Journal of Magnetic Resonance 202 (2010) 109–116

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins

Anna Zawadzka-Kazimierczuk, Krzysztof Kazimierczuk, Wiktor Koźmiński *

Faculty of Chemistry, University of Warsaw, Pasteura 1, PL-02-093 Warsaw, Poland

ARTICLE INFO

Article history:
Received 15 September 2009
Revised 13 October 2009
Available online 31 October 2009

Keywords:
CsPin
Maltose binding protein
Proteins
NMR
Sparse multidimensional Fourier transform
Resonance assignment
Random sampling

ABSTRACT

This paper presents examples of techniques based on the principle of random sampling that allows acquisition of NMR spectra featuring extraordinary resolution. This is due to increased dimensionality and maximum evolution time reached. The acquired spectra of CsPin protein and maltose binding protein were analyzed statistically with the aim to evaluate each technique. The results presented include exemplary spectral cross-sections. The spectral data provided by the proposed techniques allow easy assignment of backbone and side-chain resonances.

1. Introduction

Nuclear Magnetic Resonance (NMR) is a very powerful technique that allows examination of molecular structure and dynamics at atomic resolution. One of the main objects of such studies are proteins. Relationship between protein structure and function, although complex and not straightforward, makes structural research an important stage on the way to explanation of biological mechanisms.

In general, protein structural research employing NMR consists of two steps: signal assignment and structure determination [1,2]. In the first step, a set of various multidimensional spectra is examined with the aim to assign each resonance (i.e. spectral peak) to a particular spin system in the molecule. Nuclei within a spin system are correlated by the coherence transfer employed in a given experiment. Peak positions in a multidimensional frequency space correspond to the chemical shifts of nuclei coupled by the mutual interaction used to trigger the coherence transfer. This relationship allows assigning the chemical shift values to the nuclei. In the next step, molecule structure is determined by collecting the information on structure-dependent spectral parameters, followed by structure modeling. Some information on secondary structure is derived from chemical shifts of H_α, C_α, C_β and CO nuclei [3]. Tertiary structure can be determined using molecular dynamics with constraints obtained from NMR spectra; intensity and multiplicity of assigned peaks contain information about structural parameters, e.g. distances between nuclei and torsion angles of the peptide chain. Distances up to 5 Å can be determined using intensities of correlation signals in NOESY experiment. Torsion angles can be estimated from J-coupling values, with the use of Karplus relationship [4]. Global structural information, e.g. about relative orientation of two domains, may be obtained from residual dipolar couplings [5]. Accuracy and precision of peak frequency measurements is crucial for both signal assignment and determination of all the above parameters. Therefore, acquisition of spectra with high signal-to-noise ratio and well-resolved peaks is of utmost importance.

In this study, we have addressed the issue of signal assignment by optimizing parameters of the spectra used. Usually, the assignment of protein resonances starts with acquisition of a set of spectra that have to be analyzed simultaneously. There are many acquisition techniques, and usually the protein size indicates which of them is the most suitable. For small proteins (of molecular weight $\leqslant$ 10 kDa), homonuclear ^1H–^1H correlation experiments may be sufficient. For bigger molecules, heteronuclear techniques that require isotope labeling with ^{13}C and/or ^{15}N are needed, and very large systems may require also ^2H labeling to reduce relaxation rate of measured signal. Of the heteronuclear techniques, most widely used are three-dimensional (3D) ones, because of sampling requirements (see below) that grow with the dimensionality and severely limit the resolution in higher-dimensional spectra. Combination of spectra, which contain both intra- and inter-residual correlations, enable piecing together longer sequences of amino

* Corresponding author. Fax: +48 22 822 5996.
E-mail address: kozmin@chem.uw.edu.pl (W. Koźmiński).

1090-7807/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2009.10.006

acids and completing the assignment of the entire protein chain. There are many strategies both for manual [6] and automatic [7] signal assignment. The former approach requires usually weeks or even months of arduous work, while the latter, although much faster, usually requires manual correction of the results. Most difficulties arise from low resolution and the resulting peak overlap in spectra obtained in a conventional way.

2. Methods

2.1. Random sampling

Although increasing experiment dimensionality diminishes peak overlap, its applicability is limited in the case of conventional experiments due to practical aspects. In this kind of experiments, free induction decay signal is sampled at equal intervals (i.e. with samples placed on a Cartesian grid), which approach requires fulfilling of Nyquist theorem. The theorem stands that the sampling frequency in a particular dimension must be twice the highest frequency expected in the respective signal. Otherwise, false peaks will appear in the spectrum; this phenomenon is called aliasing or peak folding. On the other hand, according to Fourier uncertainty principle [8], the higher the maximum evolution time, the narrower are the peaks present in the spectrum, with relaxation-constrained linewidth, of course. This implies that, in multidimensional experiments based on conventional on-grid sampling, and having limited experimental time, one cannot usually obtain as narrow peaks as those determined by relaxation rate alone. In practice, conventional 4D experiments provide spectra with extremely wide lines (usually of the order of hundreds Hz), whereas conventional experiments of higher dimensionality are not feasible.

Recently [9] we have presented the idea of random sampling of evolution time space and multidimensional Fourier transform (MFT) processing. Other approaches based on random sampling schedules were also reported, which employed different processing methods: maximum entropy [10,11], multidimensional decomposition [12] or covariance spectroscopy [13]. Distributing evolution time domain points in a non-conventional way (random off-grid sampling) provides alias-free spectrum [14], even for sampling density (average sampling frequency) far below that required by Nyquist theorem. Such spectrum cannot be obtained easily from the off-grid data with sequential fast Fourier transform, as is the case for conventional distribution. Therefore, one-step MFT is to be employed, i.e. a sequence of one-dimensional Fourier integrals should be replaced by a single multidimensional integral.

Random sampling allows obtaining spectra of very high resolution in a reasonable time [15], as maximum evolution times in indirectly detected dimensions are no longer related to experimental time or spectral widths. The price to pay is the emergence of noise-like artifacts in such spectra. The peak-to-artifact ratio is proportional to the square root of the number of time domain points (similarly to peak-to-thermal noise ratio). In our most recent work [16], we have shown that increasing number of dimensions or maximum evolution times does not affect the ratio. Therefore, we find random sampling with MFT processing to be particularly suitable for NMR experiments of high dimensionality and resolution. Such techniques may help overcoming problems with spectral peaks' overlap that obstructs commonly used signal assignment strategies. Notably, the MFT approach, unlike fast Fourier transform, enables calculating spectra in arbitrary chosen frequency points, which ability is invaluable in case of spectra of high dimensionality. Below, we present and discuss a few 4D techniques that can remarkably facilitate the procedure of signal assignment due to improved resolution.

2.2. Data acquisition and processing

All techniques presented in this paper were based on well-known 3D and 4D experiments [17–21]. We expanded 3D techniques to higher dimensionality (4D) by sampling additional evolution time. The resolution was increased significantly and markedly by the use of random sampling. Being triple-resonance techniques, all our experiments required labeling samples with ^{13}C and ^{15}N. For the larger of the two proteins studied (see below), ^{2}H labeling at aliphatic positions was used as well, in which case additional ^{2}H decoupling was performed during aliphatic carbon evolution delay. Of course, ^{2}H labeling limited the choice of experiment, as it precluded coherence transfer through aliphatic protons. The use of pulsed field gradients, water flipback, sensitivity enhancement scheme and semi-constant-time evolution whenever possible brought considerable gain in signal-to-noise ratio and led to the reduction of artifacts.

High-dimensional spectra featuring narrow peaks are feasible provided one uses non-conventional sampling of evolution time space. The sampling schedule was generated using decaying (relaxation-optimized) Poisson disk distribution [22]. The number of time domain points was increased at the expense of number of transients, as this approach was optimal considering the peak-to-artifact ratio [16]. Only four transients were acquired for each time point. This circumstance limited the phase cycle to four steps, and was sufficient to reduce satisfactorily axial peak artifacts.

Importantly, the number of frequency points should be set based on the expected minimum linewidths to avoid losing the peaks located between spectral points. Assuming that three frequency points per peak would be sufficient, and considering the fact that the linewidth is never smaller than $1/t_i^{\max}$, the number of frequency points n_i in dimension i may be estimated using the following formula:

$$n_i \geqslant 3 \cdot sw_i \cdot t_i^{\max}$$

where sw_i and $t_i^{\max}$ stand, respectively, for spectral width and maximum evolution time in dimension i. Therefore, dimensionality, maximum evolution times and spectral widths determine the number of frequency points and thus the size of the resulting spectrum file. For example, 4D experiment with evolution times of tens of ms and spectral widths of a few kHz may result in spectrum size of several hundred gigabytes. Hence, it would be inconvenient, or even impossible, to store such spectra. However, peaks occupy only a small fraction of frequency space in a multidimensional spectrum, and there is no need to calculate and store "empty" parts of spectrum. We found that it is sufficient to calculate only the peak-containing 2D cross-sections of full 4D spectrum. Such sparse MFT (SMFT) procedure (described in details in [16]) was possible based on the list of peak coordinates from lower-dimensional experiment, for example 2D ^{15}N HSQC or 3D HNCO, the latter providing better peak separation. The choice of 2D cross-sections taken from higher-dimensional spectrum was based on the values of frequencies determined from the respective lower-dimensional spectrum. Similar approach has been proposed earlier for projection reconstruction using radially sampled data by Coggins et al. [23,24].

In all experiments presented below, resonance list obtained from high resolution 3D HNCO (based on random sampling) was used for SMFT processing. To obtain 2D cross-sections of 4D spectra, frequencies in amide nitrogen and amide proton dimensions were fixed (i.e. set at the resonance maxima) for SMFT processing. The number of 2D cross-sections is usually equal to the number of amino acids in the examined protein, excluding prolines.

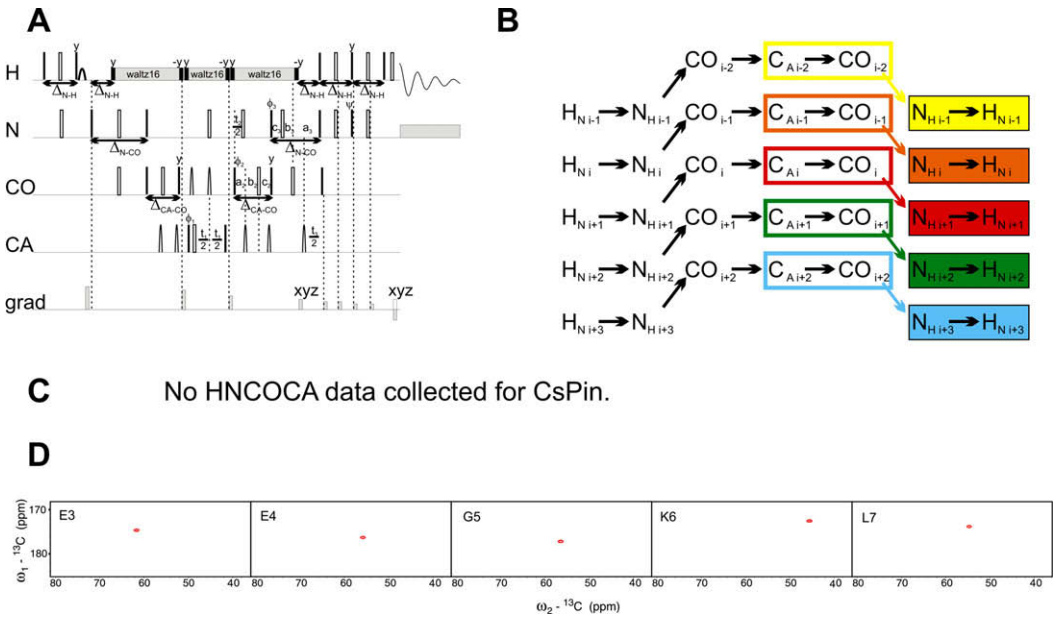

Fig. 1. HNCOCA technique: (Panel A) Pulse sequence: evolution is in the real-time mode for CA, and in semi-constant-time mode for N and CO ($a_i = (t_i + \Delta)/2$, $b_i = t_i(1 - \Delta/t_i^{max})/2$, $c_i = \Delta(1 - t_i/t_i^{max})/2$) or constant-time mode ($a_i = (\Delta t_i)/2$, $b_i = 0$, $c_i = (\Delta - t_i)/2$), where Δ stands for Δ_{N-CO} and Δ_{CA-CO}, respectively, t_i is the evolution time in ith dimension and t_i^{max} is maximal length of the evolution time delay). Delays were set as follows: $\Delta_{N-H} = 5.4$ ms, $\Delta_{N-CO} = 28$ ms, $\Delta_{CA-CO} = 6.8$ ms. (Panel B) Coherence transfer in the peptide chain. Amide nitrogen and amide proton frequencies (filled colored rectangles) are fixed for SMFT. Frame colors for CO and CA indicate the correlation. Each plane contains CO–CA peak for $i - 1$ residue. (Panel C) No HNCOCA data were collected for CsPin. (Panel D) 2D spectral planes for MBP obtained by SMFT procedure performed on the 4D HNCOCA randomly sampled signal (Poisson disk sampling) with "fixed" frequencies obtained from the respective 3D HNCO peak list. (For interpretation of color mentioned in this figure, the reader is referred to the web version of this article.)

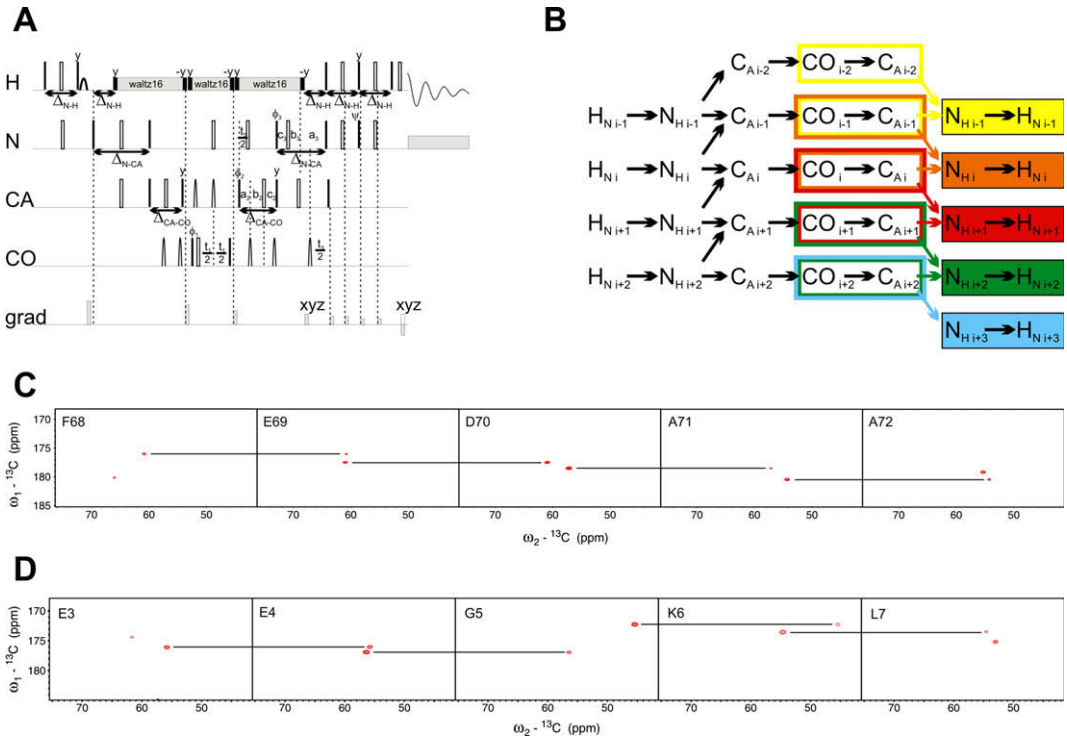

Fig. 2. 4D HNCACO technique: (Panel A) Pulse sequence, evolution for CO is in the real-time mode, and for N and CA in semi-constant-time mode ($a_i = (t_i + \Delta)/2$, $b_i = t_i(1 - \Delta/t_i^{max})/2$, $c_i = \Delta(1 - t_i/t_i^{max})/2$) or constant-time mode ($a_i = (\Delta t_i)/2$, $b_i = 0$, $c_i = (\Delta - t_i)/2$), where Δ stands for Δ_{N-CA} and Δ_{CA-CO}, respectively, t_i is the evolution time in ith dimension and t_i^{max} is the maximal length of evolution time delay. Delays were set as follows: $\Delta_{N-H} = 5.4$ ms $\Delta_{N-CA} = 22$ ms $\Delta_{CA-CO} = 6.8$ ms. (Panel B) Coherence transfer in the peptide chain. Amide nitrogen and proton frequencies (filled colored rectangles) are fixed during Fourier transformation. Frame colors for CO and CA indicate the correlation. Each plane contains CO–CA peak for i and $i - 1$ residue. (Panel C) 2D spectral planes for CsPin protein obtained by SMFT procedure performed on the 4D HNCACO randomly sampled signal (Poisson disk sampling) with "fixed" frequencies obtained from 3D HNCO peak list. (Panel D) 2D spectral planes for MBP obtained in the same manner. (For interpretation of color mentioned in this figure, the reader is referred to the web version of this article.)

2.3. Pulse sequences

HNCOCA pulse sequence (Fig. 1A) was obtained by sampling CO evolution time in 3D HN(CO)CA pulse sequence [17]. Nuclei correlated by coherence transfer were as follows: H_i–N_i–CO_{i-1}–CA_{i-1} (see the scheme in Fig. 1B). Fixing of amide proton and nitrogen frequencies at the values from 3D HNCO peak list resulted in a set of 2D CA–CO planes, each containing, as a rule, a single peak.

HNCACO pulse sequence (Fig. 2A) is commonly used with conventional on-grid sampling as 3D HN(CA)CO [18]. Inserting an additional evolution delay for CA allowed resolving spectral peaks in the fourth dimension. Although less sensitive than HNCOCA, HNCACO is more useful for signal assignment, as it yields two types of peaks: intra-residual H_i–N_i–CA_i–CO_i and inter-residual H_i–N_i–$CA_{i-1-CO}i_{-1}$, all peaks being of the same sign. Each 2D CO–CA cross-section (obtained in the same way as that used for HNCOCA) should contain two peaks, one of which is common for planes corresponding to consecutive amino acids. This "link" enables easy sequential assignment of backbone resonances. The respective coherence transfer pathway is presented in Fig. 2B.

HNCACACB pulse sequence is presented in Fig. 3A. It was based on the 3D HNCACB technique developed by Wittekind and Mueller [19]. This technique, in principle, yields four peaks per each residue (the exceptions being – of course – glycines that have no C_β), corresponding to the following correlations: H_i–N_i–CA_i–CA_i, H_i–N_i–CA_i–CB_i, H_i–N_i–CA_{i-1}–CA_{i-1} and H_i–N_i–CA_{i-1}–CB_{i-1}. Similarly to the aforementioned experiments, the amide proton and amide nitrogen frequencies should be fixed to calculate 2D cross-sections. On each CA–CAB plane, up to four peaks are expected. For any pair of adjoining residues, two of the peaks are present in both CAB planes corresponding to these residues. "Diagonal" peaks (CA–CA) are of the opposite sign than the correlation peaks (CA–CB). The respective coherence transfer pathway is shown in Fig. 3B.

The HBHACBCANH pulse sequence we present here is a random sampling version of the experiment developed by Grzesiek and Bax [20]. This pulse sequence and its respective coherence transfer scheme are shown in Fig. 4. Observed correlations include the following: H_i–N_i–CA_i–HA_i, H_i–N_i–CB_i–HB_i, H_i–N_i–CA_{i-1}–HA_{i-1} and H_i–N_i–$CB_{i-1-HB}i_{-1}$. Since there may be two H_β nuclei featuring different chemical shifts, up to six peaks may be observed on the 2D HAB–CAB plane obtained by amide proton and amide nitrogen frequency fixing.

In HN(CA)NH pulse sequence (developed originally by Weisemann et al. [21]) (see Fig. 5A), C_α nuclei take part in coherence transfer, but the CA delay is constant. The HN(CA)NH spectrum yields three types of cross-peaks: H_{i-1}–N_{i-1}–N_i–H_i, H_i–N_i–N_i–H_i,

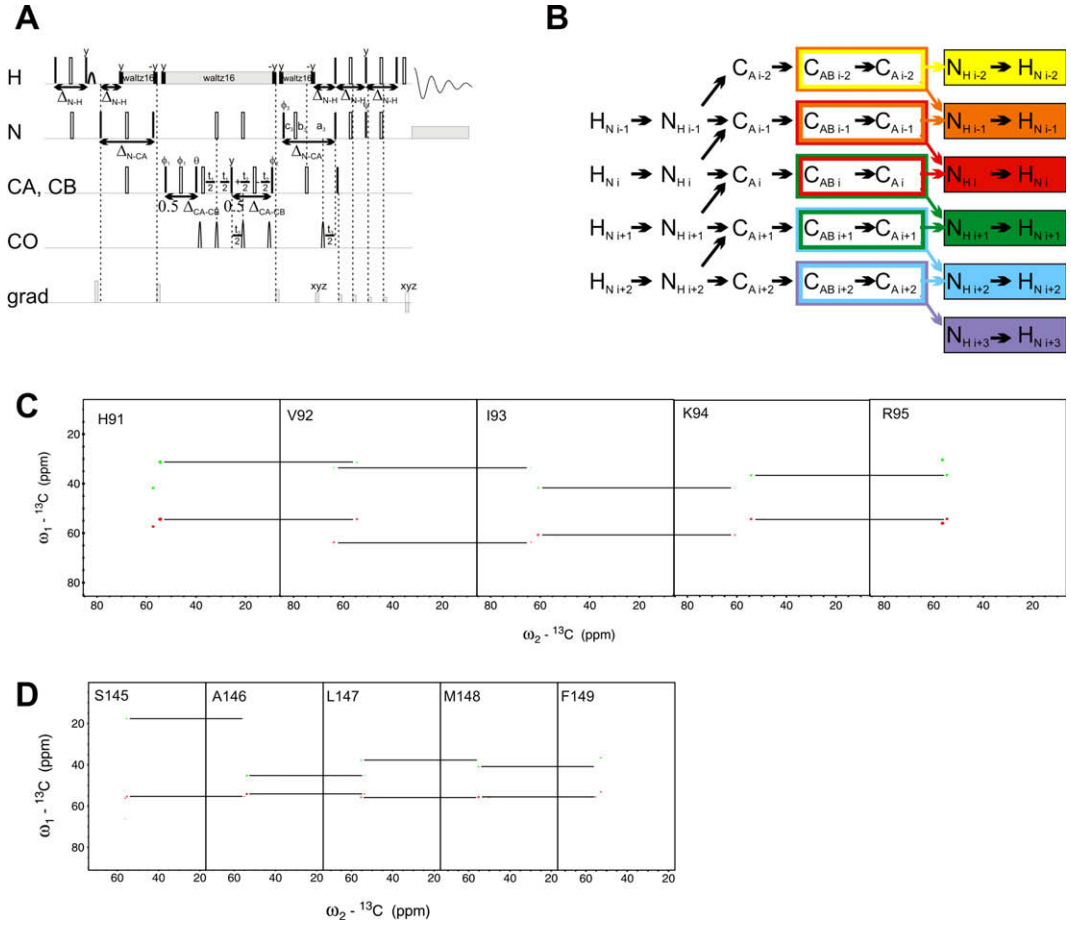

Fig. 3. 4D HNCACACB technique: (Panel A) Pulse sequence, evolution for CAB is in the real-time mode, nitrogen evolution is in semi-constant-time mode ($a_3 = (t_3 + \Delta_{N-CA})/2$, $b_3 = t_3(1 - \Delta_{N-CA}/t_3^{max})/2$, $c_3 = \Delta_{N-CA}(1 - t_3/t_3^{max})/2$) or constant-time mode ($a_3 = (\Delta_{N-CA}t_3)/2$, $b_3 = 0$, $c_3 = (\Delta_{N-CA} - t_3)/2$), and CA evolution is in constant-time mode ($\pm\frac{t_2}{2}$ means $\frac{0.5\cdot\Delta_{CA-CB}\pm t_2}{2}$), where t_i is the evolution time in ith dimension and t_i^{max} is maximal length of the evolution time delay. Delays were set as follows: $\Delta_{N-H} = 5.4$ ms, $\Delta_{N-CA} = 22$ ms, $\Delta_{CA-CB} = 14.3$ ms. (Panel B) Coherence transfer in the peptide chain. Amide nitrogen and proton frequencies (filled colored rectangles) are fixed for Fourier transform. Frame colors for CAB and CA indicate the correlation. Each plane contains CA–CA and CA–CB cross-peaks for i and $i-1$ residues. (Panel C) 2D spectral planes for CsPin protein obtained by SMFT procedure performed on the 4D HNCACACB randomly sampled signal (Poisson disk sampling) with "fixed" frequencies obtained from 3D HNCO peak list. (Panel D) 2D spectral planes for MBP obtained in the same manner. (For interpretation of color mentioned in this figure, the reader is referred to the web version of this article.)

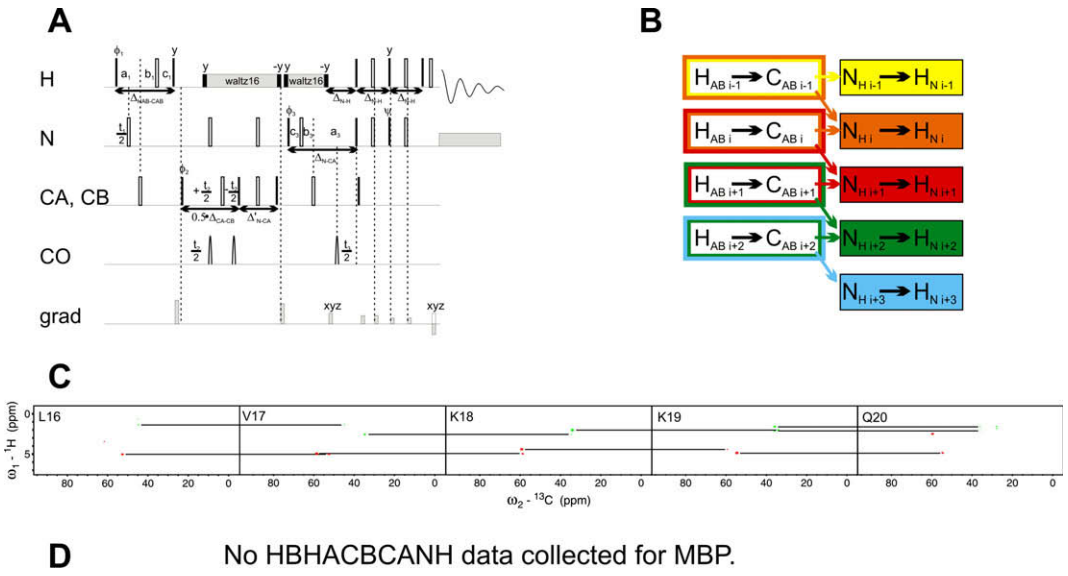

Fig. 4. 4D HBHACBCANH technique. Panel A: Pulse sequence, proton HAB and amide nitrogen evolution is in semi-constant-time mode: $a_i = (t_i + \Delta)/2$, $b_i = t_i(1 - \Delta/t_i^{max})/2$, $c_i = \Delta(1 - t_i/t_i^{max})/2$ (where Δ stands for $\Delta_{HAB-CAB}$ and Δ_{N-CA}, respectively, t_i is the evolution time in ith dimension and t_i^{max} is maximal length of the evolution time delay), carbon CAB evolution delays are in constant-time mode ($\pm\frac{t_2}{2}$ means $\frac{0.5\cdot\Delta_{CA-CB}\pm t_2}{2}$)). Delays were set as follows: Δ_{N-H} = 5.4 ms, Δ'_{N-CA} = 11 ms, Δ_{CA-CB} = 14.3 ms, $\Delta_{HAB-CAB}$ = 3.6 ms, Δ_{N-CA} = 22 ms. (Panel B) Coherence transfer in the peptide chain. Amide nitrogen and proton frequencies (filled colored rectangles) are fixed for Fourier transform. Frame colors for HAB and CAB indicate the correlation. Each plane contains HA–CA and HB–CB cross-peaks for i and $i - 1$ residue. (Panel C) 2D spectral planes for CsPin protein obtained by SMFT procedure performed on the 4D HBHACBCANH randomly sampled signal (Poisson disk sampling) with "fixed" frequencies obtained from 3D HNCO peak list. (Panel D) No HBHACBCANH data was collected for MBP. (For interpretation of color mentioned in this figure, the reader is referred to the web version of this article.)

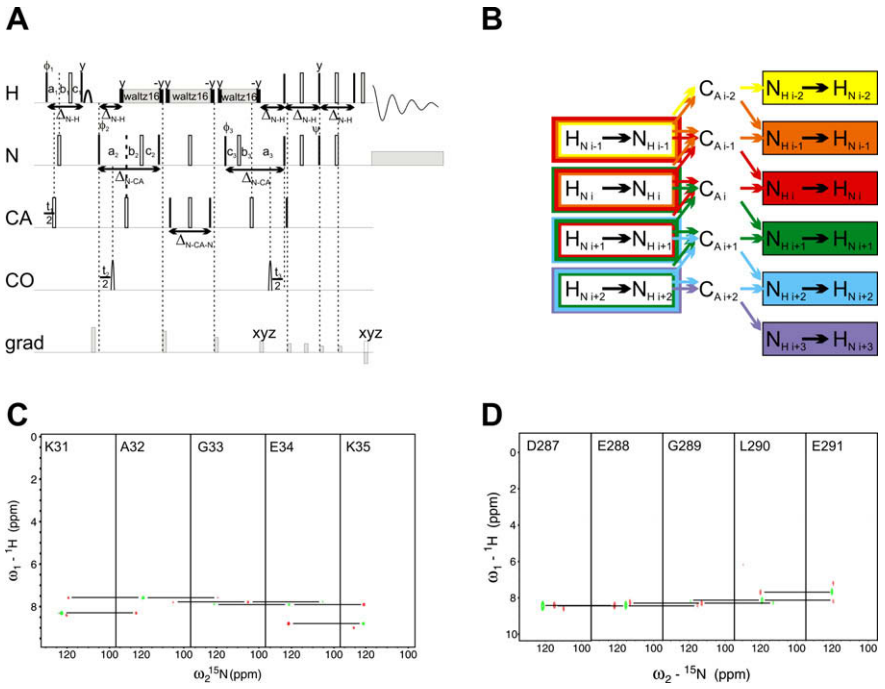

Fig. 5. 4D HN(CA)NH technique: (Panel A) Pulse sequence, proton evolution and both nitrogen evolutions are in semi-constant-time mode ($a_i = (t_i + \Delta)/2$, $b_i = t_i(1 - \Delta/t_i^{max})/2$, $c_i = \Delta(1 - t_i/t_i^{max})/2$) or in constant-time mode ($a_i = (\Delta t_i)/2$, $b_i = 0$, $c_i = (\Delta - t_i)/2$), where Δ stands for Δ_{N-H} and Δ_{N-CA}, respectively, t_i is the evolution time in ith dimension and t_i^{max} is maximal length of the evolution time delay. Delays were set as follows: Δ_{N-CA-N} = 30 ms, Δ_{N-H} = 5.4 ms, Δ_{N-CA} = 22 ms. (Panel B) Coherence transfer in the peptide chain. Amide nitrogen and proton frequencies (filled colored rectangles) are fixed during Fourier transformation. Frame colors for H and N indicate the correlation. Each plane contains H–N peak for i, $i + 1$ and $i - 1$ residues. (Panel C) 2D spectral planes for CsPin protein obtained by SMFT procedure performed on the 4D HN(CA)NH randomly sampled signal (Poisson disk sampling) with "fixed" frequencies obtained from 3D HNCO peak list. (Panel D) 2D spectral planes for MBP obtained in the same manner. (For interpretation of color mentioned in this figure, the reader is referred to the web version of this article.)

and H_{i+1}–N_{i+1}–N_i–H_i. On each 2D plane, three cross-peaks are expected. Intra-residual peaks are negative, while inter-residual peaks are positive. The exception here is glycine - the lack of C_β causes the inversion of peaks resulting from glycine C_α-mediated coherence transfer (see Fig. 5C and D). This change of sign facilitates identification of glycines positions.

3. Experimental

We tested the techniques presented above with the following two proteins that differ markedly in molecular weight and, therefore, also in relaxation rates and number of resonances: (a) [^{13}C,^{15}N]-labeled protein interacting with NIMA-kinase from *Cenarcheaum symbiosum*; (CsPin; 96 amino acid residues), 1.5 mM solution in 9:1 H_2O/D_2O, at pH 7.5 and 289 K, and (b) [^{13}C,^{15}N,^{2}H]-labeled maltose binding protein (MBP; 341 amino acid residues), 0.5 mM solution in 9:1 H_2O/D_2O, at pH 7.2 and 310 K. MBP was purchased from Cambridge Isotope Laboratory Inc., 271 Andover, MA, USA, and CsPin (BMRB accession number 11080) was prepared by prof. Peter Bayer of the Institute for Structural and Medicinal Biochemistry, Center for Medical Biotechnology, University of Duisburg-Essen, Germany, and kindly donated by Dr. I. Zhukov of the National Institute of Chemistry, Ljubljana, Slovenia. All spectra were recorded on a Varian NMR System 700 spectrometer equipped with a Performa XYZ PFG unit, using the standard 5 mm ^{1}H, ^{13}C, ^{15}N triple-resonance probe. High power ^{1}H, ^{13}C, and ^{15}N $\pi/2$ pulses of 6.7, 13.7 and 31.0 μs, and 6.0, 13.7 and 31.0 μs were used for CsPin and MBP, respectively. Selective CA and CO pulses (for off-resonance excitation or inversion) were realized as phase-modulated *sinc* shapes, with B_1 field strength set to have minimal effect on CO and CA, respectively. In all cases, acquisition time of 85 ms and relaxation delay of 1.2 s were used; for other experimental parameters see Tables 1 and 2 and figure captions. Pulse sequences were developed using own-written high-level programming library, and the data obtained were processed using own-written software. In a directly detected dimension, cosine square weighting function was applied prior to Fourier transformation with zero-filling to 1024 complex points. In indirectly detected dimensions, no weighting function was applied, as decaying distribution of points solves the problem of signal truncation. The resulting spectra were analyzed using SPARKY software [25].

4. Results and discussion

All spectra were analyzed considering sequential connections in the 2D cross-sections corresponding to consecutive amino acids. It is to be noted that during the analysis one can take the advantage of using more than one spectrum at a time, as SMFT for each kind of experiment is performed based on the same list of peaks. Therefore, 2D planes obtained from various experiments are arranged in the same order. This implies that even if the link between spectral cross-sections derived from an experiment (see Section 2.3 above) is missing, it can be revealed by another experiment. Especially useful is the combination of HNCACO and HNCOCA experiments. Although the latter does not contain sequential links, it is highly sensitive and, therefore, is a good source of additional information supporting the analysis of HNCACO spectra. Exemplary spectra from all discussed experiments are shown in Figs. 1–5. 2D cross-sections corresponding to five consecutive amino acids in CsPin and MBP are presented in panels C and panels D, respectively. For each protein studied, we have also calculated percentages of all inter-residual links that were found based on each experiment separately and all inter-residual links that were found based on combined data from all experiments. For experiments in which more than one peak should be common for neighboring 2D planes, we also have calculated the percentage of such multi-links, because the presence of the links reduces the possibility of faulty assignment. These statistics are presented in Table 3 (for CsPin) and Table 4 (for MBP).

Notably, the overlap between 2D planes occurred rarely, and usually only peaks belonging to desired spin system were observed, even in the case of MBP. The CsPin spectra feature smaller number of resonances with lower relaxation rates than the MBP spectra, thus almost complete assignment of its backbone resonances can be easily achieved based on the results of a single experiment. Each of the HNCACO, HNCACACB and HBHACBCANH experiments (separately) allowed establishing >96% of all inter-residual connections, whereas the effectiveness of the HN(CA)NH experiment was 87%. Combined analysis of the spectra resulted in 100% assignment.

Peak statistics of the MBP spectra show that it was impossible to achieve satisfying assignment employing any single technique. However, the combined analysis improved the results greatly. Simultaneous analysis of just two spectra (HNCACO and HNCOCA) allowed increasing the number of inter-residual links by almost 50% as compared to analysis of the HNCACO spectrum alone (Table 4).

Table 1
Indirect dimensions parameters for CsPin protein.

Technique	HNCACO			HNCACACB			HAHBCACBNH			HN(CA)NH		
	CO	CA	N	CACB	CA	N	HAHB	CACB	N	H	N	N
Evolution times (ms)	20	10*	30*	10	7.1†	30*	7.5†	7.1†	30*	12†	22†	22†
Spectral widths (kHz)	3	6.2	2.5	14	14	2.5	4	14	2.5	10	2.5	2.5
Number of pts	1800			1800			1800			1800		
Exp. time (h)	22			22			22			22		
Corresponding conventional exp. time	3410			12,756			2733			4437		

* Semi-contant-time.
† Constant-time.

Table 2
Indirect dimensions parameters for MBP.

Technique	HNCOCA			HNCACO			HNCACACB			HN(CA)NH		
	CA	CO	N	CO	CA	N	CACB	CA	N	H	N	N
Evolution times (ms)	6.7	12†	30*	12	6.7†	30*	10	7.1†	22†	5.4†	21.5†	21.5†
Spectral widths (kHz)	8	3	2.8	3	8	2.8	16	10	2.8	8	2.8	2.8
Number of pts	3500			9500			9500			9500		
Exp. time (h)	43			116			116			116		
Corresponding conventional exp. time	3872			1710			45,097			10,089		

* Semi-contant-time.
† Constant-time.

Table 3
Peak linkage statistics for CsPin protein.

Experiment(s)	Percentage of links observed (%)	Percentage of multiple links* (%)
HNCACO	96.34	n.a.
HNCACACB	96.34	92.75
HN(CA)NH	87	84.51
HBHACBCANH	97.56	66.25
HNCACO + HNCACACB + HN(CA)NH	98.78	–
HNCACO + HNCACACB + HN(CA)NH + HBHACBCANH	100	–

* Percentage of double or triple links relative to total number of links observed.

Table 4
Peak linkage statistics for MBP.

Experiment(s)	Percentage of links observed (%)	Percentage of multiple links* (%)
HNCACO	62.75	n.a.
HNCACO + HNCOCA	89.47	n.a.
HN(CA)NH	61.54	52.63
HNCACACB	59.51	55.2
HNCACO + HNCACACB + HN(CA)NH	80.16	–
HNCACO + HNCACACB + HN(CA)NH + HNCOCA	91.5	–

* Percentage of double or triple links relative to total number of links observed.

Comparison of the results and experimental conditions for CsPin and MBP revealed the sensitivity problems that became significant in the case of the latter (larger) protein. Over fivefold longer experiment on MBP sample yielded significantly higher signal-to-artifact ratio despite threefold increase in the number of resonances that, because of the nature of random sampling, generate artifacts. However, the signal-to-thermal noise ratio, which is determined by absolute number of time domain samples only, was the factor that limited spectral quality. Thus, it seems that – for larger proteins – sampling sparseness is limited rather by sensitivity than by signal-to-artifact ratio. Therefore, in the case of MBP, we preferred using constant-time rather than semi-constant-time evolution to avoid additional loss of signal due to fast transverse relaxation.

There are some possibilities for improving the techniques presented in this report. Sensitivity can be gained with the use of a cryogenically cooled probe. In case of ^{2}H-labeled proteins, the use of the TROSY principle [26] would additionally increase sensitivity. Moreover, all "out and back" experiments presented above can be combined with SOFAST approach [27] to accelerate data acquisition for non-deuterated systems.

5. Conclusions

Random sampling allows developing novel NMR experiments that would not be feasible with the use of conventional sampling approach. The possibility of performing high-dimensional experiments of high resolution in reasonable time opens the perspective for novel pulse sequences resulting in spectra of significantly improved peak dispersion. It should be emphasized that the resonance assignment based on these spectra can be achieved in short time and ambiguities are very rare, due to extraordinary resolution and increased dimensionality. Moreover, the risk of faulty assignment is further reduced for experiments providing spectra with multiple inter-residual links. It seems that because of these attributes the sets of 2D spectral cross-sections obtained by the techniques shown above may be very useful for automatic assignment strategies. Also manual assignment is significantly improved by the described approach. Having only few peaks on each 2D spectral plane corresponding to separate spin system should allow one to manually assign major part of resonances within few hours.

Acknowledgments

This work was partially supported by the Grant No. N301 07131/2159 from the Ministry of Science and Higher Education. The NMR measurements were run at the Structural Research Laboratory, Faculty of Chemistry, University of Warsaw, Warsaw, Poland. The authors thank Dr. Igor Zhukov of the National Institute of Chemistry in Ljubljana, Slovenia, for the CsPin sample, and Dr. Stanisław Chrapusta of the Mossakowski Medical Research Center for his assistance in language correction. K.K. thanks the Foundation for Polish Science for supporting him with the START stipend for young scientists. A.Z.-K. thanks the Foundation for Polish Science for supporting her with the MPD Programme that was co-financed by the EU European Regional Development Fund.

References

[1] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley-Interscience, New York, 1986.
[2] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, M. Rance, Protein NMR Spectroscopy: Principles and Practice, second ed., Academic Press, Boston, 2007.
[3] G. Cornilescu, J.L. Marquardt, M. Ottiger, A. Bax, Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase, J. Am. Chem. Soc. 120 (1998) 6836–6837.
[4] M. Karplus, Vicinal proton coupling in nuclear magnetic resonance, J. Am. Chem. Soc. 85 (1963) 2870–2871.
[5] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science 278 (1997) 1111–1114.
[6] M. Ikura, L.E. Kay, A. Bax, A novel approach for sequential assignment of 1H, C-13, and N-15 spectra of larger proteins: heteronuclear triple resonance three-dimensional NMR spectroscopy. Application to calmodulin, Biochemistry 29 (1990) 4659–4667.
[7] H.N.B. Moseley, G.T. Montelione, Automated analysis of NMR assignments and structures for proteins, Curr. Opin. Struct. Biol. 9 (1999) 635.
[8] C. Szantay, NMR and the uncertainty principle: How to and how not to interpret homogeneous line broadening and pulse nonselectivity. III. Uncertainty?, Concepts Magn Reson. A 32A (2008) 302–325.
[9] 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.
[10] D. Rovnyak, D.P. Frueh, M. Sastry, Z.Y.J. Sun, A.S. Stern, J.C. Hoch, G. Wagner, Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction, J. Magn. Reson. 170 (2004) 15–21.
[11] A.S. Stern, J.C. Hoch, Maximum entropy reconstruction in NMR, in: D.M. Grant, R. Harris (Eds.), Encyclopedia of Nuclear Magnetic Resonance, vol. 8, John Wiley & Sons, Chichester, UK, 1996.
[12] V. Tugarinov, L.E. Kay, I. Ibraghimov, V.Y. Orekhov, High-resolution four-dimensional H-1–C-13 NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition, J. Am. Chem. Soc. 127 (2005) 2767–2775.
[13] F.L. Zhang, R. Bruschweiler, Indirect covariance NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 13180–13181.
[14] I. Bilinskis, Digital Alias-free Signal Processing, Wiley, 2007.
[15] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Determination of spin–spin couplings from ultra-high resolution 3D NMR spectra obtained by optimized random sampling and multidimensional Fourier transformation, J. Am. Chem. Soc. 130 (2008) 5404–5405.
[16] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, Narrow peaks and high dimensionalities: exploiting the advantages of random sampling, J. Magn. Reson. 197 (2009) 219–228.
[17] A. Bax, M. Ikura, An efficient three-dimensional NMR technique for correlating the proton and nitrogen backbone amide resonances with the alpha carbon of the preceding residue in uniformly ^{13}C/^{15}N enriched proteins, J. Biomol. NMR 1 (1991) 99–104.
[18] R.T. Clubb, V. Thanabal, G. Wagner, A constant time three-dimensional triple resonance pulse scheme to correlate intraresidue HN, ^{15}N and ^{13}C′ chemical shifts in ^{15}N–^{13}C labeled proteins, J. Magn. Reson. 97 (1992) 213–217.
[19] M. Wittekind, L. Mueller, HNCACB: a high-sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins, J. Magn. Reson. B101 (1993) 201–205.

[20] S. Grzesiek, A. Bax, An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, J. Magn. Reson. 99 (1992) 201–207.
[21] R. Weisemann, H. Ruterjans, W. Bermel, 3D triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins, J. Biomol. NMR 3 (1993) 113–120.
[22] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, Optimization of random time domain sampling in multidimensional NMR, J. Magn. Reson. 192 (2008) 123–130.
[23] B.E. Coggins, R.A. Venters, P. Zhou, Generalized reconstruction of *n*-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain, J. Am. Chem. Soc. 126 (2004) 1000–1001.
[24] B.E. Coggins, P. Zhou, PR-CALC: a program for the reconstruction of NMR spectra from projections, J. Biomol. NMR 34 (2006) 179–195.
[25] T.D. Goddard, D.G. Kneller, SPARKY 3, University of California, San Francisco.
[26] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proc. Natl. Acad. Sci. USA 94 (1997) 12366–12371.
[27] P. Schanda, E. Kupce, B. Brutscher, SOFAST–HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, J. Biomol. NMR 33 (2005) 199–211.