Figure 6. (A) Small part of the DQ-α, DQ-β, ZQ-α, and ZQ-β J-HNCO(H) spectra recorded with the pulse sequence of Figure 2 on a sample of Hg(II)-MerAa in anisotropic solution. The correlation peaks of four residues are visible in these spectra. The coupling constants J_{DQ} and J_{ZQ} (in Hz), as indicated in the plots of the β sub-spectra, are obtained from the peak separation along the ^{15}N dimension between the α- and β- sub-spectra. They can be converted to values for the spin coupling constants J_{NH} and $J_{C'H}$ using Equations 9 and 10. In the α sub-spectra, the corresponding residue type and number annotate the observed signals. (B) Structural homology model for MerAa based on the solution structure of the copper-bound first domain of the Ccc2 ATPase from yeast (Banci et al., 2001). Only RDC data measured for residues labeled in black on the ribbon representation are used for the model validation. Arrows indicate the orientation of the best-fit alignment tensor. The axial and rhombic tensor components are $A_a = 7.4e\text{-}4$ and $A_r = 2.9e\text{-}4$. (C) Correlation plots between experimental and calculated residual dipolar couplings. Linear regression yields correlation coefficients r and slopes m of r = 0.92 and m = 1.02 for $D_{C'H}$, and r = 0.96 and m = 0.95 for D_{NH}.

good starting point for further structural investigations by NMR.

Conclusions

We have presented a set of sensitivity and resolution optimized two-dimensional H-N-X correlation experiments for fast sequential resonance assignment of small to medium-sized proteins. These experiments allow recording of high-resolution spectra in an experimental time, which is not determined by spectral resolution requirements but by the signal to noise ratio obtained for a given sample at a given spectrometer. As long as sensitivity is not the limiting factor, this 2D approach will yield a drastic reduction in experimental time when compared to a conventional 3D-based assignment strategy. For many protein samples, the use of a high field NMR spectrometer eventually equipped with a cryogenic probe will provide the necessary sensitivity to record the set of 2D experiments, required for sequential resonance assignment, in typically less than one day. These data immediately provide experimental determination of the location of secondary structural elements, which are a valuable input for fold prediction algorithms. In a high-throughput structural-genomics context, these fold predictions can then be used for choosing the most interesting targets for further experimental structure analysis. We have also shown that a set of two backbone dipolar couplings per peptide plane, obtained from the same set of experimental data, allows validation or rejection of a structural model obtained from homologous protein structures. Furthermore, resonance assignment provides the basis for the mapping of molecular interfaces and ligand screening by chemical shift perturbation methods, as well as the characterization of local backbone dynamics by ^{15}N spin relaxation measurements. All these data can be recorded and analyzed in a short time, which makes NMR an attractive technique for high-throughput protein studies, highly complementary to X-ray crystallography.

Acknowledgements

This work was supported by the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique. E.R. acknowledges the receipt of a Ph.D. fellowship from the Région Rhônes-Alpes.

References

Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D.L. and O'Halloran, T.V. (2002) *Genome Res.*, **12**, 255–271.

Banci, L., Bertini, I., Ciofi-Baffoni, S., Huffman, D.L. and O'Halloran, T.V. (2001) *J. Biol. Chem.*, **276**, 8415–8426.

Bax, A., Kontaxis, G. and Tjandra, N. (2001) *Meth. Enzymol.*, **339**, 127–174.

Bodenhausen, G. and Ernst R.R. (1982) *J. Am. Chem. Soc.*, **104**, 1304–1309.

Brutscher, B., Morelle, N., Cordier, F. and Marion D. (1995) *J. Magn. Reson.*, **B109**, 238–242.

Brutscher, B., Simorre, J.-P., Caffrey, M. and Marion, D. (1994) *J. Magn. Reson.*, **B105**, 77–82.

Brutscher, B. (2001) *J. Magn. Reson.*, **151**, 332–338.

Brutscher, B. (2002) *J. Magn. Reson.*, **156**, 155–159.

Ding, K. and Gronenborn, A. (2002) *J. Magn. Reson.*, **156**, 262–268.

Dosset, P., Hus J.-C. and Blackledge, M. (2001) *J. Biomol. NMR*, **20**, 927–936.

Kim, S. and Szyperski, T. (2003) *J. Am. Chem. Soc.*, **125**, 1385–1393.

Montelione, G.T., Zhen, T., Huang, Y.J., Gunsalus, K.C. and Szyperski, T. (2000) *Nat. Struct. Biol.*, **7**, 982–984.

Morelle, N., Brutscher, B., Simorre, J.-P. and Marion, D. (1995) *J. Biomol. NMR*, **5**, 154–160.

Nietlispach, D., Yutaka, I. and Laue, E. (2002) *J. Am. Chem. Soc.*, **124**, 11199–11207.

Ottiger, M., Delaglio, F. and Bax, A. *J. Magn. Reson.*, **131**, 373–378.

Palmer III, A.G., Cavanagh, J., Wright, P. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.

Permi, P. (2002) *J. Biomol. NMR*, **23**, 201–209.

Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.

Rückert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.

Simorre, J.-P., Brutscher, B., Caffrey, M. and Marion, D. (1994) *J. Biomol. NMR*, **4**, 325–333.

Szyperski, T., Pellecchia, M. and Wüthrich, K. (1994) *J. Magn. Reson.*, **B105**, 188–191.

Szyperski, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993) *J. Am. Chem. Soc.*, **115**, 9307–9308.

Wang, A.C. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 715–720.