



## Communication

## Sparsely sampled high-resolution 4-D experiments for efficient backbone resonance assignment of disordered proteins

Jie Wen<sup>a</sup>, Jihui Wu<sup>a,b,\*</sup>, Pei Zhou<sup>c,\*</sup><sup>a</sup> Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui, PR China<sup>b</sup> School of Life Science, University of Science and Technology of China, Hefei, Anhui, PR China<sup>c</sup> Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

## ARTICLE INFO

## Article history:

Received 21 October 2010

Revised 17 December 2010

Available online 4 January 2011

## Keywords:

Fast NMR

Sparse sampling

Sequential assignment

FFT-CLEAN

IDP

## ABSTRACT

Intrinsically disordered proteins (IDPs) play important roles in many critical cellular processes. Due to their limited chemical shift dispersion, IDPs often require four pairs of resonance connectivities ( $H^\alpha$ ,  $C^\alpha$ ,  $C^\beta$  and CO) for establishing sequential backbone assignment. Because most conventional 4-D triple-resonance experiments share an overlapping  $C^\alpha$  evolution period, combining existing 4-D experiments does not offer an optimal solution for non-redundant collection of a complete set of backbone resonances. Using alternative chemical shift evolution schemes, we propose a new pair of 4-D triple-resonance experiments – HA(CA)CO(CA)NH/HA(CA)CONH – that complement the 4-D HNCACB/HN(CO)CACB experiments to provide complete backbone resonance information. Collection of high-resolution 4-D spectra with sparse sampling and FFT-CLEAN processing enables efficient acquisition and assignment of complete backbone resonances of IDPs. Importantly, because the CLEAN procedure iteratively identifies resonance signals and removes their associating aliasing artifacts, it greatly reduces the dependence of the reconstruction quality on sampling schemes and produces high-quality spectra even with less-than-optimal sampling schemes.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Intrinsically disordered proteins (IDPs) represent a large fraction of the eukaryotic proteome (>30%) [1]. These proteins can function without folding into a specific and inherently stable structure. IDPs are increasingly recognized for their important roles in the regulation of vital cellular processes, including gene expression, molecular recognition and cellular signaling. Dysfunction of IDPs has been linked to many human diseases, including cancer, diabetes, neurodegenerative diseases, and cardiovascular disease. Consequently, IDPs have become valuable drug targets [2]. Because X-ray crystallography is unable to capture the dynamic information of IDPs in disordered states, high-resolution NMR spectroscopy has become the leading tool for probing dynamics and binding-induced structural changes of IDPs in solution.

A prerequisite step for protein dynamics and functional studies is to obtain complete backbone resonance assignments. A lack of secondary structures in IDPs unfortunately results in limited

chemical shift dispersion, thus hindering the assignment process. Increasing the digital resolution or dimensionality, or both, can be utilized to overcome the challenge of limited chemical shift dispersion and to achieve sequential assignment of IDPs. Both of these approaches however require prolonged acquisition times. On the other hand, IDPs possess slow-relaxing NMR signals due to their inherent flexibility, which offers adequate sensitivity for using the fast NMR methods to achieve signal separation through higher-dimensional spectroscopy at high-resolution. Indeed, several approaches based on automated projection spectroscopy, hyperdimensionality, and discrete Fourier transform of sparsely sampled time domain data have been successfully applied to the assignment of IDPs, offering a significant reduction of measurement times and greatly expanding the range of IDPs suitable for NMR studies [3–8].

Compared to well-folded proteins, for which the  $C^\alpha/C^\beta$  information is often sufficient to establish connectivity, IDPs frequently require all four pairs of nuclei ( $H^\alpha$ ,  $C^\alpha$ ,  $C^\beta$  and CO) for assignments. In several studies, we and others have demonstrated the benefit of high-resolution 4-D experiments with sparse sampling schemes [9–15]. So far, the application of 4-D triple-resonance experiments to protein sequential assignments has been largely restricted to the application of existing 4-D conventional experiments, which, due to the overlapping evolving nuclei (mostly  $C^\alpha$ ) among these experiments, do not offer the most efficient approach for collecting

\* Corresponding authors. Addresses: Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui, PR China (J. Wu), 270 Sands Bldg., Research Drive, Durham, NC 27710, USA (P. Zhou). Fax: +86 551 3600374 (J. Wu), +1 919 684 8885 (P. Zhou).

E-mail addresses: [wujihui@ustc.edu.cn](mailto:wujihui@ustc.edu.cn) (J. Wu), [peizhou@biochem.duke.edu](mailto:peizhou@biochem.duke.edu) (P. Zhou).

a complete set of backbone resonances. For example, it frequently takes three pairs of 4-D triple-resonance experiments, HNCACB/HN(CO)CACB, HNCACO/HNCOCA, HNCAHA/HN(CO)CAHA to collect a complete set of backbone resonances. To avoid redundant information collection, we demonstrate the use of two pairs of high-resolution 4-D experiments, HNCACB/HN(CO)CACB and HA(CA)CO(CA)NH/HA(CA)CONH with sparse sampling, which exploit alternative and non-overlapping chemical shift evolution periods of existing 4-D experiments for efficient collection of four pairs of sequential connectivity for IDPs. We demonstrate this strategy on the N-terminal domain of SKIPN (residues 59–129, SKIPN; Fig. 1), an IDP that plays an important role in the splicing process [16]. The SKIPN was previously assigned using four pairs of conventional 3-D experiments with an overall data acquisition time over 4 days [16]. By using the concentric shell sampling and FFT-CLEAN processing [9], we were able to achieve complete backbone resonance assignments using two pairs of high-resolution 4-D experiments, with each dataset taking 2.64 h and an overall acquisition time less than 11 h.

## 2. Methods

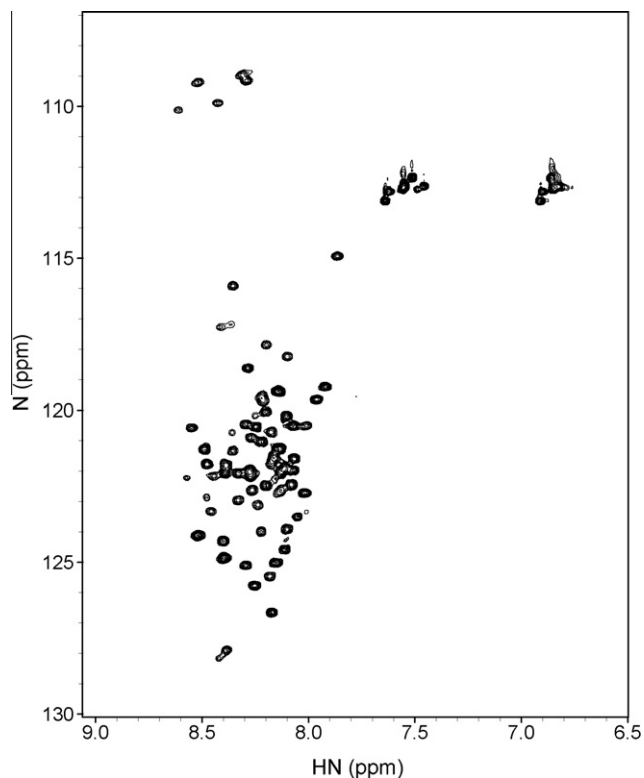
A pair of 4-D sparsely sampled HNCACB and 4-D-HN(CO)CACB experiments were modified from their conventional 4-D counterparts in order to accommodate predefined sparse sampling schemes for the collection of  $C^\alpha/C^\beta$  correlations [17,18]. To avoid collecting redundant chemical shift information (i.e.,  $C^\alpha$ ), instead of running 4-D HNCACO/HNCOCA and HNCAHA/HN(CO)CAHA experiments, we present a modified version of the (HACA)CO(CA)NH and HACA(CO)NH experiments, which were initially proposed by Löhner and Rüterjans [19] and by Boucher et al. [20], respectively. In both experiments, by allowing  $H^\alpha$  and CO chemical shift evolutions, one can readily construct new types of 4-D experiments—HA(CA)CO(CA)NH and HA(CA)CONH (Fig. 2)—for efficient

collection of  $H^\alpha$  and CO resonances as a complementary pair for the 4-D HNCACB/HN(CO)CACB experiments [17].

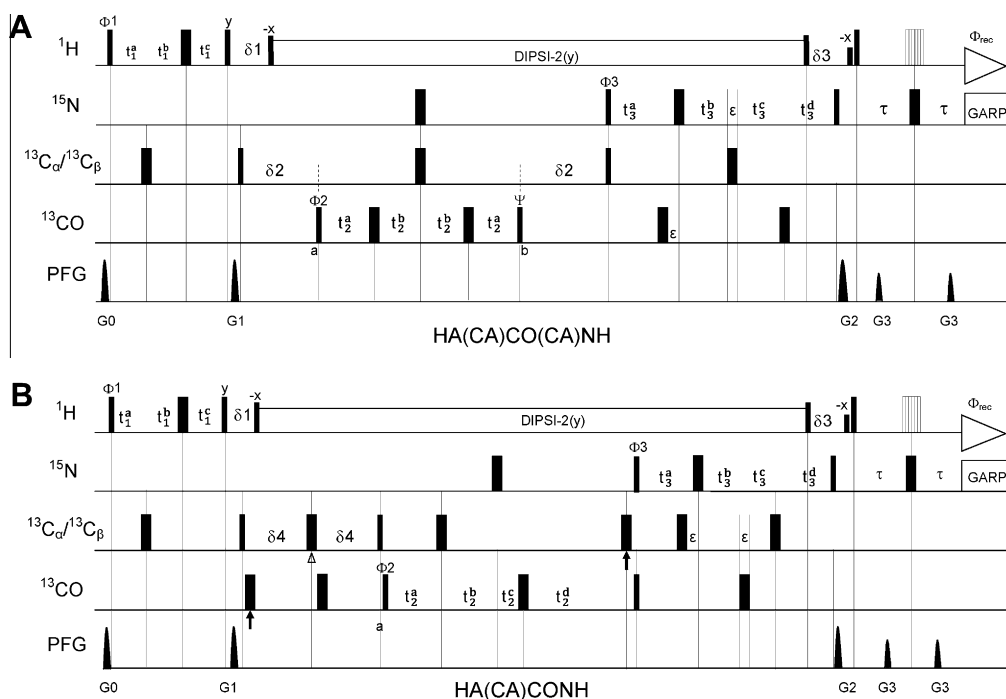
In the HA(CA)CO(CA)NH experiment, the magnetization starts from  $H_i^\alpha$ . After chemical shift evolution, the  $H_i^\alpha$  magnetization is transferred to  $C_i^\alpha$  and then to  $CO_i$  in a multi-quantum manner. After recording the  $CO_i$  chemical shift, the  $CO_i$  magnetization is returned to  $C_i^\alpha$  and then transferred to  $N_{i+1}$  via  $^1J_{NC\alpha}$  and  $^2J_{NC\alpha}$  couplings. The chemical shifts of  $N_{i+1}$  are recorded, and magnetizations are transferred to  $H_{i+1}^N$  for detection, thus providing  $H_i^\alpha CO_i N_{i+1} H_{i+1}^N$  correlations. In the original implementation of the 4-D (HA)CACO-CANH experiment [19], the CO evolution was recorded in a constant time mode, which is sufficient for the limited digital resolution achievable by conventional 4-D experiments. Such a limited CO evolution period, however, is insufficient to accommodate the high digital resolution made available by sparse sampling and required for assignment of IDPs. Therefore, the CO evolution was recorded in a semi-constant time fashion to allow for high-resolution CO separation as initially demonstrated in the 3-D version of the same experiment [19]. Due to the semi-constant time nature, the CO signals are partially modulated by  $^1J_{C\alpha C\beta}$  in addition to  $^1J_{NC\alpha}$  or  $^2J_{NC\alpha}$  couplings. Additionally, the  $C^\alpha$  magnetization stays in the transverse plane for  $\sim 28$  ms in order to refocus the  $^1J_{C\alpha C\beta}$  coupling. A lack of  $^1J_{C\alpha C\beta}$  coupling for glycine residues during this period renders glycine resonances the opposite sign to other residues. The sensitivity of this experiment can be further improved by using selective  $C^\alpha$  inversion pulses or selective  $C^\beta$  decouplings to remove the  $^1J_{C\alpha C\beta}$  coupling. Because such an approach typically results in signal losses for certain types of residues, such as Ser, Leu, Gly, Val, and Pro [21–23], it is not employed here. Technically, the magnetization transfers from  $C_i^\alpha$  to  $CO_i$  and back to  $C_i^\alpha$  can also be implemented in a single-quantum fashion to avoid  $C_i^\alpha$  relaxation during the  $CO_i$  chemical shift evolution period. In such a case, the  $C_i^\alpha$  magnetization would have to stay in the transverse plane for an additional delay of  $\sim 9$  ms ( $0.5/{}^1J_{C\alpha CO}$ ) compared to the MQ implementation described above. This can be particularly detrimental for sensitivity especially if one is interested in collecting complete resonance information regardless of residue types by avoiding the use of selective  $C^\alpha$  inversion pulses or  $C^\beta$  decoupling schemes.

In the HA(CA)CONH experiment, the magnetization also starts from  $H_i^\alpha$ . After chemical shift evolution, the  $H_i^\alpha$  magnetization is transferred to  $C_i^\alpha$  and then to  $CO_i$ , where the  $CO_i$  chemical shift is recorded. The  $CO_i$  magnetization is further transferred to  $N_{i+1}$  and  $H_{i+1}^N$  for chemical shift evolutions. This process produces a 4-D spectrum correlating the  $H_i^\alpha CO_i N_{i+1} H_{i+1}^N$  chemical shifts. Due to the non-redundant signal evolution periods, this new pair of experiments complements the 4-D HNCACB/HN(CO)CACB experiments to enable complete backbone resonance assignment of  $H^N$ , N,  $H^\alpha$ ,  $C^\alpha$ ,  $C^\beta$  and CO.

We demonstrate this approach on human SKIPN, an IDP involved in the splicing process. A high-resolution  ${}^1H$ - ${}^{15}N$  HSQC spectrum with a spectral width of 1621 Hz and a maximum evolution time of 61.7 ms for the nitrogen dimension was collected using a  ${}^{15}N$ -labeled sample (Fig. 1). The limited resonance dispersion of amide signals is characteristic for intrinsically disordered proteins. Two pairs of sparsely sampled 4-D triple-resonance experiments, as described above, were recorded using a 1 mM sample of  ${}^{15}N/{}^{13}C$ -labeled SKIPN. The NMR buffer contains 50 mM sodium phosphate, 50 mM NaCl, 1 mM EDTA, 10% (v/v)  $D_2O$ , pH 6.5. A sparse sampling scheme containing 201 points was generated using cosine-weighted concentric shells as described previously [9]. The orientations of different shells were randomized to disrupt the coherent interference of artifacts. These sampling points were then adapted to a regular grid of  $50 \times 50 \times 50$  points for indirect dimensions, corresponding to 0.16% of the sampling points used conventionally. Eight FIDs were recorded for each time point to achieve the quadrature detection of



**Fig. 1.**  ${}^1H$ - ${}^{15}N$  HSQC spectrum of SKIPN. The limited dispersion of resonance signals is characteristic for intrinsically disordered proteins.



**Fig. 2.** Pulse sequences for the (A) HA(CA)CO(CA)NH and (B) HA(CA)CONH experiments. Narrow and wide bars represent  $90^\circ$  and  $180^\circ$  pulses, respectively. The open bars represent a “3–9–19” water suppression unit [31]. All pulses are along the  $x$ -axis unless indicated otherwise.  $^1\text{H}$  decoupling is achieved by using DIPSI-2. Shaped pulse G4 [32] (with a duration of  $400\ \mu\text{s}$  and an effective inversion window of about  $110\ \text{ppm}$ ) and Q3 [33] (with a duration of  $300\ \mu\text{s}$  and an effective inversion window of about  $70\ \text{ppm}$ ) was used for  $^{13}\text{C}$   $90^\circ$  and  $180^\circ$  pulses, respectively. (A) HA(CA)CO(CA)NH: The carrier frequencies of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  are  $4.7\ \text{ppm}$ ,  $43\ \text{ppm}$  and  $117\ \text{ppm}$  respectively. The  $^{13}\text{C}$  carrier frequency is changed to  $174\ \text{ppm}$  before point  $a$  and switched back to  $43\ \text{ppm}$  after point  $b$ . The delays are  $\delta_1 = 2.4\ \text{ms}$ ,  $\delta_2 = 8\ \text{ms}$ ,  $\delta_3 = 5.5\ \text{ms}$ ,  $\tau = 2.3\ \text{ms}$ ,  $\varepsilon = 300\ \mu\text{s}$ . Semi-constant time elements are used in all three indirect dimensions, and the delays are  $t_1^a = \tau_{\text{CH}} + \frac{t_1}{2}$ ,  $t_1^b = \frac{t_1}{2}(1 - \kappa_1)$ ,  $t_1^c = \tau_{\text{CH}} - \kappa_1 * \frac{t_1}{2}$ , where  $\tau_{\text{CH}} = 1.7\ \text{ms}$ ,  $\kappa_1 = \min(1, \frac{2 * \tau_{\text{CH}} * \text{SW}_1}{N_1})$ , and  $\text{SW}_1$  and  $N_1$  are the spectral width and the largest grid point of  $^1\text{H}^z$  dimension, respectively;  $t_2^a = T_C + \frac{t_2}{4}(2 - \kappa_2)$ ,  $t_2^b = T_C - \frac{t_2}{4} * \kappa_2$ , where  $T_C = 2.75\ \text{ms}$ ,  $\kappa_2 = \min(1, \frac{4 * T_C * \text{SW}_2}{N_2})$ , and  $\text{SW}_2$  and  $N_2$  are spectral width and the largest grid point of  $^{13}\text{CO}$  dimension, respectively;  $t_3^a = T_N - \kappa_3 * \frac{t_3}{2}$ ,  $t_3^b = (1 - \kappa_3) * \frac{t_3}{2}$ ,  $t_3^c = \kappa_3 * \frac{t_3}{2}$ ,  $t_3^d = T_N + (1 - \kappa_3) * \frac{t_3}{2}$ , where  $T_N = 12.4\ \text{ms}$ ,  $\kappa_3 = \min(1, \frac{2 * T_N * \text{SW}_3}{N_3})$ , and  $\text{SW}_3$  and  $N_3$  are the spectral width and the largest grid point of  $^{15}\text{N}$  dimension. The delay during which  $^{13}\text{C}$  magnetization remain transverse was set to  $\sim 1/J_{\text{CC}\beta}$  ( $28\ \text{ms}$ ). A  $45^\circ$  phase shift  $\psi$ , which was determined experimentally, is used to compensate the Bloch–Siegert effect. The phase cycling is  $\Phi_1 = x$ ;  $\Phi_2 = x, -x$ ;  $\Phi_3 = x, x, -x, -x$ ;  $\Phi_r = x, -x, -x, x$ . Quadrature detections in  $^1\text{H}^z$ ,  $^{13}\text{CO}$  and  $^{15}\text{N}$  dimensions are achieved using the States-TPPI method by changing  $\Phi_1$ ,  $\Phi_2$  and  $\Phi_3$  respectively. (B) HA(CA)CONH: The carrier frequencies of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  are  $4.7\ \text{ppm}$ ,  $43\ \text{ppm}$  and  $117\ \text{ppm}$  respectively. The  $^{13}\text{C}$  carrier frequency is changed to  $174\ \text{ppm}$  before point  $a$ . The arrow-marked pulses are used to compensate the Bloch–Siegert effect. The  $\Delta$ -marked pulse represents an off-resonance Q3 [33]  $\text{C}^z$  selective pulse centered at  $56\ \text{ppm}$  with an effective inversion window of  $30\ \text{ppm}$ . The delays are  $\delta_1 = 2.4\ \text{ms}$ ,  $\delta_3 = 5.5\ \text{ms}$ ,  $\delta_4 = 4\ \text{ms}$ ,  $\tau = 2.3\ \text{ms}$ ,  $\varepsilon = 300\ \mu\text{s}$ . Semi-constant time element is used in all three indirect dimensions, and the delays are  $t_1^a = \tau_{\text{CH}} + \frac{t_1}{2}$ ,  $t_1^b = \frac{t_1}{2}(1 - \kappa_1)$ ,  $t_1^c = \tau_{\text{CH}} - \kappa_1 * \frac{t_1}{2}$ , where  $\tau_{\text{CH}} = 1.7\ \text{ms}$ ,  $\kappa_1 = \min(1, \frac{2 * \tau_{\text{CH}} * \text{SW}_1}{N_1})$ , and  $\text{SW}_1$  and  $N_1$  are the spectral width and the largest grid point of  $^1\text{H}^z$  dimension, respectively;  $t_2^a = \tau_{\text{CC}} + \frac{t_2}{2}$ ,  $t_2^b = \tau_{\text{CN}} - \tau_{\text{CC}}$ ,  $t_2^c = \frac{t_2}{2}(1 - \kappa_2)$ ,  $t_2^d = \tau_{\text{CN}} - \kappa_2 * \frac{t_2}{2}$ , where  $\tau_{\text{CC}} = 4\ \text{ms}$ ,  $\tau_{\text{CN}} = 12\ \text{ms}$ ,  $\kappa_2 = \min(1, \frac{2 * \tau_{\text{CN}} * \text{SW}_2}{N_2})$ , and  $\text{SW}_2$  and  $N_2$  are the spectral width and the largest grid point of  $^{13}\text{CO}$  dimension, respectively;  $t_3^a = T_N - \kappa_3 * \frac{t_3}{2}$ ,  $t_3^b = (1 - \kappa_3) * \frac{t_3}{2}$ ,  $t_3^c = \kappa_3 * \frac{t_3}{2}$ ,  $t_3^d = T_N + (1 - \kappa_3) * \frac{t_3}{2}$ , where  $T_N = 12\ \text{ms}$ ,  $\kappa_3 = \min(1, \frac{2 * T_N * \text{SW}_3}{N_3})$ , and  $\text{SW}_3$  and  $N_3$  are the spectral width and the largest grid point of  $^{15}\text{N}$  dimension. The phase cycling is  $\Phi_1 = x$ ;  $\Phi_2 = x, -x$ ;  $\Phi_3 = x, x, -x, -x$ ;  $\Phi_r = x, -x, -x, x$ . Quadrature detections in the  $^1\text{H}^z$ ,  $^{13}\text{CO}$  and  $^{15}\text{N}$  dimensions are achieved using the States-TPPI method by changing  $\Phi_1$ ,  $\Phi_2$  and  $\Phi_3$ , respectively.

indirect dimensions. Details of experimental set up were summarized in Table 1. All of the experiments were recorded at  $298\ \text{K}$  on a Bruker AVANCE 500 MHz spectrometer equipped with a cryoprobe. The 4-D spectra were processed with the FFT-CLEAN algorithm [9]. A gain of 30% was used for the CLEAN processing. The final matrix sizes are  $100 \times 100 \times 100 \times 128$  points for all spectra, with 128 points for extracted amide signals of  $7.7$ – $8.8\ \text{ppm}$  and 100 points for all indirect dimensions.

### 3. Results and discussion

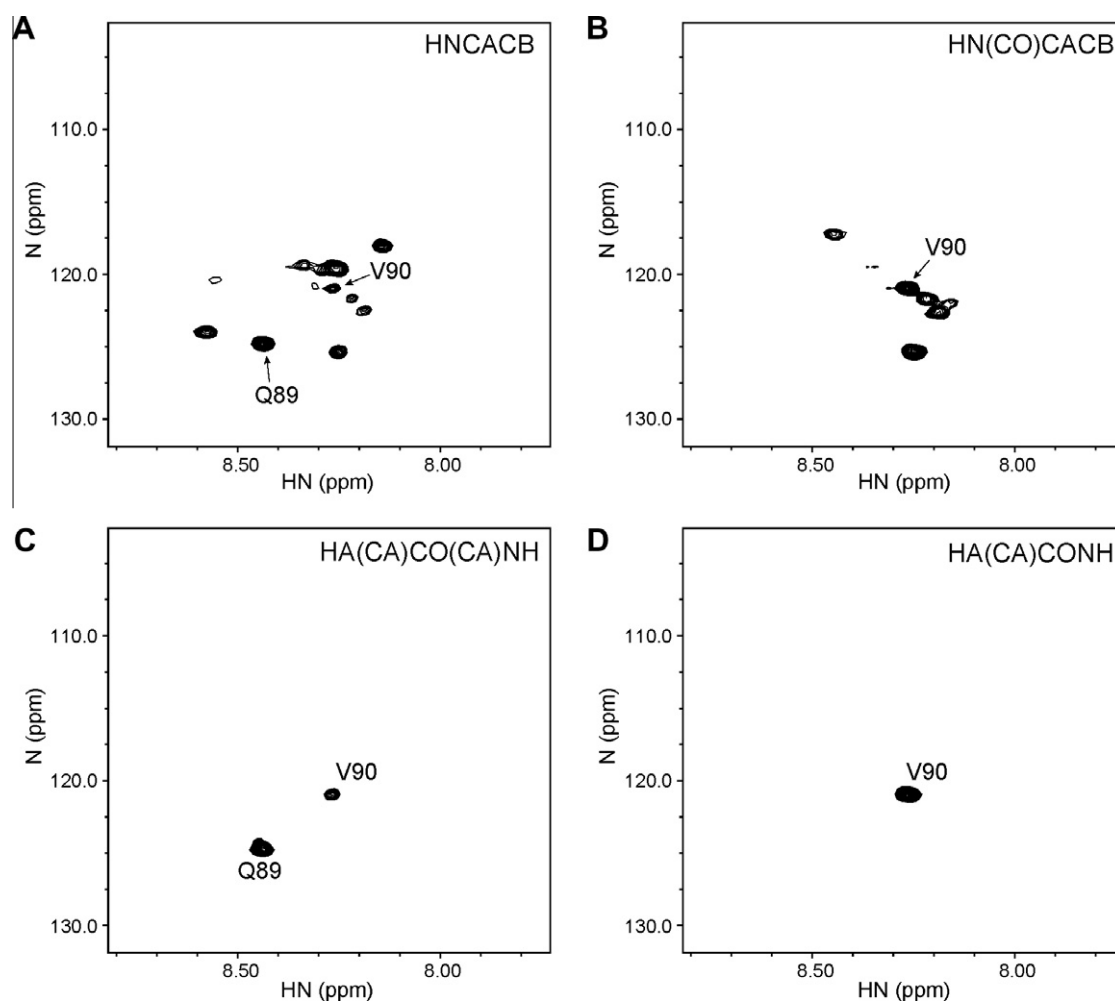
As reported previously, the  $\text{C}^\alpha$ ,  $\text{H}^\alpha$  and  $\text{C}^\beta$  chemical shift dispersion of IDPs is much poorer than the CO chemical shifts. As a result, although  $\text{C}^\alpha$  and  $\text{C}^\beta$  information is informative on the residue types, such information does not provide sufficient separation to allow for a unique connectivity. In contrast, the HA(CA)CO(CA)NH and HA(CA)CONH experiments, which correlate the  $\text{H}^\alpha$ , CO and the  $\text{H}^N$  and N chemical shifts, often provide distinct sequential connectivity information with much reduced peak overlap. This is well illustrated in Fig. 3: when the  $\text{H}^N$ –N plane was plotted at the selected  $\text{C}^\alpha$ – $\text{C}^\beta$  frequencies of Q89, one can readily see many  $\text{H}^N$ –N resonances, all sharing similar  $\text{C}^\alpha$ – $\text{C}^\beta$  chemical shifts; in contrast,

only a single pair of  $\text{H}^N$ –N signals were observed in the  $\text{H}^N$ –N plane plotted at the corresponding  $\text{H}^\alpha$ /CO frequencies of Q89 in the HA(CA)CO(CA)NH spectrum and a single peak in the HA(CA)CONH spectrum (Fig. 3). The significantly reduced overlap in the HA(CA)CO(CA)NH/HA(CA)CONH spectra greatly facilitates the process of sequential assignment for IDPs. Indeed, with a combined use of the 4-D HNCACB/HN(CO)CACB and HA(CA)CO(CA)NH/HA(CA)CONH experiments, we were able to assign 93% of the backbone resonances, obtaining nearly complete sequential assignments (Fig. 4).

Recently, a number of fast NMR methods have been applied to resonance assignment of IDPs [3–8]. Most impressively, Narayanan et al. [8] demonstrated automated resonance assignment of the 441-residue, intrinsically disordered protein Tau using 5-D and 7-D experiments based on the APSY approach [5] with 5 days of measurement time. Because the signals must be observed on every projection planes, there is a lack of signal accumulation within the overall measurement time, and the sensitivity of APSY is limited to the sensitivity of individual projections. Discrete Fourier transform of sparsely sampled time domain data has been shown to allow for signal accumulation over the entire measurement period [24–26], and Pannetier and co-workers first demonstrated its benefit in

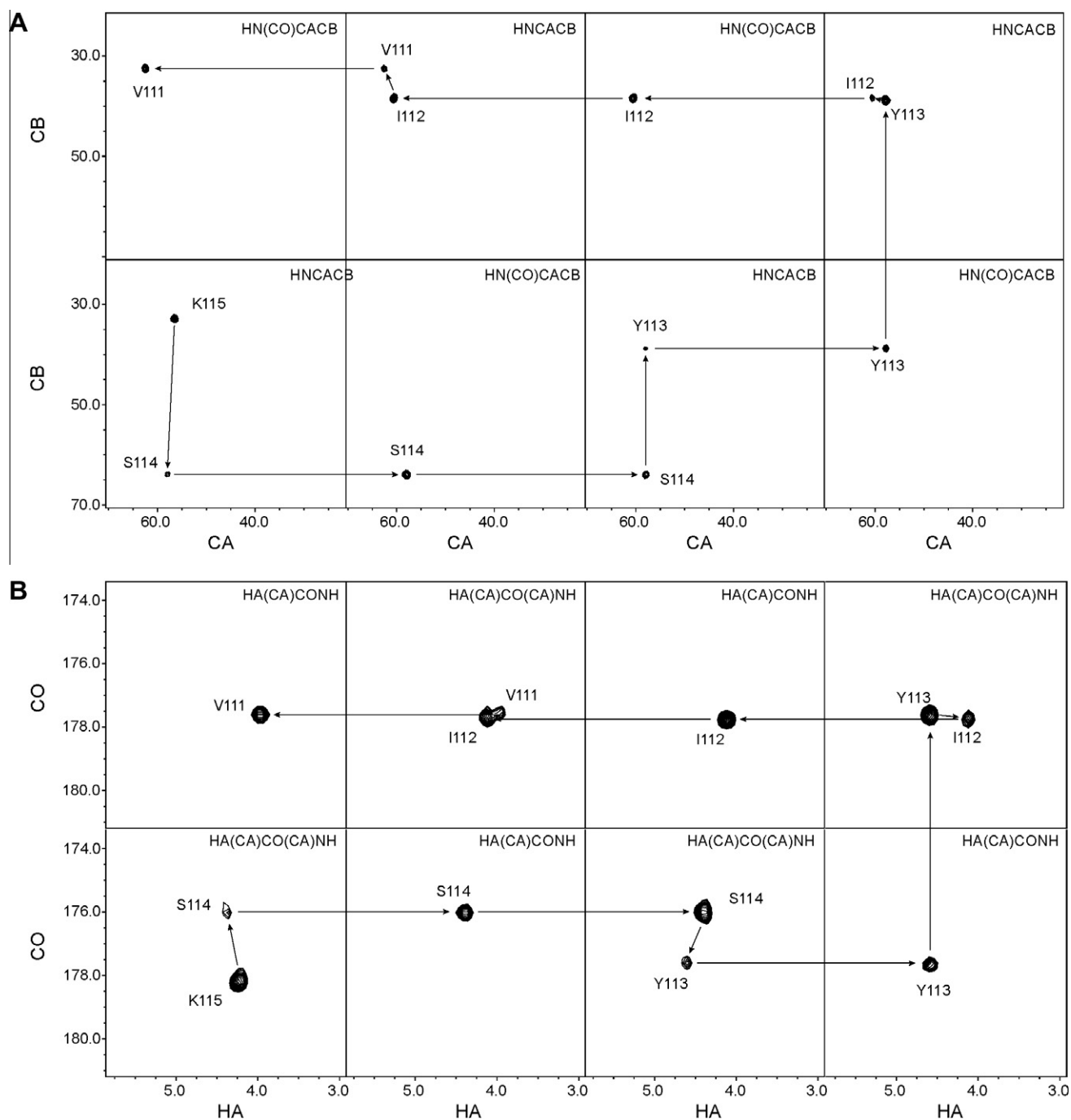
**Table 1**  
Details of NMR experiments.

Experiments	Dimension	Spectral width (ppm)	$t_{\max}$ (ms)	Sampling points (indirect dimensions)	Number of scans per FID	Relaxation delay (s)	Experimental time (h)
HNCACB	$H^N$	11.7	87.6	201	4	1.2	2.64
	N	29.6	33.3				
	$C^\alpha$	49.7	8				
	$C^\beta$	49.7	8				
HN(CO)CACB	$H^N$	11.7	87.6	201	4	1.2	2.64
	N	29.6	33.3				
	$C^\alpha$	49.7	8				
	$C^\beta$	49.7	8				
HA(CA)CO(CA)NH	$H^N$	11.7	87.6	201	4	1.2	2.64
	N	29.6	33.3				
	CO	15.9	25				
	$H^z$	6	16.7				
HA(CA)CONH	$H^N$	11.7	87.6	201	4	1.2	2.64
	N	29.6	33.3				
	CO	15.9	25				
	$H^z$	6	16.7				

**Fig. 3.**  $H^N$ - $N$  planes from the reconstructed high-resolution 4-D triple-resonance spectra. 2D  $H^N$ - $N$  planes are selected at  $C^\alpha = 55.56$  ppm,  $C^\beta = 29.67$  ppm (residue Q89) from the HNCACB (A) and HN(CO)CACB (B) spectra, and at CO = 177.5 ppm,  $H^z = 4.4$  ppm (residue Q89) from the HA(CA)CO(CA)NH (C) and HA(CA)CONH (D) spectra.

studies of partially unfolded proteins using 3-D NMR [3]. Very recently, Motáčková and co-workers presented a strategy for complete resonance assignment of disordered proteins based on the discrete Fourier transform of sparsely sampled 5-D data [7]. It should be noted that discrete Fourier transform of sparsely sampling time domain inevitably introduces aliasing artifacts due to the violation of the Nyquist sampling theorem. Although the alias-

ing artifacts can appear as pseudo-noise using optimized sampling schemes, they are often far greater than the thermal noise. Because the spectral noise is the sum of aliasing artifacts and the thermal noise, the presence of aliasing artifacts can significantly reduce the dynamic range of detectable signals and result in poor sensitivity. The signal-to-aliasing noise ratio of a single, non-decaying signal is proportional to the square root of the number of sampling

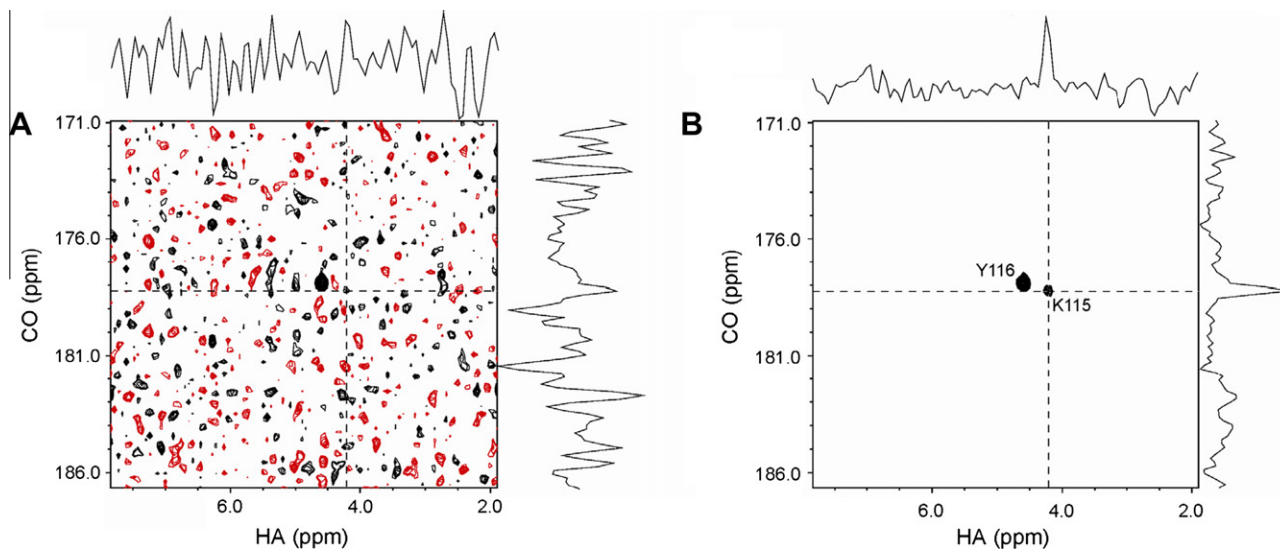


**Fig. 4.** Sequential assignment using sparsely sampled 4-D HNCACB/HN(CO)CACB and HA(CA)CO(CA)NH/HA(CA)CONH experiments. Sequential  $C^\alpha$ - $C^\beta$  and  $H^\alpha$ -CO connectivities of residue V111-K115 are shown in panels (A) and (B), respectively.

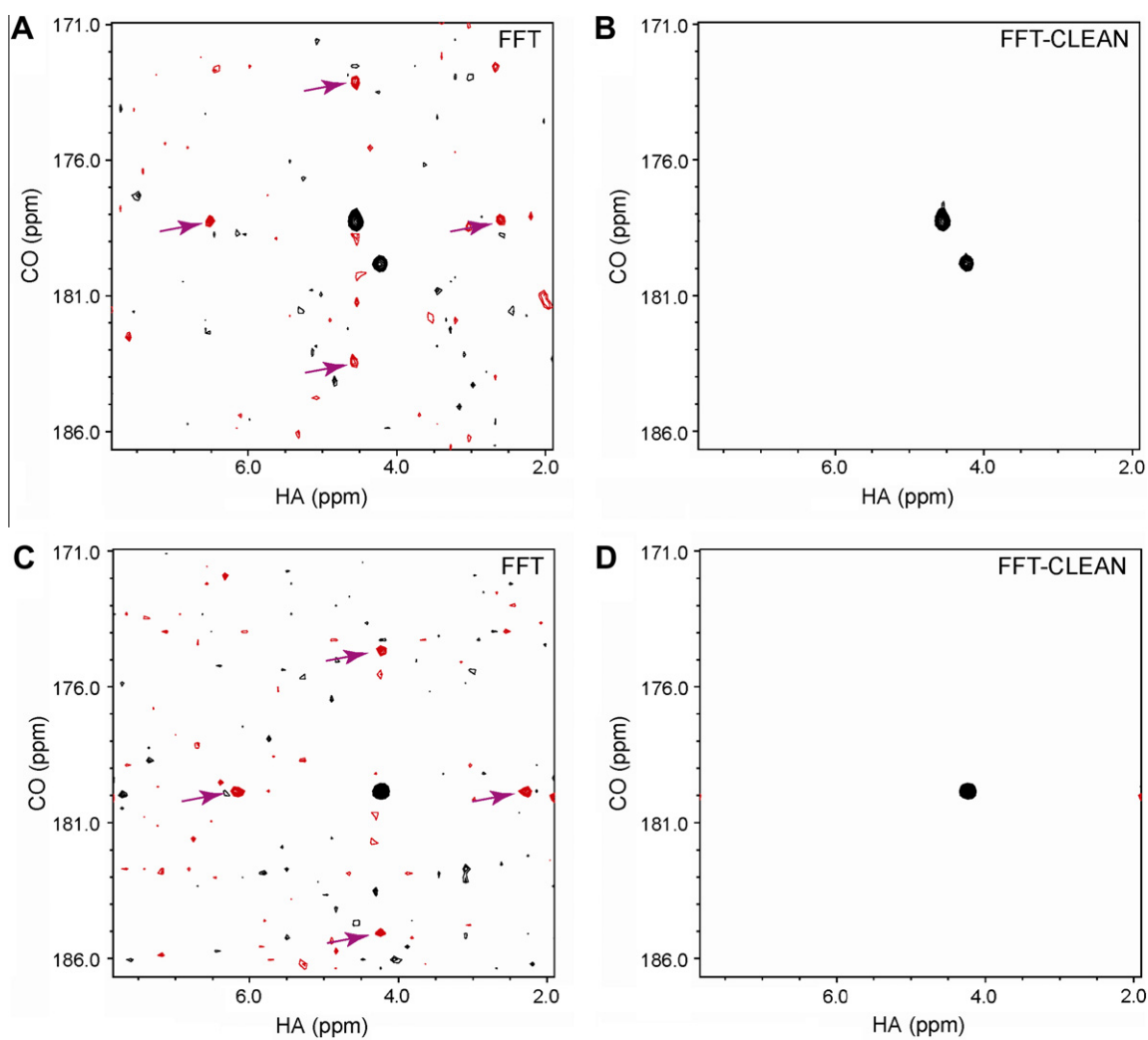
points ( $S/N_{\text{alias}} \propto \sqrt{n_{\text{sampling}}}$ ) [27]. It can be shown that the presence of multiple signals of similar magnitudes further decreases the dynamic range ( $S/N_{\text{alias}} \propto \sqrt{n_{\text{sampling}}/m}$ , where  $m$  is the number of signals present for the reconstruction). Consequently, data collection for these 5-D experiments typically took 740–2260 complex points in the indirect dimensions, occupying 20–62 h of measurement time for each experiment to allow for sufficient separation of signals from aliasing artifacts. The implementation of the iterative CLEAN procedure however systematically removes the aliasing artifacts associated with the underlying sparse sampling scheme, reduces the background “noise” to the level of true thermal noise, and thus greatly enhances the sensitivity of the experiments [9]. Indeed, using sparse sampling with FFT-CLEAN processing, we

were able to collect high-resolution 4-D spectra of IDPs using only  $\sim 200$  sampling points in the indirect dimensions, with each 4-D experiment taking 2.64 h and the entire set of 4-D data less than 11 h. Typically, the CLEAN processing reduces the background noise by 50% on average, and as much as 82% reduction in background noise has been observed for individual cubes of the 4-D spectra. A representative plane from the 4-D HA(CA)CO(CA)NH experiment before and after CLEAN is shown in Fig. 5, demonstrating the ability of CLEAN to dramatically reduce the aliasing artifacts and identify weak signals that would otherwise be concealed under the aliasing noise.

Several types of sampling patterns have been discussed in the literature with an emphasis on reducing aliasing artifacts



**Fig. 5.** The CLEAN algorithm significantly reduces the aliasing artifacts and allows for identification of weak signals. (A) A representative  $H^2$ -CO plane from the 4-D HA(CA)CO(CA)NH spectrum before CLEAN processing. (B) The same  $H^2$ -CO plane after CLEAN processing.



**Fig. 6.** The CLEAN algorithm minimizes the effect of imperfect sampling schemes. Selected 2D  $H^2$ -CO planes of residue D128 from sparsely sampled 4-D HA(CA)CO(CA)NH and HA(CA)CONH experiment without (left) and with (right) the CLEAN algorithm. The less-than-ideal sampling pattern generated stronger-than-average, negative aliasing artifacts in the FFT spectra (panels A and C, indicated by arrows), which were systematically eliminated by CLEAN (panels B and D).

associated with the discrete Fourier transform [3,9,28–30]. Because all of these optimized sampling patterns employ some degree of randomization in order to disrupt coherent interference of aliasing artifacts, a direct comparison of these sampling patterns is difficult. However, when an iterative procedure such as FFT-CLEAN is employed for eliminating the aliasing artifacts, the effect of the sampling pattern is much diminished. This is due to the fact that the aliasing artifacts are inherently connected with the underlying sampling pattern. Once the *bona fide* signals are identified, their associating aliasing artifacts, no matter how they appear in the spectrum, are systematically removed by the CLEAN procedure, thus greatly reducing the dependence of the reconstruction spectral quality on the sampling scheme. This result is well illustrated in Fig. 6. In our initial studies, we accidentally employed a 1618-point sampling scheme that was far from being optimal. The point spreading function of this sampling scheme contains strong artifacts – albeit still much lower than the real signals – that are prominently present along the axes around the real peak in the FFT spectrum (Fig. 6A,C, arrows). These strong artifacts however have no consequence on the CLEAN algorithm, which systematically identifies all of the components of real signals, eliminates their associating aliasing artifacts, and improves the signal-to-noise ratio in the final reconstructed spectrum (Fig. 6B and D).

#### 4. Conclusion

A lack of a stable conformation of intrinsically disordered proteins leads to severe signal overlaps in 3-D triple-resonance spectra. Here, we present a pair of 4-D high-resolution experiments tailored for backbone resonance assignment of intrinsically disordered proteins. By selecting alternative sampling schemes involving  $H^z$ , CO, N and  $H^N$  nuclei based on the (HACA)CO(CA)NH and HACA(CO)NH experiments, we created new types of 4-D triple-resonance experiments – HA(CA)CO(CA)NH/HA(CA)CONH – that complement the 4-D HNCACB/HN(CO)CACB experiments to enable efficient and non-redundant collection of a complete set of backbone resonances of IDPs. Additionally, we show that iterative CLEAN algorithm not only improves the spectral sensitivity by removing the aliasing artifacts, but also achieves such results even with less-than-optimal sampling schemes. The experiments presented here do not require special hardware and can be widely used to study inherently unstructured proteins.

#### Acknowledgments

This work was supported by the National Institutes of Health (GM079376) to P. Zhou. J. Wu thanks the financial support from the Chinese National Fundamental Research Project (Grants 2011CB966302 and 2006AA02A315). The authors thank Drs. Jingjun Hong and Xingsheng Wang for assistance in sample preparation, and Dr. Ronald A. Venters (Duke University NMR Center) for critical reading of the manuscript.

#### References

- [1] A.L. Fink, Natively unfolded proteins, *Curr. Opin. Struct. Biol.* 15 (2005) 35–41.
- [2] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, *Annu. Rev. Biophys.* 37 (2008) 215–246.
- [3] N. Pannetier, K. Houben, L. Blanchard, D. Marion, Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins, *J. Magn. Reson.* 186 (2007) 142–149.
- [4] V.A. Jaravine, A.V. Zhuravleva, P. Permi, I. Ibragimov, V.Y. Orekhov, Hyperdimensional NMR spectroscopy with nonlinear sampling, *J. Am. Chem. Soc.* 130 (2008) 3927–3936.
- [5] S. Hiller, F. Fiorito, K. Wüthrich, G. Wider, Automated projection spectroscopy (APSY), *Proc. Natl. Acad. Sci. USA* 102 (2005) 10876–10881.
- [6] S. Hiller, C. Wasmer, G. Wider, K. Wüthrich, Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy, *J. Am. Chem. Soc.* 129 (2007) 10823–10828.
- [7] V. Motáčková, J. Nováček, A. Zawadzka-Kazimierczuk, K. Kazimierczuk, L. Židek, H. Šanderová, L. Krásný, W. Koźmiński, V. Sklenář, Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments, *J. Biomol. NMR* 48 (2010) 169–177.
- [8] R.L. Narayanan, U.H. Durr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, Automatic assignment of the intrinsically disordered protein Tau with 441-residues, *J. Am. Chem. Soc.* 132 (2010) 11906–11907.
- [9] B.E. Coggins, P. Zhou, High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN, *J. Biomol. NMR* 42 (2008) 225–239.
- [10] B.E. Coggins, R.A. Venters, P. Zhou, Filtered backprojection for the reconstruction of a high-resolution (4,2)D CH<sub>3</sub>-NH NOESY spectrum on a 29 kDa protein, *J. Am. Chem. Soc.* 127 (2005) 11562–11563.
- [11] J.W. Werner-Allen, B.E. Coggins, P. Zhou, Fast acquisition of high resolution 4-D amide-amide NOESY with diagonal suppression, sparse sampling and FFT-CLEAN, *J. Magn. Reson.* 204 (2010) 173–178.
- [12] D. Marion, Combining methods for speeding up multi-dimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins, *J. Magn. Reson.* 206 (2010) 81–87.
- [13] A. Zawadzka-Kazimierczuk, K. Kazimierczuk, W. Kozminski, A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins, *J. Magn. Reson.* 202 (2010) 109–116.
- [14] M. Mobli, A.S. Stern, W. Bermel, G.F. King, J.C. Hoch, A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment, *J. Magn. Reson.* 204 (2010) 160–164.
- [15] V. Tugarinov, L.E. Kay, I. Ibragimov, V.Y. Orekhov, High-resolution four-dimensional  $^1H$ - $^{13}C$  NOE spectroscopy using methyl-TROSY sparse data acquisition and multidimensional decomposition, *J. Am. Chem. Soc.* 127 (2005) 2767–2775.
- [16] X. Wang, S. Zhang, J. Zhang, X. Huang, C. Xu, W. Wang, Z. Liu, J. Wu, Y. Shi, A large intrinsically disordered region in SKIP and its disorder-order transition induced by PP1L1 binding revealed by NMR, *J. Biol. Chem.* 285 (2010) 4951–4963.
- [17] M. Wittekind, L. Mueller, HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins, *J. Magn. Reson. Ser. B* 101 (1993) 201–205.
- [18] R.A. Venters, B.E. Coggins, D. Kojetin, J. Cavanagh, P. Zhou, (4, 2)D Projection – reconstruction experiments for protein backbone assignment: application to human carbonic anhydrase II and calbindin D (28K), *J. Am. Chem. Soc.* 127 (2005) 8785–8795.
- [19] F. Löhr, H. Rüterjans, A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins, *J. Biomol. NMR* 6 (1995) 189–197.
- [20] W. Boucher, E.D. Laue, S. Campbell-Burk, P.J. Domaille, Four-dimensional heteronuclear triple resonance NMR methods for the assignment of backbone nuclei in proteins, *J. Am. Chem. Soc.* 114 (1992) 2262–2264.
- [21] H. Matsuo, E. Kupče, H. Li, G. Wagner, Use of selective C alpha pulses for improvement of HN(CA)CO-D and HN(COCA)NH-D experiments, *J. Magn. Reson. B* 111 (1996) 194–198.
- [22] D. Yang, R.A. Venters, G.A. Mueller, W.Y. Choy, L.E. Kay, TROSY-based HNCOC pulse sequences for the measurement of  $^1HN$ - $^{15}N$ ,  $^{15}N$ - $^{13}CO$ ,  $^1HN$ - $^{13}CO$ ,  $^{13}CO$ - $^{13}C^z$  and  $^1HN$ - $^{13}C^z$  dipolar couplings in  $^{15}N$ ,  $^{13}C$ ,  $^2H$ -labeled proteins, *J. Biomol. NMR* 14 (1999) 333–343.
- [23] J.W. Werner-Allen, L. Jiang, P. Zhou, A 'just-in-time' HN(CA)CO experiment for the backbone assignment of large proteins with high sensitivity, *J. Magn. Reson.* 181 (2006) 177–180.
- [24] B.E. Coggins, P. Zhou, Polar Fourier transforms of radially sampled NMR data, *J. Magn. Reson.* 182 (2006) 84–95.
- [25] K. Kazimierczuk, W. Koźmiński, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, *J. Magn. Reson.* 179 (2006) 323–328.
- [26] D. Marion, Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of a reconstruction, *J. Biomol. NMR* 36 (2006) 45–54.
- [27] K. Kazimierczuk, A. Zawadzka, W. Kozminski, Narrow peaks and high dimensionalities: exploiting the advantages of random sampling, *J. Magn. Reson.* 197 (2009) 219–228.
- [28] B.E. Coggins, P. Zhou, Sampling of the NMR time domain along concentric rings, *J. Magn. Reson.* 184 (2007) 219–233.
- [29] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, Optimization of random time domain sampling in multidimensional NMR, *J. Magn. Reson.* 192 (2008) 123–130.
- [30] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, *J. Biomol. NMR* 36 (2006) 157–168.
- [31] V. Sklenář, M. Piotto, R. Leppik, V. Saudek, Gradient-Tailored water suppression for  $^1H$ - $^{15}N$  HSQC experiments optimized to retain full sensitivity, *J. Magn. Reson. A* 102 (1993) 241–245.
- [32] L. Emsley, G. Bodenhausen, Gaussian pulse cascades: new analytical functions for rectangular selective inversion and in-phase excitation in NMR, *Chem. Phys. Lett.* 165 (1990) 469–476.
- [33] L. Emsley, G. Bodenhausen, Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators, *J. Magn. Reson.* 97 (1992) 135–148.