J Biomol NMR (2013) 57:353–361
DOI 10.1007/s10858-013-9793-z

ARTICLE

High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins

Wolfgang Bermel · Isabella C. Felli · Leonardo Gonnelli · Wiktor Koźmiński · Alessandro Piai · Roberta Pierattelli · Anna Zawadzka-Kazimierczuk

Received: 1 August 2013 / Accepted: 23 October 2013 / Published online: 8 November 2013

Abstract We present three novel exclusively heteronuclear 5D ^{13}C direct-detected NMR experiments, namely ($H^{N\text{-flip}}N$)CONCACON, (HCA)CONCACON and (H)CACON(CA)CON, designed for easy sequence-specific resonance assignment of intrinsically disordered proteins (IDPs). The experiments proposed have been optimized to overcome the drawbacks which may dramatically complicate the characterization of IDPs by NMR, namely the small dispersion of chemical shifts and the fast exchange of the amide protons with the solvent. A fast and reliable automatic assignment of α-synuclein chemical shifts was obtained with the Tool for SMFT-based Assignment of Resonances (*TSAR*) program based on the information provided by these experiments.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9793-z) contains supplementary material, which is available to authorized users.

W. Bermel
Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

I. C. Felli (✉) · L. Gonnelli · A. Piai · R. Pierattelli (✉)
CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy
e-mail: felli@cerm.unifi.it

R. Pierattelli
e-mail: pierattelli@cerm.unifi.it

I. C. Felli · R. Pierattelli
Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino, Italy

W. Koźmiński · A. Zawadzka-Kazimierczuk
Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

Keywords Intrinsically disordered proteins · ^{13}C direct-detection NMR · Non-uniform sampling · Longitudinal relaxation enhancement · Multidimensional NMR experiment · Automatic assignment

Introduction

Intrinsically disordered proteins (IDPs) are a class of flexible proteins characterized by the lack of stable secondary and tertiary structures. Despite the absence of a defined fold, IDPs are able to perform a great variety of essential functions in the cell (Wright and Dyson 1999; Uversky et al. 2000; Dunker et al. 2001; Tompa 2002, 2009; Dyson and Wright 2005), leading to a reconsideration of the well-established structure–function paradigm.

X-ray crystallography is not an appropriate technique to characterize IDPs as the extreme mobility prevents the formation of suitable crystals, and NMR spectroscopy assumes the role of best candidate to investigate their disorder. In fact, NMR can provide structural and dynamic information at atomic resolution, giving a meaningful description of the properties of all the conformations sampled by the protein.

The most accessible and informative observable that can be obtained by NMR spectroscopy and used to characterize IDPs is the chemical shift of the signals of the amino-acids nuclei. These, once assigned, can in fact be used to calculate conformational properties, to evaluate propensities for transient secondary structures and to map possible interaction/binding regions. In contrast to what is usually found for globular and folded proteins, IDPs chemical shifts are not modulated by the presence of secondary structure and their values tend to be averaged out by the

fast conformational exchange processes of the protein. This fact of course leads to crowded spectra with high chances to encounter overlapping peaks. Moreover, amide protons not being protected by the presence of secondary structural elements, often experience chemical exchange with the solvent that can be so dramatic as to broaden their signals beyond detectable limits.

Exploiting heteronuclear chemical shifts is thus mandatory for assignment strategies of IDPs due to their high chemical shift dispersion as well as to their reduced sensitivity to exchange processes (Bax and Grzesiek 1993; Dyson and Wright 2001; Panchal et al. 2001; Hiller et al. 2005; Mittag and Forman-Kay 2007; Narayanan et al. 2010; Kumar and Hosur 2011; Mantylahti et al. 2011). Thanks to great recent improvements in instrumental sensitivity, direct detection of heteronuclei can be performed to take maximum advantage of these properties. ^{13}C direct detection exclusively heteronuclear NMR experiments were successfully used for the study of IDPs (Csizmok et al. 2008; Knoblich et al. 2009; O'Hare et al. 2009; Pérez et al. 2009; Hsu et al. 2009; Novacek et al. 2012). Several additional approaches to overcome the problem of extensive cross peaks overlap have been proposed (Bermel et al. 2012a, 2013; Pantoja-Uceda and Santoro 2012, 2013; Thakur et al. 2013), including the use of high multidimensionality experiments (4-5D) (Novacek et al. 2011, 2012, 2013; Bermel et al. 2012b; Haba et al. 2013), with the goal of increasing the resolution of the spectra. To reduce the experimental time while preserving high resolution in all indirect dimensions, non-uniform sampling (NUS) (Kazimierczuk et al. 2010) and approaches for longitudinal relaxation enhancement (LRE) proved to be very useful (Schanda et al. 2007; Bermel et al. 2009b; Solyom et al. 2013; Gil et al. 2013).

Starting from the above results, we propose here three new multidimensional NMR experiments in which ^{13}C direct-detection, NUS and LRE (when appropriate) are combined in order to overcome as much as possible limitations of IDPs. The accurate and unambiguous determination of chemical shifts guaranteed by these experiments allowed the *TSAR* program (Tool for SMFT-based assignment of resonances) (Zawadzka-Kazimierczuk et al. 2012) to perform a fast, reliable and automatic assignment of the collected frequencies. The combined use of these NMR experiments and software can thus be employed to speed up the sequence-specific resonance assignment of IDPs.

The effectiveness of the experiments proposed combined with the *TSAR* program were tested on a well-studied IDP of 140 amino acids, human α-synuclein. The strategy described here can successfully be applied to other IDPs, extending the size and complexity of proteins that can be investigated by NMR.

Materials and methods

All NMR experiments were performed at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. A sample of 1.0 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 6.5 was prepared as previously described (Huang et al. 2005). EDTA and NaCl were added to reach the final concentration of 0.5 and 200 mM, respectively, and 10 % D_2O was added for the lock. All experiments were acquired at 285.5 K. Parameters specific to each experiment are reported in the captions of the figures describing the pulse sequences, all reported in the Supplementary Material. For ^{13}C band-selective π/2 and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert both C' and C^{α} [smoothed Chirp 500 μs, 25 % smoothing, 80 kHz sweep width, 11.3 kHz RF field strength (Böhlen and Bodenhausen 1993)]. The ^{13}C band selective pulses on $C^{\alpha/\beta}$, C^{α}, and C' were given at the center of each region, respectively, and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz16 (Shaka et al. 1983) (1.7 kHz) and garp4 (Shaka et al. 1985) (1.0 kHz) sequences, respectively. All gradients employed had 1 ms of duration and a sine-shape. Each experiment was acquired in a pseudo-2D mode, with the States method applied in all indirect dimensions to achieve quadrature detection. All experiments employ the IPAP approach (Bermel et al. 2006a, 2008) to remove the splitting in the direct acquisition dimension caused by the C^{α}–C' couplings. The in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded (Bermel et al. 2006c). All the experiments were performed using on-grid non-uniform sampling. The on-grid "Poisson disk" sampling scheme (Kazimierczuk et al. 2008) was chosen to generate the time schedules with the *RSPack* program. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with $\sigma = 0.5$. All the spectra were acquired using *Bruker TopSpin 1.3* software. The experimental data were converted with *nmrPipe* (Delaglio et al. 1995). 3D data were processed using the Multidimensional Fourier Transform (MFT) (Kazimierczuk et al. 2006) algorithm implemented in the *ToASTD* program. 4/5D data were processed using the Sparse MFT (SMFT) algorithm (Kazimierczuk et al. 2009) implemented in the *reduced* program. Both programs, *ToASTD* and *reduced*, are included

in the MFT package software, available at http://nmr.cent3.uw.edu.pl/software. Finally, *Sparky* (Goddard and Kneller 2000) and *TSAR* (Tool for SMFT-based assignment of resonances) (Zawadzka-Kazimierczuk et al. 2012) programs were used to analyze the spectra and automatically assign the resonances, respectively.

Results and discussion

NMR experiments design

Our strategy exploits the use of ^{13}C direct detection both to avoid the problem of amide proton exchange as well as to increase the chemical shift dispersion in the direct dimension of the experiments. The superior chemical shift dispersion of the heteronuclei with respect to protons is exploited by exclusively labeling the chemical shifts of the backbone's heteronuclei, using protons only as a starting source of magnetization to increase the overall sensitivity of the experiments. The resolution of the experiments has been maximized correlating five different frequencies in five different dimensions, all built up using long evolution times exploiting the favorable relaxation properties of IDPs. To substantially reduce the experimental time and maximize resolution, non-uniform sampling (NUS) was employed in all the proposed experiments. In the 5D ($H^{N\text{-flip}}N$)CONCACON experiment, additionally the H-flip approach was used, in order to achieve a longitudinal relaxation enhancement and thus reduce the duration of the inter-scan delay. The H-flip approach (Bermel et al. 2009a; b) allows to manipulate only the desired set of ^1H spins to shorten the period needed for magnetization recovery to the $+z$ axis. For IDPs, the gain in time is usually more substantial when selectively inverting $^1H^N$ rather than $^1H^\alpha$ spins, because $^1H^N$s experience exchange processes with the solvent, hence not perturbing the water proton spins is of high importance. On the contrary, $^1H^\alpha$ spins which of course are not affected by exchange, do not experience significant longitudinal relaxation enhancement by their selective manipulation due to the small ^1H–^1H NOEs resulting from fast tumbling and low proton density in the absence of a 3D structure. For that reason, the H-flip approach was implemented only in the $^1H^N$-start version of the experiments.

Since data sampled at not uniform intervals cannot be treated with the Fast Fourier transform (FFT), the 5D experiments were processed using the Sparse Multidimensional Fourier Transform (SMFT) algorithm (Kazimierczuk et al. 2009). The features of this type of processing greatly simplifies the visualization of the multidimensional spectra, since the analysis of the 5D spectra is reduced to the inspection of a series of 2D cross-sections extracted from the full spectra at some fixed triads of frequencies. The fixed frequencies have been collected from a 3D (H)CACON lower dimensionality spectrum (the "basis spectrum"), because the correlations it provides are also shared by all the 5D experiments discussed here. Therefore, once the peaks of the 3D basis spectrum are collected, the analysis of the 5D spectra merely consists of the examination of a series of 2D planes. In these planes the two extra dimensions (with respect to the basis spectrum), containing the sequential correlations, are displayed. The information provided by the experiments is summarized in Fig. 1, while the coherence transfer pathways are reported in Fig. 2. In the latter it is also specified which dimensions were fixed with the SMFT algorithm and which were then extracted and analyzed in the 2D cross-sections.

Since the 3D (H)CACON experiment correlates the C^α and C' of a residue with the N of the following one (C^α_{i-1}–C'_{i-1}–N_i), the 5D experiments were designed with the idea of establishing the sequential correlations along C' and N frequencies, the most dispersed in terms of chemical shift (Bermel et al. 2013). The C'–N frequencies of two neighboring residues were thus connected, adding the chemical shift labeling of the C^α to minimize the occurrence of overlap. This choice has the great advantage that the same strategy can be used also when the 4D version of these experiments is acquired, since the same sequential connections can be established, even if without the help of the C^α dimension. In this case, it is sufficient to record a 2D CON experiment (to obtain the basis spectrum) and use the 4D spectra to sequentially link all the CON peaks (C'_{i-1}–N_i).

The magnetization transfer pathway of the 5D ($H^{N\text{-flip}}N$)CONCACON experiment (Fig. 1), once the common C^α_{i-1}–C'_{i-1}–N_i frequencies (in dimensions F3, F5 and F4 respectively, Fig. 2) are fixed, provides 2D cross-sections (F1–F2 planes) containing two sequential CON peaks (C'_{i-1}–N_i and C'_{i-2}–N_{i-1}). One of them correlates the same C'_{i-1}–N_i frequencies that are retrieved in the 3D (H)CACON experiment and, together with the C^α_{i-1} frequency, fixed while performing the SMFT algorithm, whereas the other provides the chemical shifts of the preceding CON peak (C'_{i-2}–N_{i-1}). The sequence-specific assignment can then be completed just linking together the 2D planes, as shown in Fig. 3, exploiting the chemical shift dispersion of C'_{i-1}–N_i correlations. The simultaneous exploitation of the C' and N frequencies reduces the ambiguities, since sequential correlations are established along both dimensions, and the frequency labeling of the C^α contributes to resolve overlaps in 2D CON cross-sections. However, the presence of proline residues interrupts the linkage between the residues, since this experiment exploits the amide protons as a starting point of the magnetization transfer pathway. This drawback is not so dramatic if the proline content of the protein sequence is not

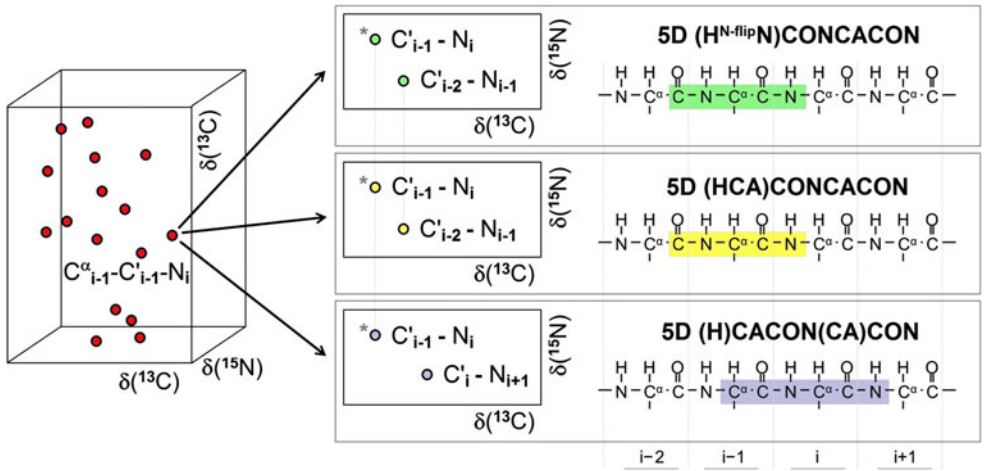

Fig. 1 The experimental strategy proposed together with the summary of the correlations provided by each experiment is schematically illustrated. Since the 5D experiments share C^{α}_{i-1}–C'_{i-1}–N_i frequencies, the 3D (H)CACON spectrum (reported on the *left*) is used as "basis spectrum" to collect these common correlations. After having fixed these frequencies with the SMFT algorithm, the information content of the various 5D experiments is easily retrieved inspecting a series of 2D cross-sections, reported in the middle part of the illustration, where the correlations provided by each experiment are shown. The CON peak which correlates the same C'_{i-1}–N_i frequencies as the associated peak in the basis spectrum is marked by an asterisk. On the right part of the figure, the nuclei involved in the coherence transfer pathways are highlighted on the backbone of the protein

large but prolines can be very abundant in IDPs (Radivojac et al. 2007). For such cases, the 5D (HCA)CONCACON experiment has to be preferred.

The 5D (HCA)CONCACON experiment can be considered as a variant of the 5D ($H^{N\text{-flip}}$N)CONCACON experiment previously described, since the magnetization transfer pathways, except for the first two steps, is the same (Fig. 1). However, the fact that this experiment exploits H^{α} protons as starting source of polarization ensures that also correlations involving the nuclei of prolines are detected. This experiment is thus recommended especially for proline-rich biomolecules and, used in combination with the previously described one, provides robust redundant information to reduce ambiguities in the assignment. Again, the sequence specific assignment is performed linking together the 2D cross-sections (F1–F2 planes) as shown in Fig. 4, where the peaks belonging to the nuclei of prolines are visible.

Both the above experiments were designed in order to provide negative peaks if the magnetization transfer pathway involves C^{α} of glycines, allowing their straightforward identification. This is obtained setting the C^{α} evolution time to $1/J_{C\alpha\text{-}C\beta}$. As an example, in Fig. 3 six sequential cross-sections of the 5D ($H^{N\text{-flip}}$N)CONCACON spectrum in which a glycine is present are reported.

Due to the fact that both experiments provide the same sequential links, one may opt for acquiring only one instead of both experiments. In this case, to choose between the 5D ($H^{N\text{-flip}}$N)CONCACON and 5D (HCA)CONCACON, one should consider proline content of the protein, level of chemical exchange, the available NMR time, stability and concentration of the sample.

Finally, a different version of the 5D (HCA)CONCACON experiment was also designed, in which the first C^{α} is frequency labeled instead of the second. This experiment, named 5D (H)CACON(CA)CON, provides different information with respect to the previously described ones, since in the 2D cross-sections (F5–F4 planes) C'_{i-1}–N_i and C'_i–N_{i+1} correlations are detectable. The sequential assignment can thus be accomplished starting from the N-terminal part and going towards the C-terminal end of the protein, whereas in the previous two experiments it proceeds in the opposite direction. Apart from this, the way to connect the cross-sections is essentially the same, since in each of them two sequential CON peaks are present. Since the H^{α} protons are used at the beginning of the magnetization transfer pathway, the prolines can be detected in this experiment too. The two C^{α} evolution times are set around $1/J_{C\alpha\text{–}C\beta}$, which means that the relative sign of cross peaks cannot be used as a straightforward criterion to identify glycines as in the previous experiments, although this can be used as a criterion to cross check results. Another difference between this experiment and those previously described is that here the direct dimension (F5) is not fixed when using the SMFT algorithm, since the C^{α}_{i-1}–C'_{i-1}–N_i frequencies that have to be fixed are all present in the indirect dimensions (F1, F2 and F3 dimensions). Artifact ridges are then present along the direct dimension (F5) of the 2D cross-sections (F5–F4 planes) in which the correlations for the sequential assignment are present (Figure S5). Even if they are not a limitation to the effectiveness of the experiment, it has to be

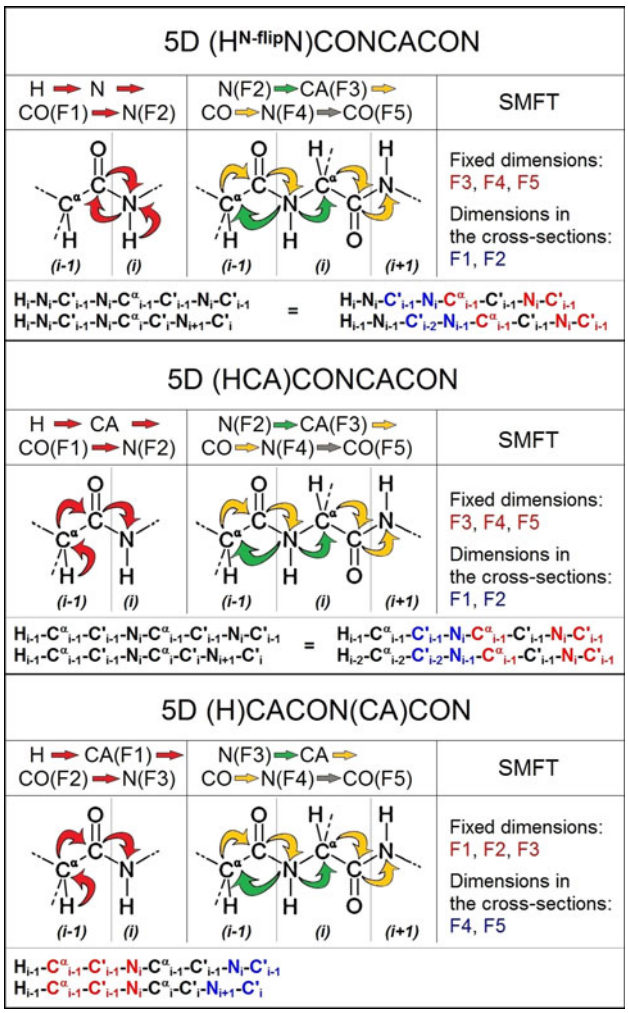

Fig. 2 The flow of the magnetization in the proposed experiments is schematically illustrated. The indirect dimensions in which chemical shift evolutions are present are indicated as F1, F2, F3 and F4, whereas the direct dimension is designated as F5. On the *left side*, the beginning of the coherence pathway of each experiment is shown: ^{1}J scalar couplings were used to transfer magnetization (*red arrows*). In the middle part of the panel, the final steps of the coherence pathways, which are common to all the experiments, are reported. Only the *arrow* representing the last transfer from N to C′ (*grey arrow* in the summary), prior to acquisition, is not shown along the backbone. Also in this case, *arrows* symbolize magnetization transfers mediated by scalar couplings: *green* when ^{1}J and ^{2}J were involved and yellow when only ^{1}J were exploited. On the *right side*, a summary of which dimensions were fixed with the SMFT algorithm and which were then extracted in the cross-sections is reported. Finally, in the bottom of each panel, the correlations retrieved in the experiment are listed explicitly

reported that they may complicate the analysis of the cross-sections and the peak picking procedure.

Automatic assignment using TSAR

Sequence-specific assignment of resonances based on the proposed experiments can be performed automatically by

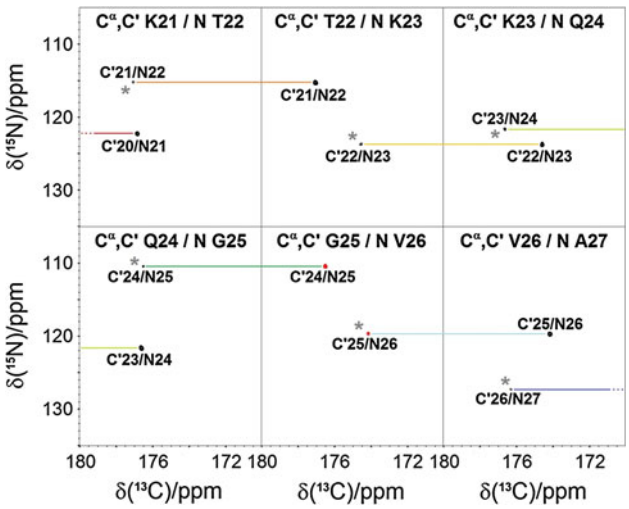

Fig. 3 Six 2D cross-sections of the (H$^{N\text{-}flip}$N)CONCACON spectrum are shown. In the top of each panel, the C^{α}_{i}–C'_{i}–N_{i+1} frequencies that have been fixed to extract the particular 2D plane are reported. In each cross-section, two sequential CON peaks are present. The CON peak which correlates the C'_{i-1}–N_i frequencies that have been used together with C^{α}_{i-1} frequency to extract the particular cross-section is labeled with an *asterisk*. *Black* peaks are positive, *red* negative. Inversion of peak sign is expected when the C^{α} belonging to a glycine is frequency labeled. Negative peaks thus allow to immediately identify glycines. The sequential specific assignment is performed connecting the CON peaks of two consecutive cross-sections, as shown with *colored lines* in the figure. Please note that the peaks are correlated using both N and C′ chemical shift, thus making the sequential assignment extremely reliable. As clearly visible, the use of five different dimensions minimize the chance of possible overlaps

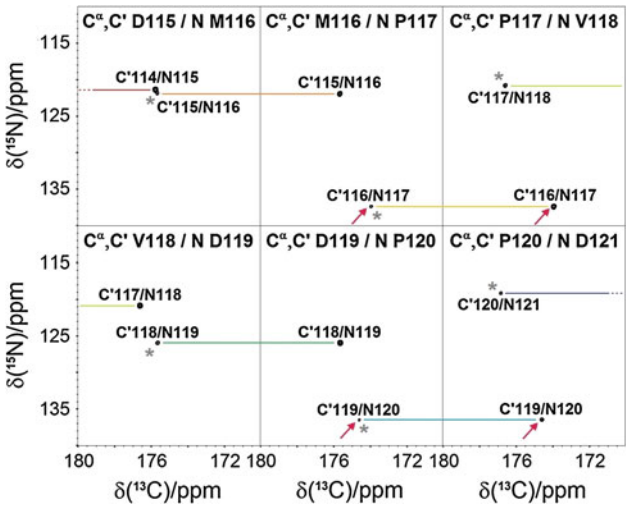

Fig. 4 Six 2D cross-sections of the (HCA)CONCACON spectrum are shown. All the labels, *lines* and *asterisks* have the same meaning as described in the captions of Fig. 3. The peaks in which the nitrogen belonging to a proline is frequency labeled (highlighted with a *red arrow*), not present in the (H$^{N\text{-}flip}$N)CONCACON spectrum, are here visible. The sequential specific assignment is thus not interrupted. The use of five different dimensions, together with the high resolution obtained in each indirect dimension, ensure to resolve very close peaks (as the case reported in the *top left panel*)

using the *TSAR* program (Zawadzka-Kazimierczuk et al. 2012), available at http://nmr.cent3.uw.edu.pl/software. This program is designed to analyze the multidimensional data processed with SMFT. What distinguishes *TSAR* from other automatic assignment algorithms is a very informative output, including also information on the correct order of cross-sections. This provides easy access to spectral data thus facilitating manual verification and/or completion of assignment. Input for the *TSAR* program includes the peak list of the basis spectrum and one or more peak lists obtained from higher-dimensional experiments. The latter ones contain only the two dimensions not fixed for SMFT, as three other frequencies are already specified in the peak list of the basis spectrum. The experiments used for *TSAR* assignment should be defined in a text file, where peak types together with their intensity signs (positive, negative, unknown or changing under specific conditions) are described (for details see Supplementary Material).

One major modification of the *TSAR* program was made to adjust it to the newly proposed experiments. In the previous version *TSAR* accepted peaks originating from three consecutive residues (on each cross-section). Now this range was enlarged to accept up to four consecutive residues. This was necessary as the simultaneous analysis of e.g. 5D (HCA)CONCACON and 5D (H)CACON(CA)CON data imply that the cross-sections obtained by fixing C^{α}_{i-1}–C'_{i-1}–N_i frequencies contain C'_{i-2}–N_{i-1}, C'_{i-1}–N_i and C'_i–N_{i+1} peaks, thus involving residues $i-2$, $i-1$, i and $i+1$.

Operation of *TSAR* includes construction of cross-section spin systems (CSSSs), where the information of corresponding cross-sections from various spectra are collected, formation of chains of CSSSs and mapping of these chains onto the primary sequence of the protein. The last step is based on identification of amino-acid type(s) for some of the CSSSs. Thus, to operate correctly, *TSAR* requires at least one experiment providing sequential connectivities and at least one experiment providing information on amino-acid type of some CSSSs. Two types of methods for amino-acid type recognition are available: statistical and topological. The former one utilizes C^{β} and/or H^{β} chemical shifts and analyzes them using the *Biological Magnetic Resonance Bank* (BMRB, Ulrich et al. 2007) statistics. The latter one exploits information on amino acid structure (e.g. absence of H^N proton in proline or presence of two possibly different H^{β} protons in all residues except for alanine, isoleucine, threonine and valine). Within the topological methods one routine is especially useful: recognition of glycine, based on peak amplitude sign. As mentioned above, it utilizes the fact that in some techniques scalar coupling between C^{α} and C^{β} carbon nuclei evolves for approximately $1/^1J_{C\alpha-C\beta}$, thus the sign of a peak corresponding to the glycine residue, not possessing the C^{β}, is inverted with respect to those of all other amino acid residues. In perspective, variants of the pulse sequences which select correlations deriving only from specific aminoacid-types can provide very useful information to contribute to the unambiguous identification of selected residue-types (Bermel et al. 2012a).

All the 5D experiments proposed in the current paper provide the sequential connectivities. Two experiments (5D ($H^{N\text{-flip}}$N)CONCACON and 5D (HCA)CONCACON) enable peak-sign-based glycine recognition, which is sufficient for CSSS's chain mapping if the content of glycine residues in the studied protein is substantial. The information on amino-acid type provided by 5D (H)CACON(CA)CON experiment cannot be exploited in the current version of the TSAR program, due to ambiguities in its interpretation: negative peak means that glycine is at $i-1$ or i position (but not at both of these positions).

Since the test protein (α-synuclein) contains almost 13 % of glycine residues in its primary sequence, glycine recognition is in principle sufficient for mapping the chains of cross-sections. Due to the protein size and to the fact that it is fully disordered, we decided however to use at least two sequential experiments able to provide links, reducing possible ambiguities in chains formation. We have constructed two data sets consisting of data from two experiments, plus 3D (H)CACON as a basis spectrum. Data set 1 contained 5D ($H^{N\text{-flip}}$N)CONCACON and 5D (H)CACON(CA)CON, while data set 2 contained 5D (HCA)CONCACON and 5D (H)CACON(CA)CON. In both cases the first experiment ensured peak-sign based glycine recognition and both experiments provided the sequential connectivities. The assignment was successful: 87.1 and 91.8 % of correctly assigned resonances were obtained for data set 1 and 2, respectively.

Despite the apparent redundancy of the information (both experiments from each data set provide the sequential links via N and C′ chemical shifts), adding the 5D (H)CACON(CA)CON to the single-experiment data sets enabled resolving the ambiguities. This is because this experiment allows to make the forward link (C'_{i-1}–N_i, C'_i–N_{i+1}), while two others show backward links (C'_{i-2}–N_{i-1}, C'_{i-1}–N_i).

Both data sets above allowed to assign the chemical shifts of N, C′ and C^{α} nuclei. When H^N, H^{α}, H^{β} and/or C^{β} chemical shifts are required as well, some other experiments have to be employed additionally. Preferably, ^{13}C-detected experiments which can be SMFT-processed using the same basis peak list (H)CACON should be employed. For instance these can be the previously published (Bermel et al., 2012b) 5D $H^{N\text{-flip}}$NCACON (containing H^N_i–N_i and H^N_{i-1}–N_{i-1} peaks on each cross-section) and 4D HCBCACON (containing H^{α}_{i-1}–C^{α}_{i-1} and H^{β}_{i-1}–C^{β}_{i-1} peaks on each cross-section). To assess the assignment capability using these sets of

Table 1 Summary of the types of experiments reported in this work

Experiment label	Name	Dimensionality	Experimental time (h)
A (basis spectrum)	(H)CACON	3D	13
B	($H^{N\text{-flip}}$N)CONCACON	5D	58
C	(HCA)CONCACON	5D	71
D	(H)CACON(CA)CON	5D	54
E	$H^{N\text{-flip}}$NCACON	5D	14
F	HCBCACON	4D	28

Table 2 *TSAR* assignment results for different data sets

Data set number	Experiments[a]	Assigned nuclei	Total exp. time (h)	Number of resonances			% of correctly assigned resonances
				Assigned correctly	Assigned incorrectly	All	
1	A + B+D	N, C′,C^{α}	123	363	10	417	87.1
2	A + C+D	N, C′,C^{α}	138	383	9	417	91.8
3	A + C+D + E	H^{N}, N, C′,C^{α}	152	511	11	551	92.7
4	A + B+D + F	N, C′,C^{α},C^{β},H^{α},H^{β}	153	707	6	798	88.6
5	A + D+E + F	H^{N}, N, C′,C^{α},C^{β},H^{α},H^{β}	109	808	7	932	86.7

[a] See Table 1 for experiments' labeling

information, three further data sets were constructed: data set 3 additionally providing H^{N} chemical shifts (5D (H)CACON(CA)CON, 5D (HCA)CONCACON and 5D $H^{N\text{-flip}}$NCACON experiments), data set 4 additionally providing H^{α}, H^{β} and C^{β} chemical shifts (5D ($H^{N\text{-flip}}$N)CONCACON, 5D (H)CACON(CA)CON and 4D HCBCACON experiments) and data set 5 providing all the backbone chemical shift (5D (H)CACON(CA)CON, 5D $H^{N\text{-flip}}$NCACON and 4D HCBCACON experiments). The results were as follows: 92.7 % of correctly assigned resonances for data set 3, 88.6 % for data set 4 and 86.7 % for data set 5. Please note, that the percentages were calculated with respect to the number of all resonances possible to assign using the given techniques, which was different for each data set (see Table 2). Thus, larger number of correctly assigned resonances does not mean that the fraction of correctly assigned resonances will also be larger. The fractions of incorrect assignments varied from 0.8 up to 2.4 % of assigned resonances. The reason for which the fractions for data sets 4 and 5 are relatively low is as follows. Both of these data sets contain 4D HCBCACON experiment that provides H^{α}, H^{β} and C^{β} chemical shifts. Each of these frequencies is known just from a single cross-section. This is not the case with N, C′ and C^{α} resonances, which are usually known from two or three cross-sections corresponding to consecutive residues. That is why if one cross-section is missing in assignment, it does not necessarily mean that any N, C′ or C^{α} resonances will be missing (usually they can be obtained from adjacent cross-sections), but it always means that some H^{α}, H^{β} and C^{β} resonances will be missing.

All the discussed data sets were gathered in Tables 1 and 2, together with experimental times required to record the respective data sets. Correctly and incorrectly assigned chemical shifts by *TSAR* were recognized upon comparison with the ones deposited in the BMRB, entry 6968 (Bermel et al. 2006b, 2012b), where the assignment of all backbone, C^{β} and H^{β} chemical shifts of α-synuclein is present.

The indicator of the reliability of the result of *TSAR* program are the lengths of assigned chains of cross-sections (Table S1): the longer they are, the more reliable the result is. In all the data sets some incorrect assignments were observed, generally for short chains (up to three cross-sections). Another problem of *TSAR* operation was observed in a case of chemical shift degeneration of pairs of residues, as in two fragments of α-synuclein: 34Lys-35Glu-36Gly and 45Lys-46Glu-47Gly. In this case the same chain of three cross-sections should be assigned to both of these sequence fragments. However, in the current version, *TSAR* cannot assign one cross-section to two different residues, thus leaving both above fragments unassigned. Importantly, if the assignment is not complete, the program provides information facilitating manual accomplishment of the task.

Conclusions

The three novel exclusively heteronuclear ^{13}C direct-detected 5D NMR experiments proposed here can be acquired in a short time preserving excellent resolution in all indirect dimensions, thanks to the use of NUS, H-flip

(when appropriate) and SMFT processing of the data. A fast and reliable automatic assignment of chemical shifts can be accomplished with the *TSAR* program. The strategy discussed here, from the acquisition of the data to the assignment of the frequencies, can be applied to investigate IDPs of increasing size and complexity.

Supplementary Material

5D ($H^{N\text{-}flip}N$)CONCACON, 5D (HCA)CONCACON, 5D (H)CACON(CA)CON and 3D (H)CACON pulse sequences; experimental parameters for the recorded experiments. Explanation of the *TSAR* input data, example of an input file with definitions of all experiments used for assignment in the article.

Acknowledgments This work was supported in part by the EC 7th Framework program BioNMR (contract 261863), by the EC Marie Curie ITN program IDPbyNMR (contract 264257) and by grant number IP2012 062772, funded by Polish Ministry of Science and Higher Education for years 2013–2014. AZK thanks the Foundation for Polish Science for support with the START and the POMOST programs.

References

Bax A, Grzesiek S (1993) Methodological advances in protein NMR. Acc Chem Res 26:131–138

Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006a) 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

Bermel W, Bertini I, Felli IC, Lee Y-M, Luchinat C, Pierattelli R (2006b) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128:3918–3919

Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006c) ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

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

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009a) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bermel W, Bertini I, Felli IC, Pierattelli R (2009b) Speeding up 13C direct detection Biomolecular NMR experiments. J Am Chem Soc 131:15339–15345

Bermel W, Bertini I, Chill JH, Felli IC, Kumar VMV, Haba N, Pierattelli R (2012a) Aminoacid-types selective ^{13}C detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). ChemBioChem 13:2425–2432

Bermel W, Bertini I, Gonnelli L, Felli IC, Kozminski W, Piai A, Pierattelli R, Stanek J (2012b) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S (2013) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR 55:231–237

Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR. J Am Chem Soc 130:16873–16879

Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX Pipes. J Biomol NMR 6:277–293

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59

Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–271

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

Gil S, Hosek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. Angew Chem Int Ed. doi:10.1002/anie.201304272

Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco

Haba NY, Gross R, Novacek J, Shaked H, Zidek L, Barda-Saad M, Chill JH (2013) NMR determines transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein. Biophysical J 105:481–493

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

Hsu ST, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in Escherichia coli. Protein Expr Purif 42:173–177

Kazimierczuk K, Zawadzka A, Kozminski W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

Kazimierczuk K, Zawadzka A, Kozminski W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–130

Kazimierczuk K, Zawadzka A, Kozminski W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197:219–228

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Kozminski W (2010) Random sampling in multidimensional NMR spectroscopy. Prog NMR Spectrosc 57:420–434

Knoblich K, Whittaker S, Ludwig C, Michiels P, Jiang T, Schafflhausen B, Günther U (2009) Backbone assignment of the N-terminal polyomavirus large T antigen. Biomol NMR Assign 3:119–123

Kumar D, Hosur RV (2011) hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone ((1) H(N), (15) N and (13) C') resonances in (15) N/(13) C-labeled proteins. Magn Reson Chem 49:575–583

Mantylahti 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

Mittag T, Forman-Kay J (2007) Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 17:3–14

Narayanan RL, Duerr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Novacek J, Zawadzka-Kazimierczuk A, Papoušková V, Zidek L, Sanderová H, Krasny L, Kozminski W, Sklenar V (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Novacek J, Haba NY, Chill JH, Zidek L, Sklenar V (2012) 4D Non-uniformly sampled HCBCACON and (1) J(NC (α))-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

Novacek J, Janda L, Dopitova R, Zidek L, Sklenar V (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

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. J Biomol NMR 20:135–147

Pantoja-Uceda D, Santoro J (2012) New amino acid residue type identification experiments valid for protonated and deuterated proteins. J Biomol NMR 54:145–153

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

Pérez Y, Gairi M, Pons M, Bernadò P (2009) Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. J Mol Biol 391:136–148

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456

Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. Proc Natl Acad Sci USA 104:11257–11262

Shaka AJ, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

Thakur A, Chandra K, Dubey A, D'Silva P, Atreya HS (2013) Rapid characterization of hydrogen exchange in proteins. Angew Chem 52:2440–2443

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533

Tompa P (2009) Structure and function of intrinsically disordered proteins. CRC Press, Boca Raton

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Shulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2007) BioMagResBank. Nucleic Acids Res 36:D402–D408

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427

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

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