J Biomol NMR (2015) 62:179–190
DOI 10.1007/s10858-015-9932-9

ARTICLE

^{13}C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing

Paweł Dziekański[1,2] · Katarzyna Grudziąż[1] · Patrik Jarvoll[3] · Wiktor Koźmiński[1] · Anna Zawadzka-Kazimierczuk[1]

Received: 11 February 2015 / Accepted: 15 April 2015 / Published online: 23 April 2015

Abstract Intrinsically disordered proteins (IDPs) have recently attracted much interest, due to their role in many biological processes, including signaling and regulation mechanisms. High-dimensional ^{13}C direct-detected NMR experiments have proven exceptionally useful in case of IDPs, providing spectra with superior peak dispersion. Here, two such novel experiments recorded with non-uniform sampling are introduced, these are 5D HabCabCO (CA)NCO and 5D HNCO(CA)NCO. Together with the 4D (HACA)CON(CA)NCO, an extension of the previously published 3D experiments (Pantoja-Uceda and Santoro in J Biomol NMR 59:43–50, 2014. doi:10.1007/s10858-014-9827-1), they form a set allowing for complete and reliable resonance assignment of difficult IDPs. The processing is performed with sparse multidimensional Fourier transform based on the concept of restricting (fixing) some of spectral dimensions to a priori known resonance frequencies. In our study, a multiple-fixing method was developed, that allows easy access to spectral data. The experiments were tested on a resolution-demanding alpha-synuclein sample. Due to superior peak dispersion in high-dimensional spectrum and availability of the sequential connectivities between four consecutive residues, the overwhelming majority of resonances could be assigned automatically using the TSAR program.

Keywords Intrinsically disordered proteins · ^{13}C direct-detection NMR · High-dimensional NMR experiment · Non-uniform sampling · Automatic assignment · Sparse multidimensional Fourier transform

Introduction

Intrinsically disordered proteins (IDPs) have recently attracted much interest of researchers studying biological mechanisms (Oldfield and Dunker 2014). Notably, ca. 25–30 % of proteins found in Eukaryota are intrinsically disordered (Oldfield et al. 2005). IDPs are extremely flexible and reveal only transient secondary structure, which allows them to play many important roles in living organisms (Wright and Dyson 1999), often connected with signaling and regulation processes. They are also associated with many human diseases, like cancer (Iakoucheva et al. 2002) or neurodegenerative diseases (Uversky and Fink 2004). Much effort has been recently put into improving experimental techniques of studying IDPs. Importantly, X-ray crystallography that is often considered to be the leader in structural proteomics cannot be applied. High mobility of the polypeptide chain usually prevents crystallization, which is essential for X-ray crystallography. On the contrary, NMR provides both structural and dynamic information with atomic resolution. It allows determining the conformational propensities or finding the regions involved in ligand binding. This makes NMR the main tool for studying IDPs.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9932-9) contains supplementary material, which is available to authorized users.

✉ Anna Zawadzka-Kazimierczuk
anzaw@chem.uw.edu.pl

[1] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

[2] Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

[3] Agilent Technologies, 10 Mead Road, Yarnton OX5 1QU, UK

The first step of every NMR-based protein research is a sequence-specific assignment of resonances. A large variety of experiments appropriate for assignment of IDPs is available (Kazimierczuk et al. 2010; Mäntylahti et al. 2011; Zawadzka-Kazimierczuk et al. 2012b; Solyom et al. 2013; Pantoja-Uceda and Santoro 2013; Liu and Yang 2013; Hellman et al. 2014; Pantoja-Uceda and Santoro 2014; Reddy and Hosur 2014; Piai et al. 2014; Yao et al. 2014; Yoshimura et al. 2015). Among them ^{13}C direct-detected high-dimensional (≥4D) techniques (Nováček et al. 2011, 2012, 2013; Bermel et al. 2012, 2013a) are especially useful, due to their specific features described below.

The high dimensionality of the experiments gives the possibility to circumvent the problem of poor dispersion of chemical shifts, caused by high flexibility of IDP backbone. However, the high flexibility of IDPs has also a positive aspect—it results in reduced transverse relaxation rates when compared to structured proteins of similar size. Thus, even long pulse sequences (employed in high-dimensional experiments) are still practical.

^{13}C detection is helpful if amide protons (typically used for NMR signal detection) undergo fast chemical exchange. It also allows efficient assignment of proline-rich proteins. Proline residue does not give a signal in HN-detected spectra, creating a break in chains of sequentially-linked residues. Notably, prolines, as disorder-promoting residues, are more abundant in IDPs than in folded proteins (Dunker et al. 2008). Another benefit of carbon detection is that it takes an advantage of peak dispersion in CO dimension that is superior to HN, usually used for NMR signal detection.

High-dimensional and ^{13}C-detected experiments require special approach to data acquisition and processing. As the linewidths are inversely proportional to maximum evolution times (in each spectral dimension), conventional sampling of the evolution time space generally does not provide the sufficiently narrow peaks for ≥4D experiments within practically achievable experiment duration. This is caused by the fact that the distance between the equally-distanced sampling points is defined by the Nyquist theorem (Nyquist 2002). Therefore, to acquire high-dimensional data, non-uniform sampling (NUS) of the evolution-time space should be employed. For that, one has to choose a method of processing of such data: projection reconstruction (Freeman and Kupče 2012), automated projection spectroscopy, APSY (Hiller and Wider 2012), maximum entropy (Mobli and Hoch 2008), multi-way decomposition (Malmodin and Billeter 2005), multi-dimensional decomposition (Orekhov and Jaravine 2011), multidimensional Fourier transform (Kazimierczuk et al. 2012), compressed sensing (Holland and Gladden 2014). Finally, there remains an important problem of handling the high-dimensional spectrum. No software for displaying 5D and higher-dimensional spectra is so far available. Moreover, the full-dimensional spectrum would be a very large file (tens GB and more). Thus, usually the full-dimensional spectrum is not calculated. Some methods come with a software for automatic analysis of high-dimensional data [APSY–GAPRO (Hiller et al. 2005), multiway decomposition—PRODECOMP (Malmodin and Billeter 2006)], without giving access to a full spectrum. Another approach is calculating just the regions of the spectra that contain peaks (without losing any information), as in sparse multidimensional Fourier transform, SMFT (Kazimierczuk et al. 2009).

Despite all the aforementioned advantages, ^{13}C-detection is also associated with some experimental limitations. Firstly, ^{13}C detection features lower sensitivity than ^{1}H detection. This is primarily caused by the difference in gyromagnetic ratios of ^{1}H and ^{13}C nuclei. The problem has been alleviated by constant improvement in NMR instrumentation, including ^{13}C-detection optimized cryogenically cooled probes. Another problem of carbon-detected experiments is peak splitting caused by the large value of one-bond homonuclear C–C scalar coupling. This effect can be avoided by applying band-selective carbon homodecoupling during acquisition or so called virtual decoupling (Bermel et al. 2006a).

Taking advantage of the methods presented above (NUS, SMFT, virtual decoupling), several high-dimensional ^{13}C-detected experiments were recently elaborated. To establish the sequential connectivities between the residues, some of the experiments use aliphatic chemical shifts (HA_i–CA_i and/or HB_i–CB_i pairs), others—better resolved and thus more efficient CO_{i-1}–N_i pairs. Some of the techniques employ amide proton excitation, other excite alpha protons preserving proline signals. All the previously proposed techniques establish the sequential connectivities between two adjacent residues.

In the current study we present two 5D ^{13}C-detected experiments. They are combined with the 4D (HACA)-CON(CA)NCO experiment, an extension of the previously published 3D hacaCOncaNCO and 3D hacacoNcaNCO experiments (Pantoja-Uceda and Santoro 2014). The 4D (HACA)CON(CA)NCO experiment serves as a basis spectrum for SMFT processing. This allows us to calculate spectral cross-sections in several ways, which we refer to as multiple fixing. As a result, 4D (HACA)CON(CA)NCO spectrum provides the sequential connectivities (via CO_{i-1}–N_i pairs) between four consecutive residues, allowing to resolve even difficult cases with high degree of chemical shift degeneracy. 5D spectra provide additionally amide proton (HNCO(CA)NCO spectrum) and aliphatic HA, CA, HB, CB (HabCabCO(CA)NCO spectrum) chemical shifts. Moreover both 5D techniques provide sequential connectivities between two adjacent residues. The

set of experiments allows for efficient automatic assignment of IDPs' resonances.

Materials and methods

A sample of 1.2 mM ^{13}C,^{15}N-uniformly labeled alpha-synuclein protein (purchased from Giotto Biotech) in 20 mM sodium phosphate buffer, at pH 6.5 was used for all experiments. The experiments were performed at the temperature of 298 K, on Agilent NMR system spectrometer operating at 599.6 MHz ^{1}H, 150.8 MHz ^{13}C and 60.8 MHz ^{15}N resonance frequencies. The spectrometer was equipped with a HCN ^{13}C-enhanced cryogenic probe. High-power ^{1}H, ^{13}C and ^{15}N $\pi/2$ pulses of 11.35, 13.2 and 29.5 μs, respectively, were used. Selective CA and CO pulses were realized as phase-modulated (for off-resonance excitation or inversion) *sinc* shapes, with B1 field strength adjusted to have a minimal effect on CO and CA, respectively. Simultaneous inversion of CA and CO spins was achieved using six-element composite pulse (Shaka 1985). Decoupling of ^{1}H was achieved with waltz16 sequence (Shaka et al. 1983). All gradients employed had 500 μs of duration and were applied along the z axis. Each experiment was acquired in a pseudo-2D mode, with the States-TPPI method applied in all indirect dimensions to achieve quadrature detection. All experiments employed the IPAP approach (Bermel et al. 2006b, 2008) to avoid peak splitting in the direct dimension caused by the CA–CO scalar couplings. The in-phase (IP) and anti-phase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded (Bermel et al. 2006a). In all experiments four scans per increment were acquired. The acquisition time was set to 106 ms and relaxation delay to 1 s. The experiments were performed using random on-grid Poisson disk sampling with sampling density set according to a Gaussian distribution $\exp(-(t/0.5)^2)$ with regard to maximum evolution time (Kazimierczuk et al. 2008). The time schedules were generated using the *RSPack* program available from http://nmr.cent3.uw.edu.pl/software. The numbers of indirectly-detected time points, maximum evolution times and spectral widths in indirect dimensions are shown in Table 1. The pulse sequences were written for Agilent spectrometers using own-developed programming library and are available from the authors upon request.

The multidimensional Fourier transform (MFT) (Kazimierczuk et al. 2006), implemented in the *ToASTD* program, was used for processing of the basis 4D (HACA)CON(CA)NCO experiment. The sparse multidimensional Fourier transform (SMFT) (Kazimierczuk et al. 2009), implemented in the *reduced* program, with 'fixed' frequencies derived from the basis spectrum peak list, was used to obtain 2D cross-sections of all experiments. Both *ToASTD* and *reduced* programs are available from http://nmr.cent3.uw.edu.pl/software. Prior to direct-dimension processing, both programs performed the appropriate handling of in-phase and anti-phase components so that no doublets appeared in the spectrum. For processing of directly detected dimension, cosine square weighting function was used prior to Fourier transform with zero-filling to 2048 complex points. No apodization was applied in indirect dimensions.

The resulting spectra were displayed using the SPARKY software (Goddard and Kneller 2002). The assignment of protein resonances was done automatically using the TSAR program (Zawadzka-Kazimierczuk et al. 2012a), available from http://nmr.cent3.uw.edu.pl/software. The TSAR input files in which the new experiment types are defined are shown in Supplementary Materials.

Results and discussion

SMFT processing

As stated above, handling of high-dimensional spectra is problematic due to large file size and lack of appropriate software for displaying the high-dimensional spectra. Furthermore, such spectra are mostly empty: in multidimensional space (of several hundred frequency points in each dimension) typically just several hundred peaks are present (few peaks per residue). The sparse multidimensional Fourier transform, SMFT (Kazimierczuk et al. 2009), allows to limit the processing to the regions that contain peaks. The information on the location of the regions is gathered by recording certain basis spectrum that shares some dimensions with the high-dimensional spectrum. Thus, for each peak of the basis spectrum, one (or more) cross-section(s) of the high-dimensional spectrum can be calculated, by fixing the frequencies obtained from the basis peak list. Up to now, the basis spectrum was chosen in a way that allowed to calculate just a single cross-section for each peak (Kazimierczuk et al. 2009, 2010; Nováček et al. 2011; Zawadzka-Kazimierczuk et al. 2012b; Bermel et al. 2013a). In the present study, we chose the basis spectrum in a way that allows us to calculate a few different cross-sections for each basis peak. Such multiple fixing, will make it possible to do an easy expansion of the spin systems and thus facilitates the establishment of sequential connectivities.

Experimental techniques and data processing

We developed a set of ^{13}C-detected high-dimensional techniques, consisting of 4D (HACA)CON(CA)NCO,

Table 1 Experimental parameters for indirect dimensions of all experiments

Experiment	ni[a]	Measurement time	Dimension 1			Dimension 2			Dimension 3			Dimension 4		
			nucl	t^b_{max}	sw[c]	nucl	t^b_{max}	sw[c]	nucl	t^b_{max}	sw[c]	nucl	t^b_{max}	sw[c]
4D (HACA)CON(CA)NCO	2400	53 h 15 min	CO	28	2.2	N	28	2.8	N	28	2.8		n.a.	
5D HabCabCO(CA)NCO	1800[d]	78 h 40 min[d]	Hab	10	6	Cab	7.1	15	CO	28	2.2	N	28	2.8
5D HNCO(CA)NCO	1800[e]	81 h 10 min[e]	HN	10	5.2	N	28	2.8	CO	28	2.2	N	28	2.8

[a] Number of increments in all indirectly-detected dimensions together

[b] Maximum evolution time in indirectly-detected dimension, ms

[c] Spectral width in indirectly-detected dimension, kHz. The spectral widths were set in a safe manner. Unlike in conventional experiments, in NUS experiments the measurement time is very weakly dependent on the spectral widths and maximum evolution time achieved, so too wide ranges do not significantly influence quality of the spectra

[d] Identical result could be obtained using 1200 increments (the measurement time would be then 52.5 h)

[e] Identical result could be obtained using 1200 increments (the measurement time would be then c.a. 54 h)

5D HabCabCO(CA)NCO and 5D HNCO(CA)NCO. It allows the assignment of the following resonances: HN, N, CO, CA, CB, HA and HB. However, not always all of these experiments are needed to obtain the resonance assignment. As shown below, the 5D HNCO(CA)NCO experiment can be excluded and still the assignment can be obtained. All of these experiments are recorded using NUS, which allows to achieve extraordinary resolution in a reasonable experimental time. The first two of the above experiments exploit excitation of aliphatic protons, allowing preservation of proline signals. The last one, 5D HNCO(CA)NCO, excites amide protons and is the only one which yields amide proton chemical shifts. In all three techniques the signal is acquired on carbon channel, and the detected nuclei are carbonyl carbons.

4D (HACA)CON(CA)NCO is processed with multidimensional Fourier transform, MFT (Kazimierczuk et al. 2006). The obtained 4D spectrum serves as the basis spectrum during SMFT processing. Two other experiments are processed with sparse multidimensional Fourier transform, SMFT (Kazimierczuk et al. 2009). For such spectra the multiple fixing method can be applied. Interestingly, also the basis spectrum can be SMFT-processed (also with multiple fixing), basing on itself (see below).

4D (HACA)CON(CA)NCO is an extension of the previously published 3D versions of this experiment: hacaCOncaNCO and hacacoNcaNCO (Pantoja-Uceda and Santoro 2014). The pulse sequence is shown in Fig. S1 and magnetization transfer pathway is shown in Fig. 1a. The numbering of experimental dimensions (used below) is the following: (HACA)CO^1N^2(CA)N^3CO^4 (dimension 4 is directly detected). The 4D spectrum contains peaks of two types: CO_i–N_{i+1}–N_i–CO_{i-1} (negative intensity unless residue *i* is glycine) and CO_i–N_{i+1}–N_{i+1}–CO_i (positive intensity unless residue *i* is glycine). The latter peak type is diagonal and it contains no new information with respect to the former peak type. Thus, just the peaks of the former type are incorporated into the basis peak list. Once the basis peak list is ready, the same data can be processed with SMFT. As the experiment's dimensionality is four, two frequencies have to be fixed to calculate a 2D cross-section. These two frequencies can be chosen (from the basis peak list containing CO_i–N_{i+1}–N_i–CO_{i-1} peaks) in four different ways (Fig. 1b): (1i) dimension 3 fixed based on N_i frequencies and dimension 4 fixed on CO_{i-1} frequencies: on such a cross-section peaks of two types appear: CO_{i-1}–N_i and CO_i–N_{i+1}, (1ii) dimension 3 fixed based on N_{i+1} frequencies and dimension 4 fixed on CO_i frequencies: on such a cross-section peaks of two types appear: CO_i–N_{i+1} and CO_{i+1}–N_{i+2}, (1iii) dimension 2 fixed based on N_i frequencies and dimension 1 fixed on CO_{i-1} frequencies: on such a cross-section peaks of two types appear: CO_{i-2}–N_{i-1} and CO_{i-1}–N_i, (1iv) dimension 2 fixed based on N_{i+1} frequencies and dimension 1 fixed on CO_i frequencies: on this cross-section peaks of two types appear: CO_{i-1}–N_i and CO_i–N_{i+1}. Versions (1i) and (1ii) include direct dimension fixing, while (1iii) and (1iv) do not. To calculate versions (1i) and (1iv), different basis frequencies are used, nonetheless both versions yield identical peak types. Therefore they can be used to resolve overlapping cross-sections (see "The strategy of resonance assignment and multiple fixing method" section). Altogether, in all cross-sections corresponding to a single basis peak, four consecutive CO–N pairs are present. An example of 2D cross-sections obtained from 4D (HACA)-CON(CA)NCO experiment is shown in Fig. 1c–f.

5D HabCabCO(CA)NCO pulse sequence is shown in Fig. S2 and magnetization transfer pathway is shown in Fig. 2a. The numbering of experimental dimensions (used below) is the following: Hab^{1}Cab^{2}CO^3(CA)N^{4}CO^5 (dimension 5 is directly detected). The spectrum contains peaks of four types: HA_i–CA_i–CO_i–N_i–CO_{i-1} (positive

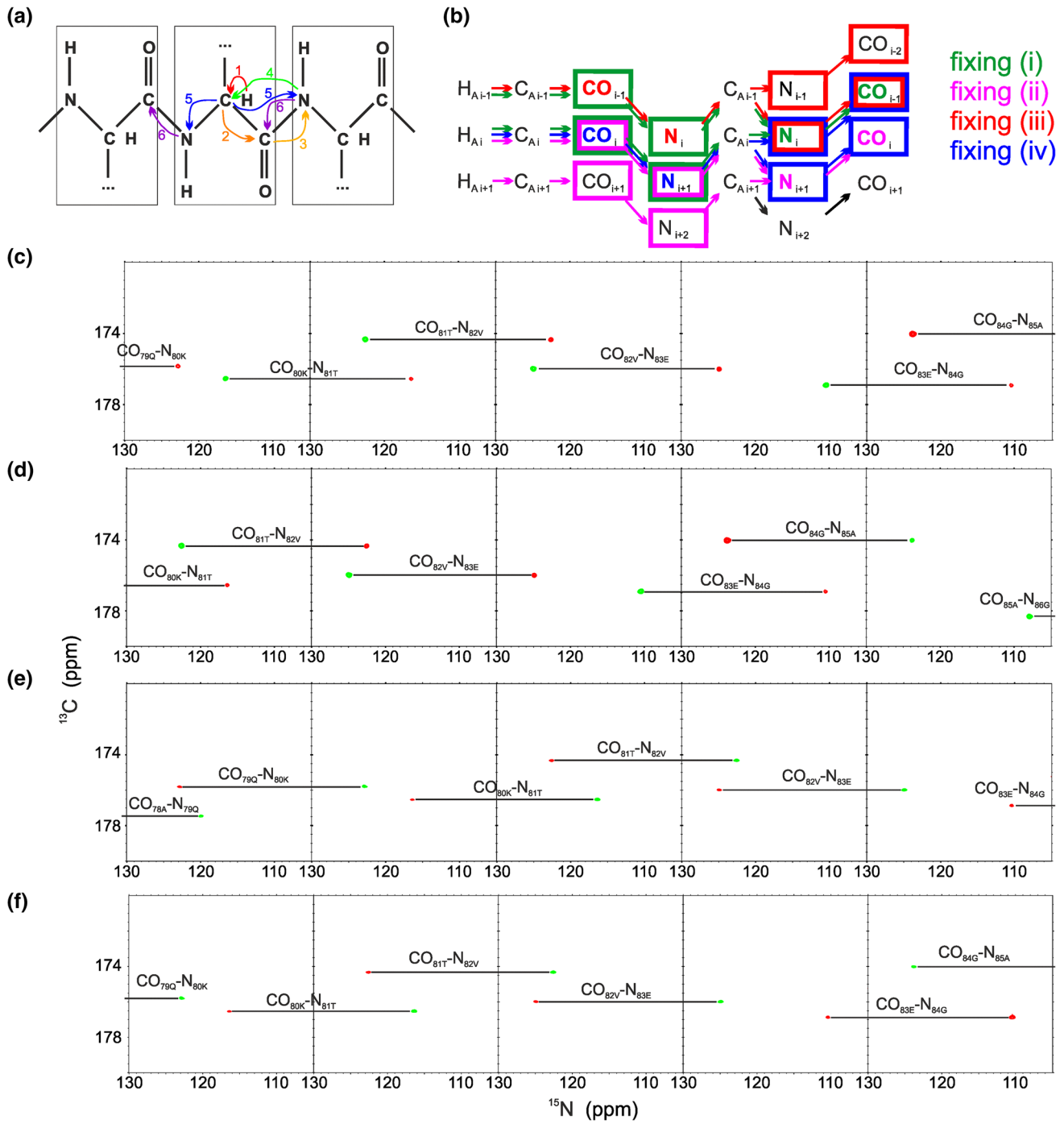

Fig. 1 4D (HACA)CON(CA)NCO technique. **a** Coherence transfer in a polypeptide chain. **b** Four ways of fixing the frequencies for SMFT processing, shown using *different colors* on a coherence transfer scheme: (*1i*)—*green*, (*1ii*)—*magenta*, (*1iii*)—*red*, (*1iv*)—*blue* (*different fixing symbols* are explained in the text). Highlighting the text with an appropriate *color* shows the fixed frequencies used during SMFT, the frequencies appearing on a cross-section are marked using frames of the *same color*. **c**–**f** 2D cross-sections of the 4D spectrum of alpha-synuclein protein, obtained by SMFT procedure, using (*1i*) fixing—panel (**c**), (*1ii*) fixing (**d**), (*1iii*) fixing (**e**) and (*1iv*) fixing (**f**). The cross-sections correspond to the 80K–84G protein fragment. Cross-sections corresponding to the same basis peak are placed one over another

intensity unless residue *i* is glycine), HB_i–CB_i–CO_i–N_i–CO_{i-1} (always positive intensity), HA_i–CA_i–CO_i–N_{i+1}–CO_i (positive intensity unless residue *i* is glycine), HB_i–CB_i–CO_i–N_{i+1}–CO_i (always positive intensity). Three versions of fixing are possible for this experiment (Fig. 2b): (2i) dimension 3 fixed based on CO_i frequencies,

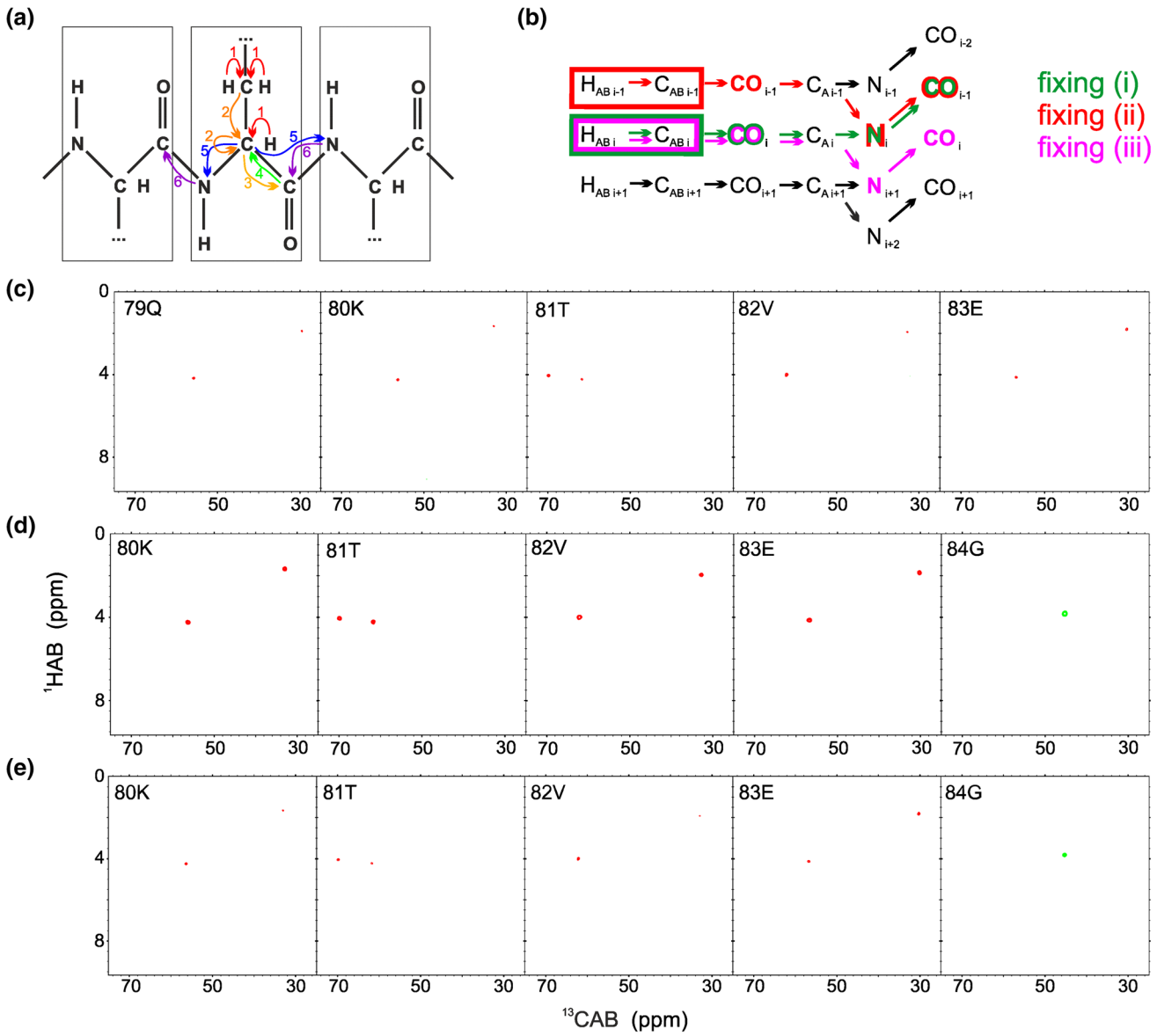

Fig. 2 5D HabCabCO(CA)NCO technique. **a** Coherence transfer in a polypeptide chain. **b** Three ways of fixing the frequencies for SMFT processing, shown using *different colors* on a coherence transfer scheme: (*1i*)—*green*, (*1ii*)—*red*, (*1iii*)—*magenta* (different fixing *symbols* are explained in the text). Highlighting the text with an appropriate *color* shows the frequencies fixed during SMFT, the frequencies appearing on a cross-section are marked using frames of the *same color*. **c**–**e** 2D cross-sections of the 5D spectrum of alpha-synuclein protein, obtained by SMFT procedure, using (*1i*) fixing (**c**), (*1ii*) fixing (**d**), (*1iii*) fixing (**e**). The cross-sections correspond to the 80K–84G protein fragment. Cross-sections corresponding to the same basis peak are placed one over another

dimension 4 fixed based on N_i frequencies and dimension 5 fixed based on CO_{i-1} frequencies: on such a cross-section peaks of two types appear: HA_i–CA_i and HB_i–CB_i, (2ii) dimensions 3 and 5 fixed based on CO_{i-1} frequencies and dimension 4 fixed based on N_i frequencies: on such a cross-section peaks of two types appear: HA_{i-1}–CA_{i-1} and HB_{i-1}–CB_{i-1}, (2iii) dimensions 3 and 5 fixed based on CO_i frequencies and dimension 4 fixed based on N_{i+1} frequencies: it yields identical peak types as version (2i), thus it can be used for cross-check in a case of overlapping planes. The overall results of this experiment are two consecutive HA–CA and HB–CB pairs, found on cross-sections corresponding to a single basis peak. Sample 2D cross-sections obtained from 5D HabCabCO(CA)NCO experiment are shown in Fig. 2c–e.

5D HNCO(CA)NCO pulse sequence is shown in Fig. S3 and magnetization transfer pathway is shown in Fig. 3a. The numbering of experimental dimensions (used below) is the following: H^N $^1N^2CO^3(CA)N^4CO^5$ (dimension 5 is directly detected). The spectrum contains peaks of two

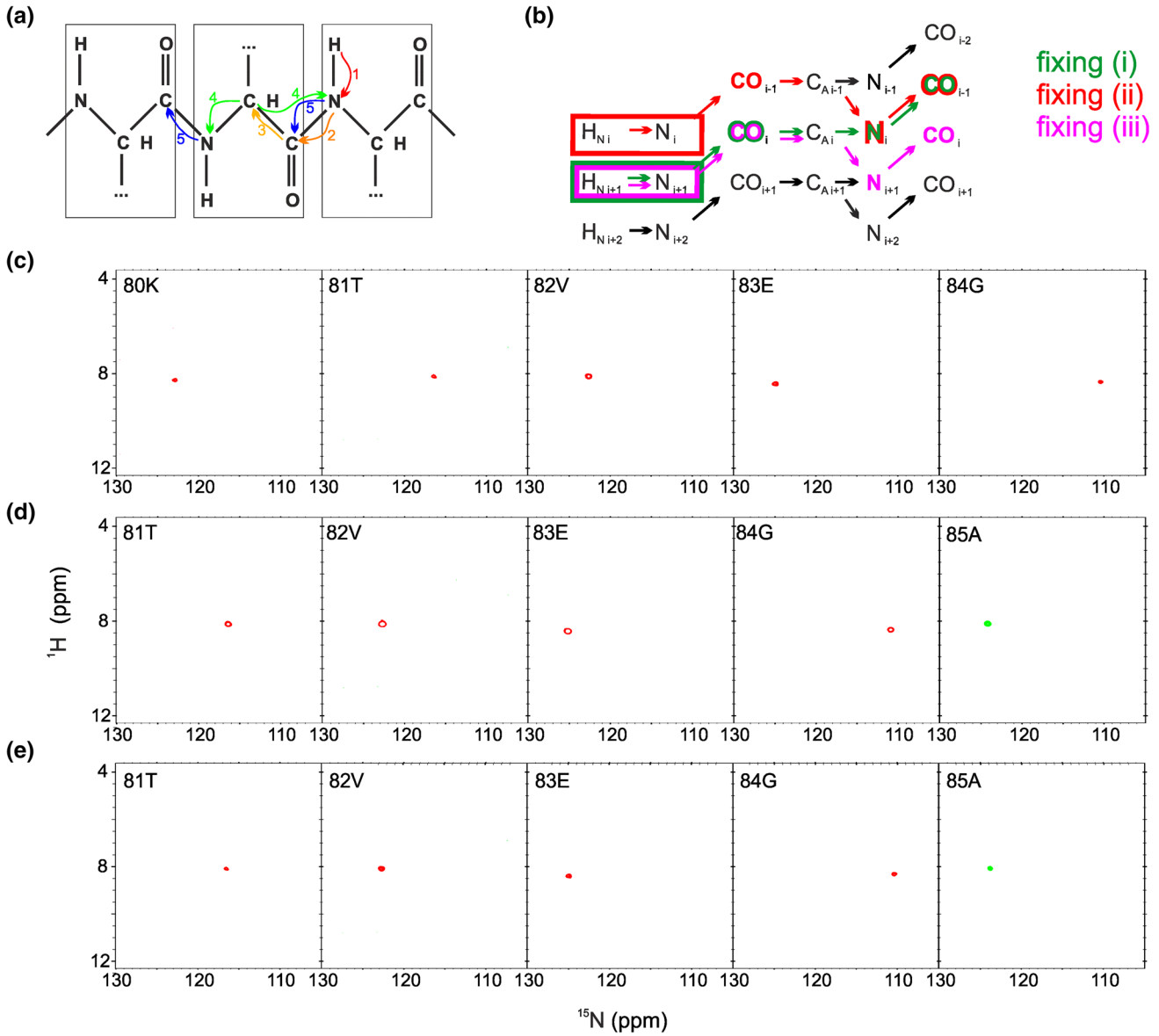

Fig. 3 5D HNCO(CA)NCO technique. **a** Coherence transfer in a polypeptide chain. **b** Three ways of fixing the frequencies for SMFT processing, shown using different *colors* on a coherence transfer scheme: (*1i*)—*green*, (*1ii*)—*red*, (*1iii*)—*magenta* (different fixing *symbols* are explained in the text). Highlighting the text with an appropriate *color* shows the frequencies fixed during SMFT, the frequencies appearing on a cross-section are marked using frames of the *same color*. **c**–**e** 2D cross-sections of the 5D spectrum of alpha-synuclein protein, obtained by SMFT procedure, using (*1i*) fixing (**c**), (*1ii*) fixing (**d**), (*1iii*) fixing (**e**). The cross-sections correspond to the 80K–84G protein fragment. Cross-sections corresponding to the same basis peak are placed one over another

types: H_i^N–N_i–CO_{i-1}–N_{i-1}–CO_{i-2} (positive intensity unless residue $i-1$ is glycine), H_i^N–N_i–CO_{i-1}–N_i–CO_{i-1} (positive intensity unless residue $i-1$ is glycine). Similarly to the 5D experiment above, also here three versions of fixing are possible (Fig. 3b): (3i) dimension 3 fixed based on CO_i frequencies, dimension 4 fixed based on N_i frequencies and dimension 5 fixed based on CO_{i-1} frequencies: on such a cross-section just a single peak appears: H_{i+1}^N–N_{i+1}, (3ii) dimensions 3 and 5 fixed based on CO_{i-1} frequencies and dimension 4 fixed based on N_i frequencies: on such a cross-section a peak of the following type appears: H_i^N–N_i, (3iii) dimensions 3 and 5 fixed based on CO_i frequencies and dimension 4 fixed based on N_{i+1} frequencies: it yields identical peak types as version (3i), thus it can be used for cross-check in a case of overlapping planes. In general, this experiment yields two consecutive H^N–N pairs, found on cross-sections corresponding to a single basis peak. Sample 2D cross-sections obtained from 5D HNCO(CA)NCO experiment are shown in Fig. 3c–e.

The strategy of resonance assignment and multiple fixing method

The strategy of resonance assignment is based on the parallel analysis of cross-sections of various multidimensional spectra. The four-level process, performed automatically by the TSAR program (Zawadzka-Kazimierczuk et al. 2012a), includes: (1) formation of cross-section spin systems (CSSSs, a data structure corresponding to each basis peak, containing the chemical shifts of the nuclei of the $i - 2$, $i - 1$, i, $i + 1$ and $i + 2$ residues), (2) finding the sequential connectivities between the CSSSs and formation of CSSSs chains, (3) recognition of amino acids that may correspond to each CSSS, (4) mapping of the CSSSs chains on the protein sequence.

Multiple fixing significantly facilitates steps (1) and (4) of the procedure. The cross-sections originating from various spectra and various ways of fixing are all calculated using the same basis peak list. Therefore, there are several 2D planes corresponding to each basis peak. These planes together contain peaks of various types and allow extending the range of chemical shifts accessible in each CSSS. For instance, the 4D (HACA)CON(CA)NCO spectrum when processed using a single fixing yields information on two consecutive CO–N pairs. When the same spectrum is processed using additionally two other fixing methods four consecutive CO–N pairs are included into the CSSS. It will give the establishment of the sequential links via three CO–N pairs instead of just one. This makes the technique applicable to the complex cases with high degree of CO–N overlap. Another example is 5D HNCO(CA)NCO spectrum, which fixed in one way yields just one H^N–N pair, not allowing to establish any sequential links. Other fixing provides the second H^N–N pair, making the sequential linking possible. In total, all the proposed experiments, processed using multiple fixing SMFT, yield the following chemical shifts: CO_{i-2}, CO_{i-1}, CO_i, CO_{i+1}, N_{i-1}, N_i, N_{i+1}, N_{i+2}, CA_{i-1}, CA_i, CB_{i-1}, CB_i, HA_{i-1}, HA_i, HB_{i-1}, HB_i, H_i^N, H_{i+1}^N (see Fig. 4).

Multiple fixing might help also in the case of cross-section overlap, when two different ways of fixing result in identical set of peaks. This is the case e.g. in 4D (HACA)CON(CA)NCO spectrum. Peaks CO_{i-1}–N_i and CO_i–N_{i+1} appear on two cross-sections, which are calculated by fixing N_i and CO_{i-1} or N_{i+1} and CO_i frequencies. It may happen that N_i–CO_{i-1} pair has chemical shifts close to another N_j–CO_{j-1}, which results in excess of peaks on both overlapping planes. It may also happen that N_{i+1}–CO_i pair is close to another N_k–CO_{k-1}, and also in these two planes there are more peaks than expected. However most likely the "additional" peaks in N_i–CO_{i-1} and N_{i+1}–CO_i planes are not identical. This difference can be used to resolve the overlap (see Fig. 5) and again confirms the high functionality of the proposed techniques for IDPs with the large extent of N–CO overlap.

It is worth mentioning that not all ways of fixing are equally useful. On those where direct dimension is not fixed the ridges of artifacts along the indirect dimension may appear, making such planes more difficult to work with than those with fixed direct dimension. The contribution of sampling artifacts to the overall spectral noise depends on the technique sensitivity. In a case of low-sensitive experiments with the low dynamic range of peak amplitudes, including typically all ^{13}C-detected ones, the artifacts constitute just a tiny proportion of all spectral noise. For more sensitive techniques, the problem of spectral artifacts can be overcome by using one of the artifact-cleaning methods (Kazimierczuk et al. 2007; Coggins and Zhou 2008; Stanek and Koźmiński 2010). Various cross-section types may differ in a degree of planes overlap. It depends on the resolution in fixed dimensions. As direct dimension features the highest resolution, usually the planes obtained without direct dimension fixing overlap more severely than those calculated with direct-dimension fixing. Usually, if two types of fixing yield identical peaks, it is more beneficial to use the one with direct-dimension fixing. In the proposed 5D experiments (HabCabCO(CA)NCO and HNCO(CA)NCO) in one fixing type (2i and 3i) three different frequencies are used for cross-sections calculation, while in two other fixing types (2ii and 3ii, 2iii and 3iii) just two different frequencies are used—dimensions 3 and 5 are fixed based on the same frequency (see Figs. 2b, 3b). Therefore, sets of frequencies used for fixing 2i and 3i are more unique then those used for other fixing types. It results with lower degree of cross-section overlap for fixing 2i and 3i.

In principle, a full-dimensional spectrum contains all the information available from the given experiment. Obviously, multiple fixing will not yield any new information, but provides an easy access to the available information. It facilitates expanding of the CSSSs to promote finding more sequential connectivities. In consequence the assignment process is more efficient.

The difficulties in the assignment process may appear when some peaks are missing in the basis spectrum. It can happen due to the low sensitivity of the basis experiment or due to overlap of the CO_i–N_{i+1}–N_i–CO_{i-1} and CO_i–N_{i+1}–N_{i+1}–CO_i peaks, which have opposite signs. Then, the whole CSSS is missing, as the corresponding cross-sections of the high-dimensional spectra are not calculated. This may interrupt the formation of chains of CSSS. Nonetheless, as the basis spectrum yields the sequential connectivities via three N–CO pairs, such a gap is not critical. Shorter chains (with a gap between) can be still sequentially linked and all the chemical shifts of the "lacking" residue can be acquired from the adjacent CSSSs.

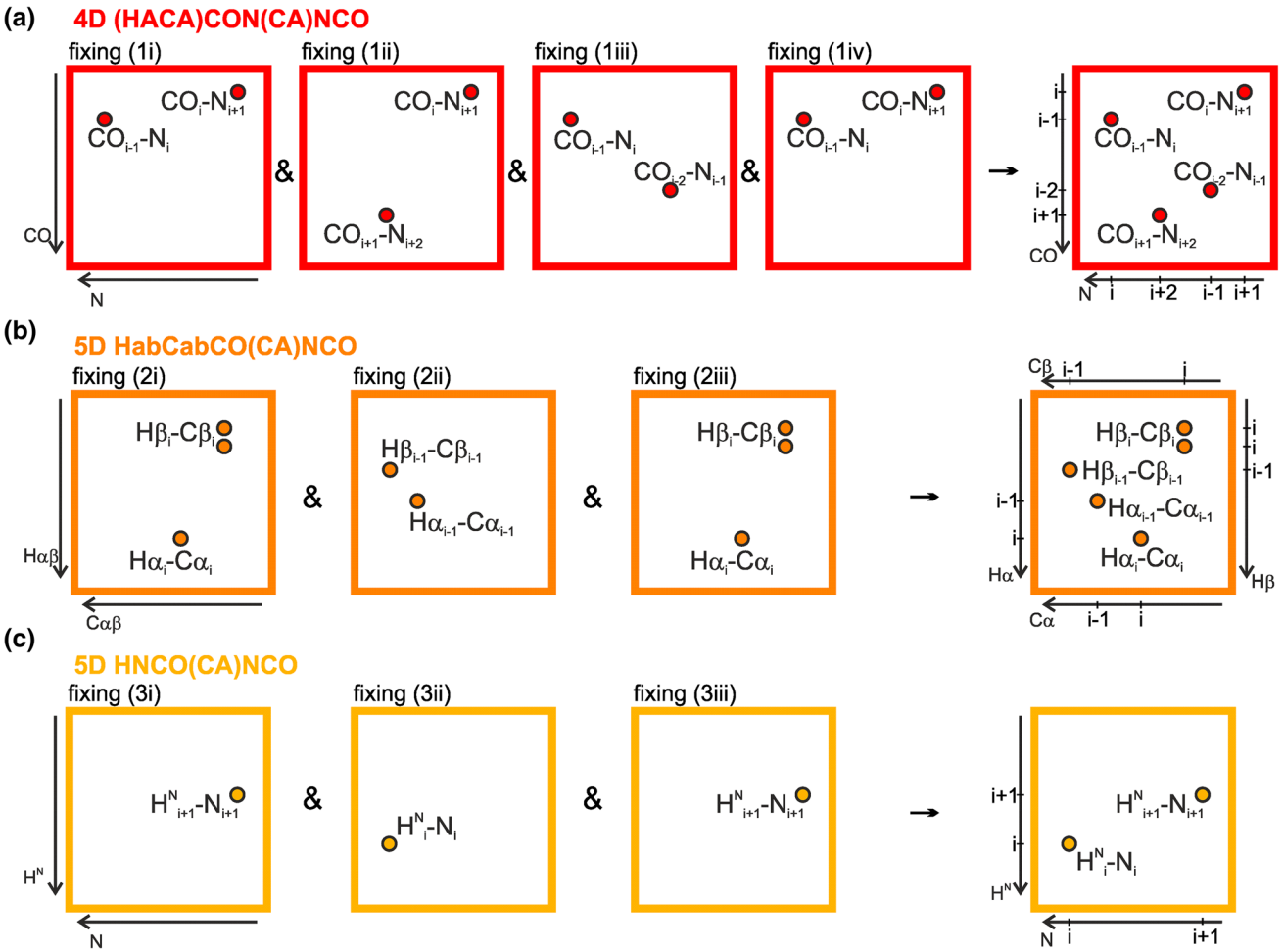

Fig. 4 Schemes of cross-sections corresponding to a *single basis peak* CO_i–N_{i+1}–N_i–CO_{i-1}. For all the techniques, cross-section of each way of fixing is shown. On a *right-hand side* of each *panel*, a hypothetical plane containing peaks from all types of fixing is shown (in practice there is no such a cross-section)

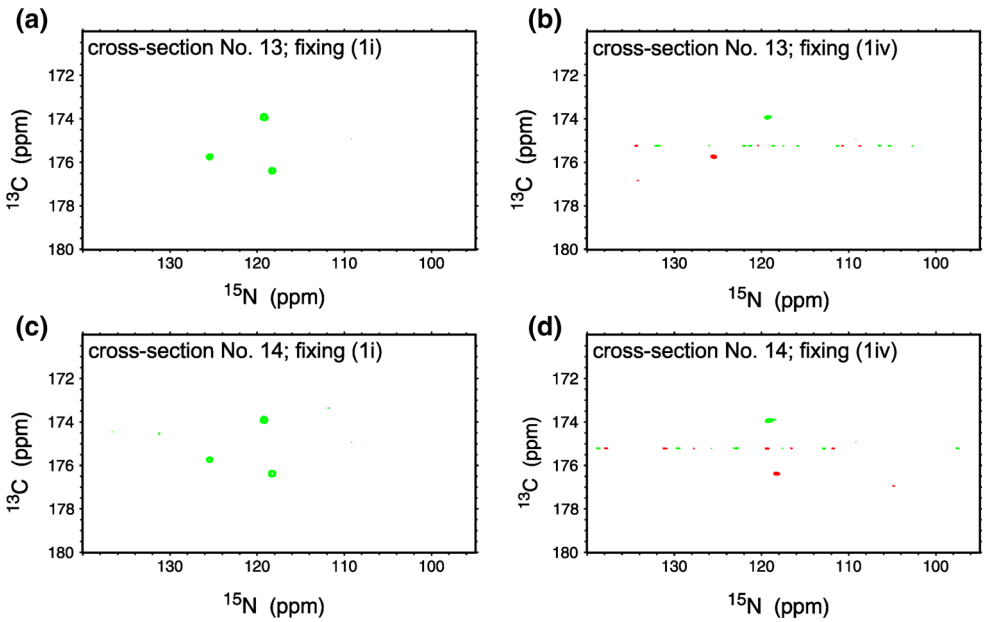

Fig. 5 An example of 4D (HACA)CON(CA)NCO (*1i*) cross-sections overlap: three peaks instead of the expected two appear on both cross-sections (number *13* and *14*). The ambiguity can be resolved by comparison with the same cross-sections calculated using (*1iv*) fixing, where identical peaks are expected

Due to high dimensionality of the experiments, the strategy is capable to cope with proteins with high degree of chemical shift overlap. Excitation of aliphatic proton nuclei (utilized in two of the three spectra used) and ^{13}C-detection makes it better suited to proline-rich IDPs than more sensitive, in the case of slow chemical exchange, H^N-excited and/or—detected techniques. Comparing to previously published strategies employing ^{13}C-detected high-dimensional experiments (Nováček et al. 2011, 2012, 2013; Bermel et al. 2012, 2013b), the proposed strategy seems to be very robust, as the sequential connectivities are established via chemical shifts of seven nuclei types: N, CO, H^N, CA, CB, HA and HB. The former three ones (N, CO, HN) are especially useful in case of IDPs. They are typically better resolved than the aliphatic carbon and proton chemical shifts, which in IDPs are strongly dependent on the particular amino acid. Moreover, two of the chemical shifts (N and CO) provide the links between four residues. Similar level of sequential linking was proposed in (Bermel et al. 2013c), which can use (dependently on the set of chosen pulse sequences) N and CO chemical shifts of three residues and H_N, CA, CB, HA and HB of two residues. The main difference between the two strategies are the fixed dimensions, which influence the level of cross-section overlap. In the strategy of Bermel et al. in 4D pulse sequences two frequencies (CO_{i-1} and N_i) are fixed for SMFT and in 5D pulse sequences CA_{i-1} is fixed additionally. In the strategy proposed in the present article, in one fixing type of the 5D techniques, CO_i, N_i and CO_{i-1} frequencies are fixed. Such triples are better resolved than the CA_{i-1}, CO_{i-1} and N_i triples utilized by Bermel et al., which is beneficial, regarding the level of cross-section overlap. On the other hand, other fixing types of the 5D spectra and the basis spectrum utilize CO_{i-1} and N_i pairs, which not always allows to resolve the cross-sections. To sum up, the two approaches are of similar quality.

Experimental results

All presented techniques were tested on alpha-synuclein protein and the data were used to assign the resonances. To get the assignment at least one experiment assuring the sequential connectivities and one experiment yielding CB chemical shifts (for amino acid recognition) are needed. Thus, assignment can be done using just the 4D (HACA)CON(CA)NCO and 5D HabCabCO(CA)NCO data (below we call it data set 1). However, such a data set does not yield H^N chemical shifts. Therefore, also the second data set (data set 2) was constructed, including additionally 5D HNCO(CA)NCO spectrum. Data set 1 was acquired within ca. 132 h of the measurement time, while data set 2—within ca. 213 h. Due to the high signal to noise ratio in the 5D spectra, the data was transformed again using smaller number of increments (on NUS data such operation does not cause resolution loss). It was found that both 5D HabCabCO(CA)NCO and 5D HNCO(CA)NCO spectra obtained using 1200 increments (2/3 of the original number of increments) contained identical set of peaks: no false peak appeared and no true peak disappeared. It means that these experiments could be acquired faster, without any information loss. In such a case, the total measurement time would be <106 h for data set 1, and almost 160 h for data set 2.

The results obtained by the TSAR program are gathered in Table 2. They are satisfactory in both cases. For data set 1 the fraction of correctly and incorrectly assigned resonances was 91.8 and 1.2 % (respectively), while for data set 2 it was 92.9 and 1.2 % (respectively). It might seem surprising that the highest percentage of incorrect assignments corresponds to CO and N, which are the nuclei that form the basis of the assignment method. To understand this fact, one should consider also the fractions of correctly assigned resonances, which are also the highest for CO and N nuclei. Simply, more spectral peaks originating from aliphatic groups than from amide and carbonyl groups were missing. Also, in some cases when one of the aliphatic peaks (HA–CA or HB–CB) was missing, then TSAR program could not decide whether the present peak was alpha or beta peak, and in consequence this peak remained unassigned. In the presented data, for some incorrectly assigned CO and N nuclei, no other chemical shift was known, thus the percentage of incorrect assignments for nuclei other than CO and N did not increase.

For both data sets the chains of cross-sections were identical. 123 cross-sections were correctly assigned. Most of them (106) were parts of long (≥8 residues) chains,

Table 2 Results obtained with the TSAR program

Data set	Total measurement (time, h)	Assigned resonances correct/incorrect (%)							
		H^N	N	CO	CA	CB	HA	HB	total
1	132 (106)[a]	n.a.	96.4/1.4	96.4/2.9	90.7/0.7	87.7/0.8	90.7/0.7	87.7/0.8	91.8/1.2
2	213 (160)[a]	94.8/0.7	96.4/1.4	96.4/2.9	90.7/0.7	87.7/0.8	90.7/0.7	87.7/0.8	92.2/1.2

[a] In parenthesis the restricted experimental times were given (the experimental times that provide identical results as full-time experiments, see the main text)

whose assignment is the most reliable. Just three of them were parts of short (1–2 residues) chains. Out of three incorrectly assigned cross-sections, two were in "chains" of length 1 and one was at the very end of a long chain.

Conclusion

In the present study we proposed two new high-dimensional experiments 5D HabCabCO(CA)NCO and 5D HNCO(CA)NCO) and a 4D (HACA)CON(CA)NCO extension of the previously published 3D experiments, dedicated for resonance assignment of proteins featuring low chemical shift dispersion, like IDPs. Since all these experiments exploit ^{13}C detection and two of them use HA and HB proton excitation, this set of experiments is suitable for samples of high content of proline residue or featuring high level of chemical exchange of amide protons. NUS allows to acquire these experiments in limited measurement time. Using sparse multidimensional Fourier transform together with multiple fixing method provides an easy access to the high-dimensional data and facilitates spectral analysis. The TSAR program can perform the assignment automatically. On the example of 140-amino-acid-long IDP it was shown that the proposed techniques enable almost complete automatic resonance assignment.

Acknowledgments Anna Zawadzka-Kazimierczuk thanks the Foundation for Polish Science for support with the POMOST program. This work was also supported by the Grant No. IP2012 062772, funded by Polish Ministry of Science and Higher Education for years 2013–2015.

References

Bermel W, Bertini I, Felli I et al (2006a) ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Prog Nucl Magn Reson Spectrosc 48:25–45. doi:10.1016/j.pnmrs.2005.09.002

Bermel W, Bertini I, Felli IC et al (2006b) Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64. doi:10.1016/j.jmr.2005.08.011

Bermel W, Felli IC, Kümmerle R, Pierattelli R (2008) ^{13}C direct-detection biomolecular NMR. Concepts Magn Reson Part A 32A:183–200. doi:10.1002/cmr.a.20109

Bermel W, Bertini I, Felli IC et al (2012) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301. doi:10.1007/s10858-012-9639-0

Bermel W, Felli IC, Gonnelli L et al (2013a) Supplementary high-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR. doi:10.1007/s10858-013-9793-z

Bermel W, Felli IC, Gonnelli L et al (2013b) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361. doi:10.1007/s10858-013-9793-z

Coggins BE, Zhou P (2008) High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. J Biomol NMR 42:225–239. doi:10.1007/s10858-008-9275-x

Dunker AK, Oldfield CJ, Meng J et al (2008) The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genom 9(Suppl 2):S1. doi:10.1186/1471-2164-9-S2-S1

Freeman R, Kupče E (2012) Concepts in projection-reconstruction. Top Curr Chem 316:1–20. doi:10.1007/128

Goddard TD, Kneller DG (2002) Sparky 3. University of California, San Francisco

Hellman M, Piirainen H, Jaakola V-P, Permi P (2014) Bridge over troubled proline: assignment of intrinsically disordered proteins using (HCA)CON(CAN)H and (HCA)N(CA)CO(N)H experiments concomitantly with HNCO and i(HCA)CO(CA)NH. J Biomol NMR 58:49–60. doi:10.1007/s10858-013-9804-0

Hiller S, Wider G (2012) Automated projection spectroscopy and its applications. Top Curr Chem 316:21–47. doi:10.1007/128

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881. doi:10.1073/pnas.0504818102

Holland DJ, Gladden LF (2014) Less is more: how compressed sensing is transforming metrology in chemistry. Angew Chemie Int Ed 53:13330–13340. doi:10.1002/anie.201400535

Iakoucheva LM, Brown CJ, Lawson JD et al (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573–584. doi:10.1016/S0022-2836(02)00969-5

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168. doi:10.1007/s10858-006-9077-y

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2007) Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets. J Magn Reson 188:344–356. doi:10.1016/j.jmr.2007.08.005

Kazimierczuk K, Zawadzka A, Koźmiński W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–130. doi:10.1016/j.jmr.2008.02.003

Kazimierczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197:219–228. doi:10.1016/j.jmr.2009.01.003

Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W (2010) Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J Magn Reson 205:286–292. doi:10.1016/j.jmr.2010.05.012

Kazimierczuk K, Misiak M, Stanek J et al (2012) Generalized Fourier transform for non-uniform sampled data. Top Curr Chem 316:79–124. doi:10.1007/128

Liu X, Yang D (2013) HN(CA)N and HN(COCA)N experiments for assignment of large disordered proteins. J Biomol NMR 57:83–89. doi:10.1007/s10858-013-9783-1

Malmodin D, Billeter M (2005) Multiway decomposition of NMR spectra with coupled evolution periods. J Am Chem Soc 127:13486–13487. doi:10.1021/ja0545822

Malmodin D, Billeter M (2006) Robust and versatile interpretation of spectra with coupled evolution periods using multi-way decomposition. Magn Reson Chem. doi:10.1002/mrc.1824

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(-CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109. doi:10.1007/s10858-011-9470-z

Mobli M, Hoch JC (2008) Maximum entropy spectral reconstruction of non-uniformly sampled data. Concepts Magn Reson Part A Bridg Educ Res 32A:436–448. doi:10.1002/cmr.a.20126

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V et al (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11. doi:10.1007/s10858-011-9496-2

Nováček J, Haba NY, Chill JH et al (2012) 4D non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148. doi:10.1007/s10858-012-9631-8

Nováček J, Janda L, Dopitová R et al (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301. doi:10.1007/s10858-013-9761-7

Nyquist H (2002) Certain topics in telegraph transmission theory. Proc IEEE 90:280–305. doi:10.1109/5.989875

Oldfield CJ, Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem 83:553–584. doi:10.1146/annurev-biochem-072711-164947

Oldfield CJ, Cheng Y, Cortese MS et al (2005) Comparing and combining predictors of mostly disordered proteins. Biochemistry 44:1989–2000. doi:10.1021/bi047993o

Orekhov VY, Jaravine VA (2011) Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Prog Nucl Magn Reson Spectrosc 59:271–292. doi:10.1016/j.pnmrs.2011.02.002

Pantoja-Uceda D, Santoro J (2013) Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR 57:57–63. doi:10.1007/s10858-013-9765-3

Pantoja-Uceda D, Santoro J (2014) New ^{13}C-detected experiments for the assignment of intrinsically disordered proteins. J Biomol NMR 59:43–50. doi:10.1007/s10858-014-9827-1

Piai A, Hošek T, Gonnelli L et al (2014) "CON–CON" assignment strategy for highly flexible intrinsically disordered proteins. J Biomol NMR 60:209–218. doi:10.1007/s10858-014-9867-6

Reddy JG, Hosur RV (2014) A reduced dimensionality NMR pulse sequence and an efficient protocol for unambiguous assignment in intrinsically disordered proteins. J Biomol NMR 59:199–210. doi:10.1007/s10858-014-9839-x

Shaka AJ (1985) Composite pulses for ultra-broadband spin inversion. Chem Phys Lett 120:201–205. doi:10.1016/0009-2614(85)87040-8

Shaka A, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340. doi:10.1016/0022-2364(83)90035-5

Solyom Z, Schwarten M, Geist L et al (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321. doi:10.1007/s10858-013-9715-0

Stanek J, Koźmiński W (2010) Iterative algorithm of discrete Fourier transform for processing randomly sampled NMR data sets. J Biomol NMR 47:65–77. doi:10.1007/s10858-010-9411-2

Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim Biophys Acta 1698:131–153. doi:10.1016/j.bbapap.2003.12.008

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331. doi:10.1006/jmbi.1999.3110

Yao X, Becker S, Zweckstetter M (2014) A six-dimensional alpha proton detection-based APSY experiment for backbone assignment of intrinsically disordered proteins. J Biomol NMR 60:231–240. doi:10.1007/s10858-014-9872-9

Yoshimura Y, Kulminskaya NV, Mulder FAA (2015) Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone (15)N or (13)C′ chemical shifts of multiple contiguous residues in highly resolved 3D spectra. J Biomol NMR. doi:10.1007/s10858-014-9890-7

Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M (2012a) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95. doi:10.1007/s10858-012-9652-3

Zawadzka-Kazimierczuk A, Koźmiński W, Sanderová H, Krásný L (2012b) High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52:329–337. doi:10.1007/s10858-012-9613-x