

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Tossavainen, Helena; Salovaara, Santeri; Hellman, Maarit; Ihalin, Riikka; Permi, Perttu

Title: Dispersion from $C\alpha$ or NH : 4D experiments for backbone resonance assignment of intrinsically disordered proteins

Year: 2020

Version: Published version

Copyright: © The Author(s) 2020

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Tossavainen, H., Salovaara, S., Hellman, M., Ihalin, R., & Permi, P. (2020). Dispersion from $C\alpha$ or NH : 4D experiments for backbone resonance assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR*, 74(2-3), 147-159. <https://doi.org/10.1007/s10858-020-00299-w>



Dispersion from C^α or N^H: 4D experiments for backbone resonance assignment of intrinsically disordered proteins

Helena Tossavainen¹ · Santeri Salovaara¹ · Maarit Hellman¹ · Riikka Ihalin² · Perttu Permi^{1,3}

Received: 30 August 2019 / Accepted: 2 January 2020 / Published online: 13 January 2020
© The Author(s) 2020

Abstract

Resonance assignment of intrinsically disordered proteins is remarkably challenging due to scant chemical shift dispersion arising from conformational heterogeneity. The challenge is even greater if repeating segments are present in the amino acid sequence. To forward unambiguous resonance assignment of intrinsically disordered proteins, we present iHACANCO, HACACON and (HACA)CONCAHA, three H^α-detected 4D experiments with C^α as an additional dimension. In addition, we present (HACA)CON(CA)NH and (HACA)N(CA)CONH, new 4D H^α-start, H^N-detect experiments which have two N^H dimensions to enhance peak dispersion in a sequential walk through C', N^H and H^N, and provide more accurate N^H/H^N chemical shifts than those that can be obtained from a crowded ¹H, ¹⁵N-HSQC spectrum. Application of these 4D experiments is demonstrated using BilRI (165 aa), an outer-membrane intrinsically disordered protein from the opportunistic oral pathogen *Aggregatibacter actinomycetemcomitans*. BilRI amino acid sequence encompasses three very similar repeats with a 13-residue identical stretch in two of them.

Keywords *Aggregatibacter actinomycetemcomitans* · BilRI · Resonance assignment · Intrinsically disordered protein · IDP

Introduction

Aggregatibacter actinomycetemcomitans is a Gram-negative opportunistic oral pathogen that is linked to periodontitis, infection of tissues supporting the teeth (for reviews, see Fine et al. 2006; Åberg et al. 2015; Fine et al. 2019). Although *A. actinomycetemcomitans* resides in subgingival multispecies biofilms, it is also able to migrate to underlying vessels and cause systemic diseases such as cardiovascular diseases (Kozarov et al. 2005; Hyvärinen et al. 2012). The host response to biofilms is mediated by inflammatory cytokines.

In healthy junctional epithelium of tooth the balance of various cytokines and chemokines ensures that the host defence works appropriately. Some periodontal pathogens are able to impair the balance. *Porphyromonas gingivalis* can suppress the expression of chemokine interleukin(IL)-8 (Takeuchi et al. 2013) and *A. actinomycetemcomitans* biofilm is able to sequester and internalize IL-1β, IL-8 and IL-6, which leads to changes in biofilm composition and metabolic activity (Paino et al. 2011, 2012; Ahlstrand et al. 2017). It has been suggested that *A. actinomycetemcomitans* bacterial IL receptor I, BilRI is associated with this sequestering activity (Paino et al. 2013; Ahlstrand et al. 2017). BilRI is an outer membrane lipoprotein able to bind IL-1β, IL-8 and IL-10 and tumor necrosis factor (TNF)-α (Ahlstrand et al. 2017). Due to its rather low binding affinity it is assumed that BilRI acts by concentrating cytokines on cell membranes, which are then transferred to other components of the uptake system (Paino et al. 2013; Ahlstrand et al. 2017). We have engaged in the structural characterization of BilRI.

BilRI is an intrinsically disordered protein (IDP), as demonstrated by its ¹H, ¹⁵N HSQC spectrum, which displays very limited signal dispersion (Ahlstrand et al. 2017). In the H^N dimension their dispersion is only 0.63 ppm. This arises from a feature typical of IDPs, namely a biased amino acid

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10858-020-00299-w>) contains supplementary material, which is available to authorized users.

✉ Perttu Permi
perttu.permi@jyu.fi

¹ Department of Chemistry, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland

² Department of Biochemistry, University of Turku, Turku, Finland

³ Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

composition with a pronounced number of polar or charged amino acids, low number of bulky hydrophobic amino acids and lack of aromatic amino acids (Dyson 2016). BilRI amino acid sequence is dominated by alanine (23%), lysine (14%) and aspartic acid (13%) residues. However, unlike the general trend (Dunker et al. 2001), BilRI sequence contains only one proline and one glycine. BilRI is positioned among IDPs in the mean net charge-hydropathy plot (Uversky et al. 2000) (Fig. 1a). The paucity in the variability of residue's nearest neighbors narrows down the chemical shift range of a particular amino acid type, for example there are 12 Lys-Asp-Ala triplets in the sequence, which most likely results in close chemical shifts for the middle aspartic acid. Additionally, the BilRI sequence contains three very similar segments of about forty residues. The longest identical stretch is of 13 residues, present in two of these repeats, covering residues 71–82 and 111–122 (Fig. 1b).

The ^1H , ^{15}N HSQC spectra of IDPs typically being very challenging to scrutinize, the resonance assignment most often relies on spectra other than the classical H^{N} -detected experiments widely used for folded proteins (for reviews see e.g. Sattler et al. 1999; Permi and Annala 2004). Indeed, the conventional approach based on the HNCACB and HN(CO)CACB/CBCA(CO)NH experiments that link

the intra- and sequential $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$ chemical shifts to $^1\text{H}^{\text{N}}$ and ^{15}N frequencies, are very inefficient for many IDPs due to severe clustering of aliphatic carbon chemical shifts for each residue type. Instead, correlation of ^{15}N and $^{13}\text{C}'$ frequencies provide much better results for IDPs (Yao et al. 1997; Mäntylähti et al. 2009; Bermel et al. 2012). Another obstacle arises from the increasing chemical exchange rate of amide protons with water at alkali pH and/or elevated measurement temperature (Mäntylähti et al. 2010). Yet another challenge for the assignment originates from the abundance of proline residues in IDPs. As an N-substituted residue, proline lacks the amide proton, which results in gaps during the resonance assignment procedure. Although this can be a benefit in the case of globular proteins, it significantly hampers the resonance assignment of disordered systems (Hellman et al. 2014). Several different approaches have been proposed to overcome these obstacles imposed during the assignment procedure. These include increase of dimensionality from conventional 3D to 4–7D spectra (Fiorito et al. 2006; Motáčková et al. 2010; Nováček et al. 2011; Kazimierczuk et al. 2013; Brutscher et al. 2015) as well as detection of non-exchangeable spins $^{13}\text{C}'$ and $^1\text{H}^{\alpha}$ instead of $^1\text{H}^{\text{N}}$ (Bermel et al. 2006a, 2009; 2012; Mäntylähti et al. 2010, 2011; Permi and Hellman 2012). Our group has been resorting both to H^{α} -start, H^{N} -detect pulse schemes (Mäntylähti et al. 2009; Hellman et al. 2014) or complete H^{α} -detection experiments (Mäntylähti et al. 2010, 2011; Permi and Hellman 2012) for the assignment of IDPs, which overcome hurdles associated with proline assignment and exchange broadening at alkali pH or elevated temperature. In addition to reduced susceptibility towards solvent exchange induced linebroadening, the H^{α} chemical shift is extremely valuable in structural analysis. The C^{α} , H^{α} and C' shifts are particularly sensitive to the ϕ/ψ angles of the protein backbone and thus the most informative in the estimation of secondary structure content in an IDP (Borchers and Daughdrill 2018).

Comprehensive assignments allow for detailed, residue-specific analysis of structure and dynamics (Konrat 2014). There are therefore grounds for an extra effort towards a more comprehensive backbone resonance assignment. IDPs are often comprised of repetitive amino acid sequences and hence higher dimensionality in combination with high resolution offer superior results. However, the increased dispersion of signals should not be obtained, if possible, at expense of sensitivity. Here we present the resonance assignment of BilRI, whose demanding amino acid sequence necessitated development of a suite of 4D pulse sequences that offer superior signal dispersion with respect to their well-established 3D counterparts without indirect sampling associated sensitivity loss.

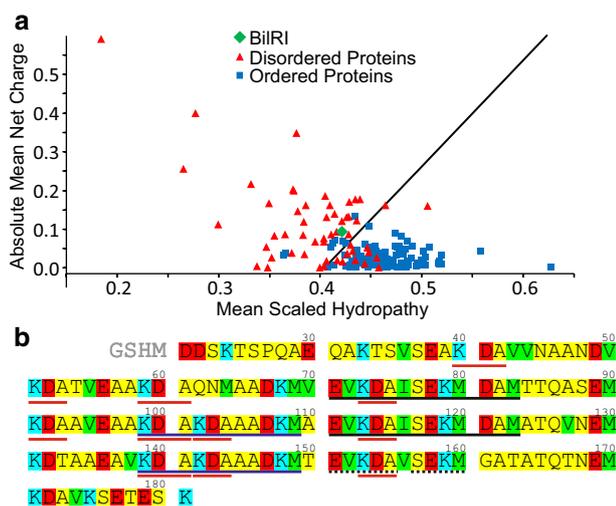


Fig. 1 **a** Mean net charge-hydropathy plot of BilRI. Data for the plot were created with the PONDR predictor (<https://www.pondr.com/>). **b** Amino acid sequence of *A. actinomycetemcomitans* BilRI. Amino acids are classified by type: yellow, small hydrophilic (A, G, N, P, Q, S, T); green, hydrophobic (I, M, V); red, negatively charged (D, E) and blue, positively charged (K) amino acids. BilRI sequence does not contain C, F, H, L, R, W or Y. The longest 13-residue repetitive segments are underlined in black, and the broken underline indicates a very similar segment. The second longest 11-residue repetitive segments are underlined in blue, and the triplet K-D-A, which recurs twelve times in the sequence is underlined in red. The first four residues are a cloning artefact

Materials and methods

Protein expression and purification

The gene encoding BilRI (residues 21–181) was cloned to pET15b vector (Novagen) into the NdeI and XhoI sites. This leads to soluble recombinant BilRI protein with N-terminal His-Tag with a thrombin cleavage site.

Production of ^{13}C , ^{15}N labeled BilRI was carried out by transforming plasmids into the BL21(DE3) cells. Cells were grown in M9 minimal media, supplemented with 1 g/l of $^{15}\text{NH}_4\text{Cl}$ and 2 g/l ^{13}C -D-glucose as the sole nitrogen or nitrogen and carbon source, respectively. Cell culture was incubated at 37 °C and temperature was decreased to 16 °C when OD of the cell culture reached 0.4 and protein production was induced with 1 mM IPTG when OD of the cell culture reached 0.6. Cells were further incubated at 16 °C for

Table 1 Data acquisition parameters

Experiment	Points in F ₁ (ms)	Points in F ₂ (ms)	Points in F ₃ (ms)	Points in F ₄ (ms)	Sampling %	Number of scans
4D iHACANCO	80 (16.6) $^{13}\text{C}'$	84 (19.1) ^{15}N	48 (3.9) $^{13}\text{C}^\alpha$	1024 (80.1) $^1\text{H}^\alpha$	10	4
4D HACACON	64 (13.2) ^{15}N	64 (14.5) $^{13}\text{C}'$	48 (6.9) $^{13}\text{C}^\alpha$	1024 (80.1) $^1\text{H}^\alpha$	10	4
4D (HACA)CONCAHA	64 (13.2) $^{13}\text{C}'$	64 (14.5) ^{15}N	48 (6.9) $^{13}\text{C}^\alpha$	1024 (80.1) $^1\text{H}^\alpha$	20	4
4D (HACA)CON(CA)NH	74 (26.3) $^{13}\text{C}'$	80 (24.7) ^{15}N	82 (25.3) ^{15}N	1024 (71.2) $^1\text{H}^\text{N}$	7	4
4D (HACA)N(CA)CONH	74 (26.3) $^{13}\text{C}'$	80 (24.7) ^{15}N	82 (25.3) ^{15}N	1024 (71.2) $^1\text{H}^\text{N}$	7	4
3D iHA(CA)NCO	116 (24.0) $^{13}\text{C}'$	200 (45.5) ^{15}N	–	1024 (80.1) $^1\text{H}^\alpha$	25	8
3D HA(CA)CON	220 (50.0) ^{15}N	230 (46.8) $^{13}\text{C}'$	–	1024 (80.1) $^1\text{H}^\alpha$	25	8
3D (HACA)CON(CA)HA	156 (32.3) $^{13}\text{C}'$	220 (50.0) ^{15}N	–	1024 (80.1) $^1\text{H}^\alpha$	25	8

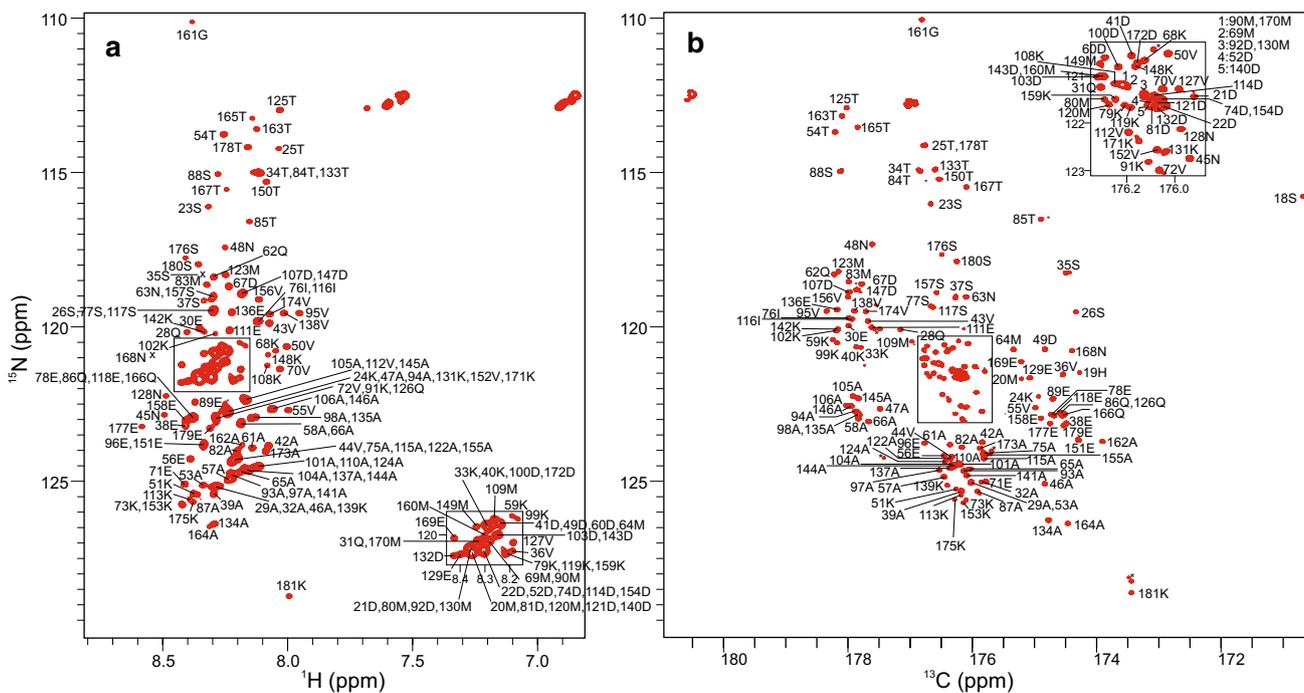
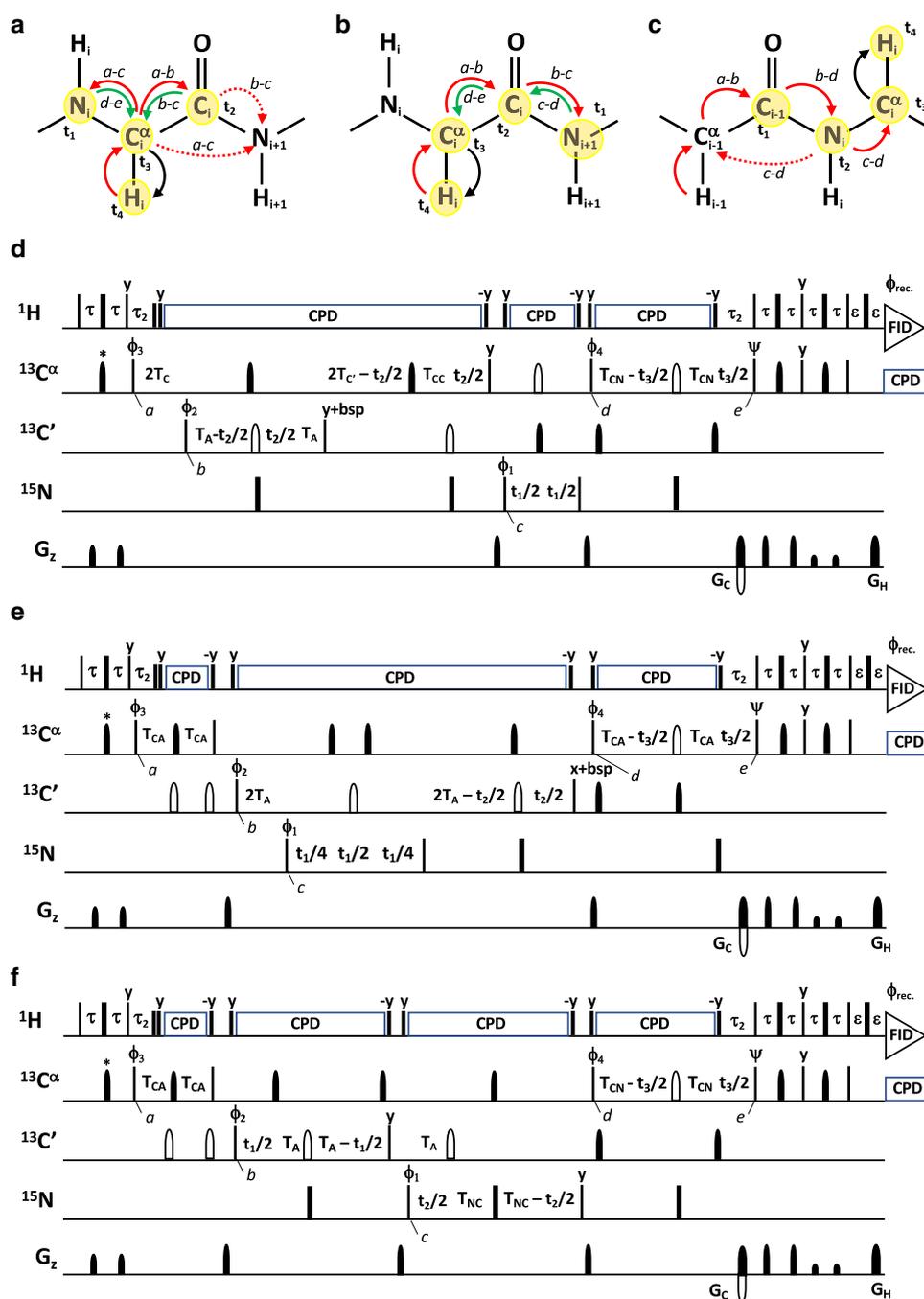


Fig. 2 **a** 2D ^1H , ^{15}N -HSQC spectrum of BilRI. The peaks are labeled with residue numbers and one-letter amino acid codes. Crosses indicate peaks found at lower contour levels. Residue numbering corresponds to that of whole BilRI protein (1–181) although the construct used was shorter (21–181). **b** 2D CON spectrum of BilRI. The peaks are labeled with residue number and amino acid code of the amide

nitrogen in the $\text{C}'\text{-N}^{\text{H}}$ pair. The peak of the only proline of the BilRI amino acid sequence, which resonates at 172.8 ($^{26}\text{Ser C}'$), 138.1 ($^{27}\text{Pro N}^{\text{H}}$) ppm is not shown. Asterisks indicate impurities. Both 2D spectra were acquired at 800 MHz ^1H frequency, 25 °C from a 1 mM BilRI sample at pH 6.5



16 h and collected by centrifugation. Cells were disrupted with sonication and resulting supernatant was clarified by centrifugation with $30,000\times g$.

Clarified supernatant of BilRI was applied to the 1-mL His GraviTrap column (GE Healthcare) and the His-Tag was removed by thrombin protease (GE Healthcare) digestion according to the manufacturer's instructions. Protease digested mixture was applied to His GraviTrap column. BilRI, without His-Tag, eluted with flow-through. Flow-through was concentrated to volume of 1 ml with Vivaspin

2 concentrator. Concentrated BilRI sample was applied into the Superdex 75 16/60 gel filtration column (GE Healthcare). Buffer used in gel filtration contained 20 mM sodium phosphate (pH 6.5) and 50 mM NaCl (NMR buffer). Fractions with pure BilRI were pooled and concentrated for NMR studies. The gel filtration was performed by using the ÄKTA Purifier FLPC purification system (GE Healthcare).

Fig. 3 a–c Schematic presentation of magnetization transfer pathway during the 4D iHACANCO **d**, 4D HACACON **e** and 4D (HACA) CONCAHA **f** experiments. Red arrows indicate direct transfer pathway from $^1\text{H}^\alpha(i)$ to $^{15}\text{N}(i)$ or $^{15}\text{N}(i+1)$ whereas green arrows indicate a nested $\text{C}^\alpha\text{C}'_z\text{N}_z \rightarrow \text{N}_z\text{C}^\alpha$ transfer in **d**) known as the *intraresidual filter* (Permi 2002; Brutscher 2002) or highly selective $\text{C}^\alpha\text{C}'_z\text{N}_z \rightarrow \text{C}^\alpha$ transfer in **e**). Arrows indicate out-and-back type magnetization transfer, whereas one-way arrows represent coherence transfer route which is unidirectional. One-letter codes above the arrows indicate time points in the pulse sequence. **d** Intraresidual iHACANCO experiment to correlate $^1\text{H}^\alpha(i)$, $^{13}\text{C}^\alpha(i)$, $^{13}\text{C}'(i)$ and $^{15}\text{N}(i)$ chemical shifts, **e** the HACACON experiment, which correlates chemical shifts of $^1\text{H}^\alpha(i)$, $^{13}\text{C}^\alpha(i)$, $^{13}\text{C}'(i)$ and $^{15}\text{N}(i+1)$ resonances. **f** The (HACA) CONCAHA experiment for correlating $^1\text{H}^\alpha(i)$, $^{13}\text{C}^\alpha(i)$, $^{13}\text{C}'(i)$ and $^{15}\text{N}(i+1)$ resonances. Narrow and wide filled bars on ^1H and ^{15}N channels correspond to rectangular 90° and 180° pulses, respectively, applied with phase x unless otherwise stated. All ^{13}C pulses are band-selective shaped pulses, denoted by filled narrow bars (90°) and filled and unfilled half ellipsoids (180°). Unfilled bars are applied on-resonance. The ^1H , ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^\alpha$ carrier positions are 4.7 (water), 118 (center of ^{15}N spectral region), 174 ppm (center of $^{13}\text{C}'$ spectral region), and 56 ppm (center of $^{13}\text{C}^\alpha$ spectral region). The ^{13}C carrier is set initially to the middle of $^{13}\text{C}'$ region (174 ppm), shifted to $^{13}\text{C}^\alpha$ region (56 ppm) prior to 90° ^{15}N pulse ϕ_1 in scheme **d**). In scheme **e**) and **f**), the carrier is initially at 56 ppm and shifted to 174 ppm prior to 90° ^{13}C pulse ϕ_2 , and shifted back to 56 ppm before 90° ^{13}C pulse ϕ_4 . The first band-selective 180° ^{13}C pulse, refocusing $^{13}\text{C}^\alpha$ magnetization (56 ppm, denoted with an asterisk) had duration of 788 μs at 800 MHz. Other band-selective 90° and 180° pulses for $^{13}\text{C}^\alpha$ (56 ppm) and $^{13}\text{C}'$ (174 ppm) were applied with durations of 240.0 μs and 192.0 μs at 800 MHz, respectively. Band-selective 90° and 180° pulses for $^{13}\text{C}'/^{13}\text{C}^\alpha$ have the shape of Q5 and Q3 (Emsley and Bodenhausen 1992) and duration of 240.0 μs and 192.0 μs at 800 MHz, respectively. The adiabatic 180° Chirp broadband inversion pulse for inverting $^{13}\text{C}^\alpha$ and $^{13}\text{C}'$ magnetization in the middle of t_1 period had duration of 500 μs at 800 MHz (Böhlen and Bodenhausen 1993). The Waltz-65 sequence (Zhou et al. 2007) with strength of 4.17 kHz was employed to decouple ^1H spins. The GARP (Shaka et al. 1985, 1987) with field strength of 4.55 kHz was used to decouple ^{13}C during acquisition. Delay durations: $\tau = 1/(4J_{\text{HC}}) \sim 1.7$ ms; $\tau_2 = 3.4$ ms (optimized for non-glycine residues) or 2.2–2.6 ms (for observing both glycine and non-glycine residues); $\varepsilon =$ duration of G_{H} + field recovery ~ 0.4 ms; $2T_{\text{C}} = 1/(2J_{\text{C}\alpha\text{C}'}) \sim 9.5$ ms; $T_{\text{CA}} = 1/(6J_{\text{C}\alpha\text{C}'}) \sim 3.3$ ms; $T_{\text{A}} = 1/(4J_{\text{C}'\text{N}}) \sim 16.6$ ms; $T_{\text{C}'} = T_{\text{C}} + T_{\text{CC}}$; $T_{\text{CC}} = 1/(J_{\text{C}\alpha\text{C}'}) - 1/(4J_{\text{C}'\text{N}}) - 1/(2J_{\text{C}\alpha\text{C}'}) \sim 0-2.5$ ms; $T_{\text{NC}} \sim 14$ ms; $T_{\text{CN}} \sim 14$ ms; $T_{\text{N}} \sim 14$ ms. Maximum t_1 , t_2 and t_3 are restrained in scheme **d**, $t_{2,\text{max}} < 2.0 * T_{\text{C}'}$, $t_{3,\text{max}} < 2.0 * T_{\text{CN}}$, in scheme **e**, $t_{2,\text{max}} < 4.0 * T_{\text{A}}$, $t_{3,\text{max}} < 2.0 * T_{\text{CA}}$, in scheme **f**, $t_{1,\text{max}} < 2.0 * T_{\text{A}}$, $t_{2,\text{max}} < 2.0 * T_{\text{NC}}$, $t_{3,\text{max}} < 2.0 * T_{\text{CN}}$. Frequency discrimination in ^{15}N and $^{13}\text{C}'$ dimensions is obtained using the States-TPPI protocol (Marion et al. 1989) applied to ϕ_1 and ϕ_2 , respectively, whereas the quadrature detection in $^{13}\text{C}^\alpha$ dimension is obtained using the sensitivity-enhanced gradient selection (Kay et al. 1992; Schleucher et al. 1994). The echo and antiecho signals in $^{13}\text{C}^\alpha$ dimension are collected separately by inverting the sign of the G_{C} gradient pulse together with the inversion of ψ , respectively. Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = x$; $\psi = x$; rec. = $x, 2(-x), x, -x, 2(x), -x$. Selective 180° pulse for $^{13}\text{C}^\alpha$ in the middle of delay $2T_{\text{A}}$ induces a Bloch-Siegert shift to $^{13}\text{C}'$ magnetization, a careful adjustment of phase (bsp) of the last $^{13}\text{C}'$ 90° (phase y) pulse is necessary in scheme **d**). Gradient strengths and durations: $G_{\text{C}} = 13$ k G/cm (1.6 ms), $G_{\text{H}} = 13$ k G/cm (0.4 ms). The pulse sequences code and parameter file for Bruker Avance system are available from authors upon request

NMR spectroscopy

BilRI NMR experiments were acquired using 0.5–1.0 mM ^{15}N , ^{13}C labeled protein samples in 5/95% $\text{D}_2\text{O}/\text{H}_2\text{O}$ at pH 6.5. Chemical shifts were referenced to external 2,2,-dimethyl-2-silapentane-5-sulfonic acid (DSS). All data were acquired at 25 °C on a Bruker AVANCE III HD 800 MHz spectrometer, equipped with a TCI $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ cryoprobe. In addition to the new 4D experiments described here, the following experiments were used in the resonance assignment: 2D ^1H , ^{15}N -HSQC, constant time ^1H , ^{13}C -HSQC, ^{13}C -detected 2D CON (Bermel et al. 2006b), 3D ^1H -detected HNCACB, CBCA(CO)NH, HNCO (reviewed in (Sattler et al. 1999; Permi and Annala 2004) and i(HACA)CO(CA)NH (Mäntylähti et al. 2009), 3D ^1H -detected HA(CA)CON, iHA(CA)NCO and (HACA)CON(CA)HA (Mäntylähti et al. 2010, 2011). All 3D/4D experiments were collected using non-uniform sampling (Table 1). Sampling densities were 25% for the 3D experiments and 7–20% for the 4D experiments. NMR data were processed with TopSpin 3.5 (Bruker Inc) and analyzed with CcpNmr Analysis v. 2.4.2 (Vranken et al. 2005). BilRI chemical shifts have been deposited to the BMRB database (www.bmrwisc.edu) with accession code 27824.

Results and discussion

Peaks in the BilRI 2D ^1H , ^{15}N -HSQC spectrum display overwhelming overlap (Fig. 2a) and a very narrow distribution in the ^1H dimension. As it was later revealed, all but three of the 21 aspartic acid residues' peaks are located in the middle region pile of peaks, 8.27–8.38, 121.2–121.9 ppm, together with 11 of the 13 methionine residues' amide peaks. The Asp and Met α signals heavily overlap also in the ^1H , ^{13}C CT-HSQC (Suppl. Fig. S1). The CON spectrum, on the other hand, shows remarkably well dispersed signals (Fig. 2b) with peaks from all $\text{C}'\text{-N}^{\text{H}}$ pairs present. We thus first attempted the assignment with spectra having the CON spectrum as the root spectrum, namely ^1H -detected experiments HA(CA)CON, iHA(CA)NCO and (HACA)CON(CA)HA (Mäntylähti et al. 2010, 2011). 3D ^1H -detected experiments HNCACB, CBCA(CO)NH, HNCO and i(HACA)CO(CA)NH (Mäntylähti et al. 2009) were acquired to collect ^1H , C' , C^α and C^β chemical shifts. With this set of seven 3D spectra, we assigned the majority of backbone resonances. However, there were several ambiguous assignments, in particular within the aforementioned repeating segments. Moreover, while precise ^1H and C' shifts were obtained from the 3D ^1H -detected experiments and the CON, due to heavy overlap, $\text{C}^\alpha/\text{C}^\beta$ shifts were far more difficult to read from the HNCACB and CBCA(CO)NH spectra. Indeed, a precise $\text{C}^\alpha/\text{C}^\beta$ chemical shift for 66% of the residues was obtained,

while for H^α , resolved peaks for 87% of the residues were observed (Suppl. Fig. S2).

To resolve these ambiguities and to extend the number of accurate chemical shifts, we resorted to 4D NMR spectroscopy. In order to bypass the C^α overlap problem in 3D H^N -detected experiments and to establish direct connectivities between H^α and C^α , and to provide dispersion to solve ambiguities arising from occasional overlap of C' , H^α resonances encountered in the 3D H^α -detected experiments, we devised H^α -detected 4D experiments with C^α as an additional dimension. In addition, we developed and employed new 4D H^α -start, H^N -detect experiments which can bridge stretches over single prolines similar to 3D experiments described in Hellman et al. (2014), but which have two N^H dimensions to enhance peak dispersion in a sequential walk through C' , N^H and H^N , and to provide for more accurate N^H/H^N chemical shifts than those that could be obtained from the crowded regions of the 1H , ^{15}N HSQC.

4D iHACANCO, HACACON and (HACA)CONCAHA experiments

The proposed 4D iHACANCO, HACACON and (HACA)CONCAHA experiments are extensions of their established 3D counterparts (Mäntylähti et al. 2010, 2011) with additional sampling on the fourth ^{13}CA dimension (Fig. 3). The coherences flow through the 4D iHACANCO, HACACON and (HACA)CONCAHA experiments in Eqs. 1, 2, 3:

$$\begin{aligned}
 &^1H^\alpha(i) \xrightarrow{2\tau(^1J_{HaCa})} {}^{13}C^\alpha(i) \xrightarrow{2T_C(^1J_{CaN}, ^1J_{CaN'}, ^1J_{CaC'})} \\
 &{}^{13}C'(i) [2T_A - t_2; ^1J_{C'N}, ^1J_{CaN}, ^2J_{CaN}] \\
 &\rightarrow {}^{13}C^\alpha(i) [2T_{C'} - t_2; ^1J_{C'\alpha N}, ^2J_{CaN}, ^1J_{CaC'}] \rightarrow {}^{15}N(i) [t_1] \\
 &\rightarrow {}^{13}C^\alpha(i) [2T_N - t_3; ^1J_{C'\alpha N}, ^2J_{CaN}] \xrightarrow{4\tau(^1J_{HaCa})} {}^1H^\alpha(i) [t_4] \quad (1)
 \end{aligned}$$

$$\begin{aligned}
 &^1H^\alpha(i) \xrightarrow{2\tau(^1J_{HaCa})} {}^{13}C^\alpha(i) \xrightarrow{2T_C(^1J_{CaC'})} {}^{13}C'(i) \xrightarrow{2T_A(^1J_{C'N})} \\
 &{}^{15}N(i+1) [t_1] \rightarrow {}^{13}C'(i) [2T_A - t_2; ^1J_{C'N}] \\
 &\rightarrow {}^{13}C^\alpha(i) [2T_{C'} - t_3; ^1J_{CaC'}] \xrightarrow{4\tau(^1J_{HaCa})} {}^1H^\alpha(i) [t_4] \quad (2)
 \end{aligned}$$

$$\begin{aligned}
 &^1H^\alpha(i-1) \xrightarrow{2\tau(^1J_{HaCa})} {}^{13}C^\alpha(i-1) \xrightarrow{2T_C(^1J_{CaC'})} \\
 &{}^{13}C'(i-1) [2T_A - t_1; ^1J_{C'N}] \\
 &\rightarrow {}^{15}N(i) [2T_{NC} - t_2; ^1J_{CaN}, ^2J_{CaN}, ^1J_{C'N}] \\
 &\rightarrow {}^{13}C^\alpha(i) [2T_A - t_2; ^1J_{C'N}] \\
 &\rightarrow {}^{13}C^\alpha(i) [2T_{CN} - t_3; ^1J_{CaN}, ^2J_{CaN}] \xrightarrow{4\tau(^1J_{HaCa})} {}^1H^\alpha(i) [t_4] \quad (3)
 \end{aligned}$$

respectively. All experiments start with the $^1H^\alpha(i) \rightarrow {}^{13}C^\alpha(i)$ transfer, and the density operator immediately after the ϕ_3 pulse is described as $H_z^\alpha(i)C_z^\alpha(i)$ (time point a). Subsequently, the magnetization is transferred to the $^{13}C'$ spin followed by the labeling of $^{13}C'$ chemical shift in t_1 or t_2 . The relevant density operator after the ϕ_2 pulse in all experiments is described as $C_z^\alpha(i)C_y'(i)$ (time point b). Next, the desired coherence is transferred to the ^{15}N spin of the sequential residue in HACACON and (HACA)CONCAHA experiments, described with the density operator $C_z^\alpha(i)C_z'(i)N_y(i+1)$ (time point c). In the iHACANCO experiment, the magnetization is solely transferred to the ^{15}N spin within the residue, described with the density operator $C_z^\alpha(i)C_z'(i)N_y(i)$ (time point c). After labeling the ^{15}N chemical shifts in t_1 (or t_2), the magnetization is transferred to the $^{13}C^\alpha$ coherence after the ϕ_4 pulse. The relevant density operators (time point d) are $C_y^\alpha(i)N_z(i)$ for iHACANCO, $C_y^\alpha(i)C_z'(i)$ for HACACON and $C_y^\alpha(i)N_z(i)$ for (HACA)CONCAHA schemes. The $^{13}C^\alpha$ chemical shift is labeled during the t_3 period between time points $d-e$.

While the HACACON (Fig. 3b) is the conventional out-and-back experiment, the (HACA)CONCAHA and iHACANCO utilize the intraresidual filter for the selective $^{13}C^\alpha(i) \rightarrow {}^{15}N(i)$ transfer (Permi 2002; Mäntylähti et al. 2010, 2011). Especially in the iHACANCO experiment, the magnetization transfer is nested and further clarification is delivered in the following. After converting magnetization to the $H_z^\alpha(i)C_z^\alpha(i)$ coherence (time point a), the $^1J_{CaC'}$, $^1J_{CaN}$, $^2J_{CaN}$ and $^1J_{CaC\beta}$ couplings are active during the time interval $(2T_C + 2T_A + 2T_{C'} + T_{CC}) = 52-57$ ms, which converts it to the $C_z^\alpha(i)N_y(i)$ coherence (time point c). However, during the delay $2T_{CC}$, that can be selected to be 0–5 ms based on the relaxation properties of $^{13}C^\alpha$ spins, only $^1J_{CaC\beta}$ is active. This is to maximize the transfer efficiency during the $(2T_C + 2T_A + 2T_{C'} + T_{CC})$ delay. Of note, to avoid chemical shift evolution of $^{13}C^\alpha$ - $^{13}C'$ multiple-quantum coherence during $2T_A - t_2$, an additional but opposite frequency labeling period for $^{13}C^\alpha$ has been implemented in the $2T_{C'}$ period. Efficiently, only the chemical shift evolution of $^{13}C'$ will take place during t_2 , but the attainable resolution is limited by $2T_{C'} (= 2T_C + T_{CC})$ i.e. $t_{2,max}$ is 19–22 ms, depending on the setting of $2T_{CC}$ (0–5 ms).

Thus, the 4D iHACANCO, HACACON and (HACA)CONCAHA experiments yield correlations at $\omega_{HA(i)}$, $\omega_{CA(i)}$, $\omega_{C'(i)}$, $\omega_{N(i)}$; $\omega_{HA(i)}$, $\omega_{CA(i)}$, $\omega_{C'(i)}$, $\omega_{N(i+1)}$, and $\omega_{HA(i)}$, $\omega_{CA(i)}$, $\omega_{C'(i-1)}$, $\omega_{N(i)}$ frequencies, respectively. Given that frequency labeling of $^{13}C^\alpha$ chemical shifts is implemented in a constant-time manner, without lengthening the actual pulse sequence and incorporating sensitivity enhanced gradient

echo in t_3 , there is no sensitivity loss involved in increasing the dimensionality of these H^α -detected experiments. Hence, the coherence transfer efficiencies provided with the corresponding 3D experiments by Mäntylähti et al. (2010, 2011) are directly comparable to 4D implementations shown in Fig. 3. Indeed, by taking into account typical values of one-bond couplings $^1J_{C\alpha C'} = 53$ Hz, $^1J_{C'N} = 15$ Hz, and $^1J_{C\alpha C\beta} = 35$ Hz, and the average random coil values for one-bond ($^1J_{C\alpha N} = 10.6$ Hz) and two-bond ($^2J_{C\alpha N} = 7.5$ Hz) couplings between backbone $^{13}C^\alpha$ and ^{15}N spins (Delaglio et al. 1991), as well as transverse relaxation times (T_2) for $^{13}C^\alpha$ ($= 100$ ms), $^{13}C'$ ($= 200$ ms) and ^{15}N ($= 200$ ms) spins, we can estimate coherence transfer efficiencies for these experiments in IDPs (Mäntylähti et al. 2010, 2011). The HACACON is superior in sensitivity ($I \sim 0.28$) in comparison to iHACANCO and (HACA)CONCAHA experiments, with coherence transfer efficiencies of 0.22 and 0.18, respectively. Particularly, for the assignment of prolines, sensitivities of iHACANCO and HACACON are superior to the (HACA)CONCAHA scheme, which yields coherence transfer of 0.026 for proline residues. The sensitivity loss is associated with the $^{15}N(i) \rightarrow ^{13}C^\alpha(i)$ transfer, $2T_{NC}$, during which the ^{15}N magnetization is further modulated by $^1J_{NC\beta}$ coupling interaction in prolines.

Figure 4 compares 3D H^α -detected with the new 4D spectra: problems associated with multifold overlap and ambiguities in choosing the right sequential connection when using 3D spectra can be surpassed, and the assignment procedure expedited by extending frequency labeling to C^α .

With the help of these H^α -detected 4D experiments, 76% of the α correlations were successfully assigned (Suppl. Fig S2). Considering that some N, C' resonances are separated by less than 0.05 ppm in the CON spectrum, e.g. corresponding resonances within $^{73/113}KDAISE^{78/118}$ differ only by 0.03–0.04 ppm, all peaks were not expected to be resolved in spectra with spectral resolutions of 0.11 (^{15}N) and 0.05 (^{13}C) ppm. The H^α , C^α shifts were as ineffective in providing the needed dispersion, the smallest peak separations being comparable, e.g. the $^{76/116}Ile$ alpha peaks differ by only 0.01 and 0.05 ppm in 1H and ^{13}C , respectively, and those of $^{77/117}Ser$ even less (Suppl. Fig. S1).

4D (HACA)CON(CA)NH and (HACA)N(CA)CONH experiments

Analogously to 4D H^α -detected experiments described above, the 4D H^N -detected experiments (Fig. 5), with an isolated proline assignment enhancement, are based on their

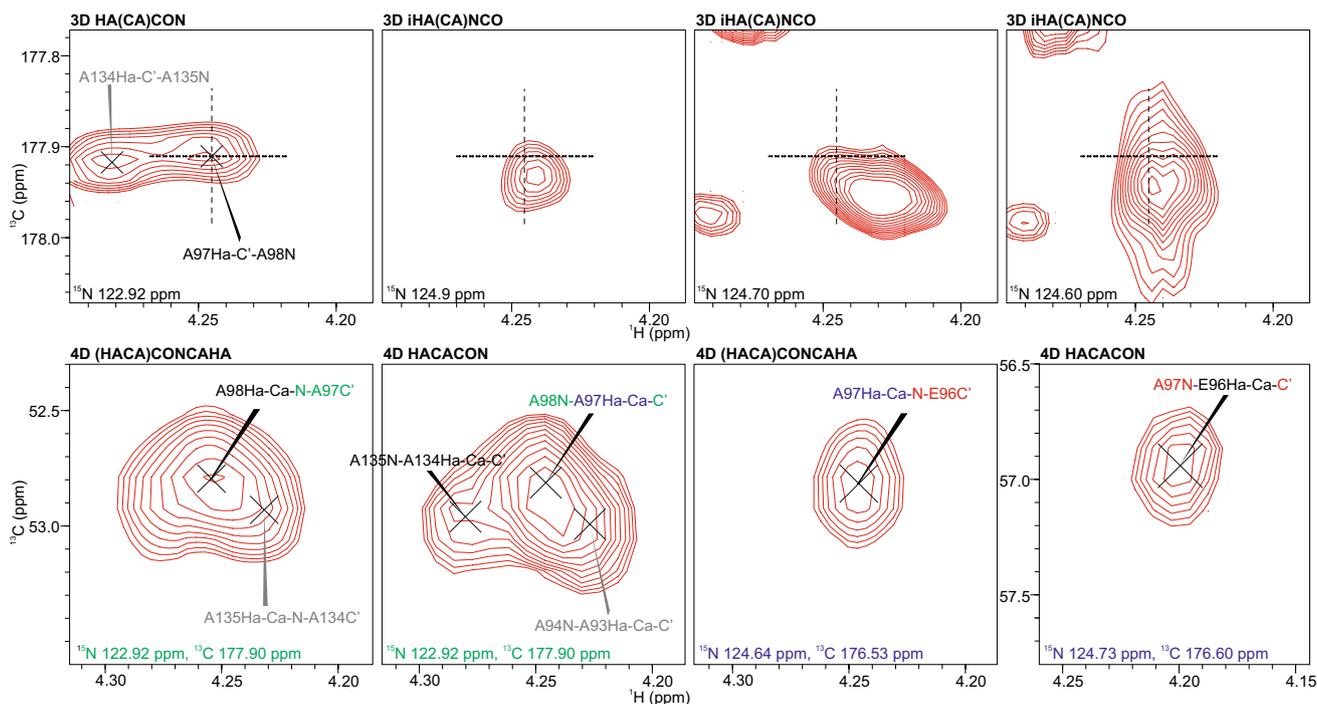
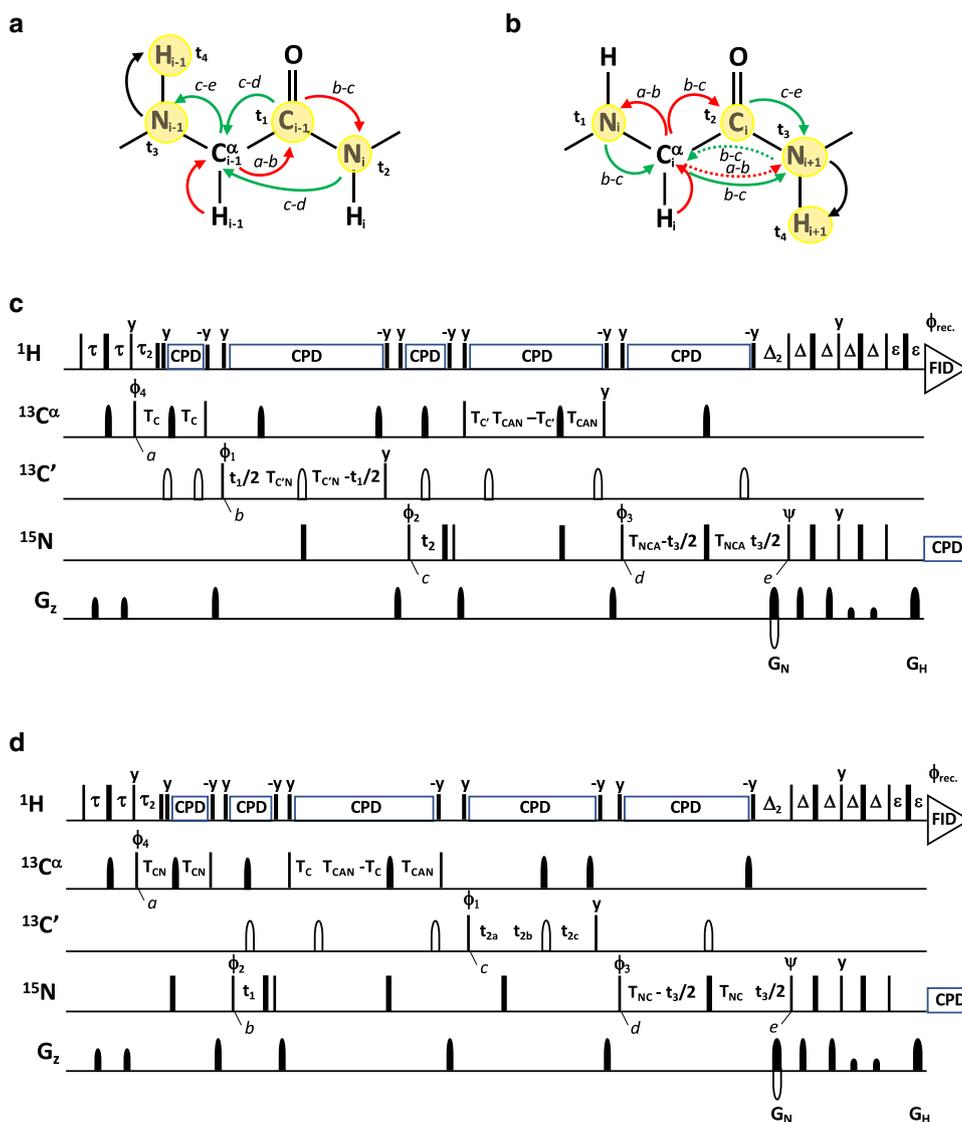


Fig. 4 Problematics in resonance assignment with 3D H^α -detected experiments can be resolved with 4D HACACON and 4D (HACA)CONCAHA experiments. The upper row of 2D planes from 3D HA(CA)CON and iHA(CA)NCO spectra shows that the sequential walk from ^{97}Ala to ^{96}Glu is ambiguous because the ^{97}Ala Ha, C' shift

pair observed in the HA(CA)CON can be found in several planes of the iHA(CA)NCO and is overlapping with other peaks. While there is overlap in the 4D spectra also, the planes are far more easily interpreted and allow for unambiguous assignment of ^{98}Ala - ^{97}Ala - ^{96}Glu . Grey labels mark peaks with maximum in adjacent ^{15}N or ^{13}C planes



3D counterparts (Hellman et al. 2014). Again, sampling of an additional ^{15}N dimension can be implemented without introducing sensitivity loss thanks to the gradient enhanced coherence order selective coherence transfer (COS-CT) (Kay et al. 1992; Schleucher et al. 1994). Magnetization transfer through (HACA)CON(CA)NH and (HACA)N(CA)CONH experiments are briefly described in Eqs. 4 and 5, respectively:

$$\begin{aligned}
 & {}^1\text{H}^\alpha(i-1) \xrightarrow{2\tau({}^1J_{\text{CaHa}})} {}^{13}\text{C}^\alpha(i-1) \xrightarrow{2T_{\text{C}}({}^1J_{\text{CaC}'})} \\
 & {}^{13}\text{C}'(i-1) [2T_{\text{C}'\text{N}} - t_1; {}^1J_{\text{C}'\text{N}}] \rightarrow {}^{15}\text{N}(i) [t_2] \\
 & \rightarrow {}^{13}\text{C}^\alpha(i-1) \xrightarrow{2T_{\text{CAN}}({}^1J_{\text{CaN}}, {}^2J_{\text{CaN}})} {}^{15}\text{N}(i-1) \\
 & [2T_{\text{NCA}} - t_3; {}^1J_{\text{CaN}}, {}^2J_{\text{CaN}}] \xrightarrow{4\Delta({}^1J_{\text{NH}})} {}^1\text{H}^N(i-1) [t_4] \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 & {}^1\text{H}^\alpha(i) \xrightarrow{2\tau({}^1J_{\text{CaHa}})} {}^{13}\text{C}^\alpha(i) \xrightarrow{2T_{\text{CN}}({}^1J_{\text{CaN}}, {}^2J_{\text{CaN}})} {}^{15}\text{N}(i) [t_1] \\
 & \rightarrow {}^{13}\text{C}^\alpha(i) \xrightarrow{2T_{\text{CAN}}, 2T_{\text{C}}({}^1J_{\text{CaN}}, {}^2J_{\text{CaN}}, {}^1J_{\text{CaC}'})} {}^{13}\text{C}'(i) \\
 & [2T_{\text{C}} - t_2; {}^1J_{\text{CaC}'}] \rightarrow {}^{15}\text{N}(i+1) [2T_{\text{NC}} - t_3; {}^1J_{\text{C}'\text{N}}] \\
 & \xrightarrow{4\Delta({}^1J_{\text{NH}})} {}^1\text{H}^N(i+1) [t_4] \quad (5)
 \end{aligned}$$

Like in the H^α -detected experiments (vide supra), the magnetization is first transferred from ${}^1\text{H}^\alpha$ to ${}^{13}\text{C}^\alpha$ spin (time point *a*), and further to either ${}^{13}\text{C}'(i-1)$ spin in (HACA)CON(CA)NH (density operator $\text{C}_z^\alpha(i-1)\text{C}'_y(i-1)$) or selectively to ${}^{15}\text{N}(i)$ spin in (HACA)N(CA)CONH (density operator $\text{C}_z^\alpha(i)\text{N}_y(i)$) at time point *b*. This is followed by the frequency labeling of ${}^{13}\text{C}'(i-1)$ and ${}^{15}\text{N}(i)$ chemical shifts during t_1 in (HACA)CON(CA)NH and

Fig. 5 **a** and **b** schematic representation of the magnetization transfer pathways during the **c** 4D (HACA)CON(CA)NH and **d** 4D (HACA)N(CA)CONH experiments, respectively. Red arrows highlight direct transfer pathway from $^1\text{H}^\alpha(i)$ to $^{15}\text{N}(i)$ or $^{15}\text{N}(i+1)$ whereas green arrows indicate a nested $\text{C}^\alpha\text{C}'_z\text{N}_z \rightarrow \text{N}_z\text{C}^\alpha$ transfer in **c** or $\text{N}_z\text{C}^\alpha \rightarrow \text{C}^\alpha\text{C}'_z$ transfer in **d**, known as the *intraresidual filter* (Brutscher 2002; Permi 2002). Black arrows stand for the coherence order selective coherence transfer, COS-CT (Kay et al. 1992; Schleucher et al. 1994). Arrows with solid line denote the actual magnetization transfer pathway, whereas an arrow with a broken line displays coherence transfer that is suppressed by the *intraresidual filter*. The 4D (HACA)CON(CA)NH experiment establishes connectivities between $^1\text{H}^\alpha(i-1)$ (t_4), $^{15}\text{N}(i-1)$ (t_3), $^{15}\text{N}(i)$ (t_2), and $^{13}\text{C}'(i-1)$ (t_1) frequencies. The 4D (HACA)N(CA)CONH scheme correlates $^1\text{H}^\alpha(i+1)$ (t_4), $^{15}\text{N}(i+1)$ (t_3), $^{13}\text{C}'(i)$ (t_2) and $^{15}\text{N}(i)$ (t_1) frequencies. These two experiments allow also bridging the gap between XPX motifs, where X is any other residue except proline. Narrow and wide bars correspond to 90° and 180° pulses for ^1H and ^{15}N , respectively. Pulses are applied with phase x unless otherwise marked. All ^{13}C pulses are band-selective shaped pulses, denoted by filled narrow bars (90°) and filled and unfilled half ellipsoids (180°). Unfilled half ellipsoids denote on-resonance 180° pulses. The ^1H , ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^\alpha$ carrier positions are 4.7 (water), 118 (center of ^{15}N spectral region), 174 ppm (center of $^{13}\text{C}'$ spectral region), and 56 ppm (center of $^{13}\text{C}^\alpha$ spectral region). The ^{13}C carrier resides at 174 ppm throughout the experiments. Band-selective 90° and 180° pulses for $^{13}\text{C}'/^{13}\text{C}^\alpha$ have the shape of Q5 and Q3 (Emsley and Bodenhausen 1992) and duration of 240.0 μs and 192.0 μs at 800 MHz, respectively. The Waltz-65 sequence (Zhou et al. 2007) with strength of 4.17 kHz was employed to decouple ^1H spins. The GARP-4 (Shaka et al. 1985) with field strength of 1.14 kHz was used to decouple ^{15}N during acquisition. Delay durations: $\tau_1 = 1/(4J_{\text{HC}}) \sim 1.7$ ms; $\tau_2 = 2.6$ ms (for the scheme c) or 2.2 ms (for the scheme d); $\Delta = 1/(4J_{\text{HN}}) \sim 2.7$ ms; $\Delta_2 = 1/(2J_{\text{HN}}) \sim 5.5$ ms; $\varepsilon = \text{duration of } G_{\text{H}} + \text{field recovery} \sim 1.2$ ms; $2T_{\text{C}} = 1/(6J_{\text{C}^\alpha\text{C}'}) \sim 6.2$ ms or 28 ms; $2T_{\text{C}'} \sim 9.4$ ms; $2T_{\text{C}'\text{N}} \sim 33$ ms; $2T_{\text{NCA}} \sim 29$ ms; The $^{13}\text{C}^\alpha$ ^{15}N transfer delay $2T_{\text{CAN}} \sim 50\text{--}56$ ms to suppress the auto-correlated pathway, or $2T_{\text{CAN}} \sim 25$ ms for observing both sequential and auto-correlated cross-peaks. Maximum t_1 and t_3 is restrained in scheme c, $t_{1,\text{max}} < 2.0 * T_{\text{C}'\text{N}}$, $t_{3,\text{max}} < 2.0 * T_{\text{NCA}}$, and t_3 is restricted in scheme d, $t_{3,\text{max}} < 2.0 * T_{\text{NC}}$. Frequency discrimination in $^{13}\text{C}'$ and ^{15}N dimensions is obtained using the States-TPPI protocol (Marion et al. 1989) applied to ϕ_1 and ϕ_2 . Frequency discrimination in the second, t_3 , ^{15}N dimension is accomplished by the COS-CT implementation (Kay et al. 1992; Schleucher et al. 1994). The echo and antiecho signals in the ^{15}N dimension are collected separately by inverting the sign of the G_{N} gradient pulse together with the inversion of ψ , respectively. Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = x$; $\phi_4 = 4(x), 4(-x)$; $\psi = x$; $\phi_{\text{rec.}} = x, 2(-x), x, -x, 2(x), -x$. Gradient strengths and durations: $G_{\text{N}} = 13$ k G/cm (1 ms), $G_{\text{H}} = 13$ k G/cm (1 ms). The pulse sequence codes and parameter files for Bruker Avance system are available from authors upon request

(HACA)N(CA)CONH experiments, respectively. Next, the desired magnetization is transferred to the sequential ^{15}N spin or intraresidual $^{13}\text{C}'$ spin, described by the density operators $\text{C}^\alpha_z(i-1)\text{C}'_z(i-1)\text{N}_y(i)$ and $\text{C}^\alpha_z(i)\text{C}'_y(i)\text{N}_z(i+1)$ in (HACA)CON(CA)NH and (HACA)N(CA)CONH, respectively (time point *c*). Frequency labeling of $^{15}\text{N}(i)$ and $^{13}\text{C}'(i)$ chemical shift takes place in t_2 . The suppression of sequential ^{15}N pathway in (HACA)N(CA)CONH is accomplished using the intraresidual filtering between time points *b* and *c*, and in (HACA)CON(CA)NH

between time points *c* and *d*. Finally, the chemical shift of $^{15}\text{N}(i-1)$ and $^{15}\text{N}(i+1)$ chemical shifts are recorded for $\text{C}^\alpha_z(i-1)\text{N}_y(i-1)$ and $\text{C}'_y(i)\text{N}_z(i+1)$ coherences in (HACA)CON(CA)NH and (HACA)N(CA)CONH between time points *d–e*.

Hence, the 4D (HACA)CON(CA)NH and (HACA)N(CA)CONH spectra exhibit correlations at the intersection of $\omega_{\text{HN}(i-1)}$, $\omega_{\text{N}(i-1)}$, $\omega_{\text{C}'(i-1)}$, $\omega_{\text{N}(i)}$ and $\omega_{\text{HN}(i+1)}$, $\omega_{\text{N}(i+1)}$, $\omega_{\text{C}'(i)}$, $\omega_{\text{N}(i)}$ frequencies, respectively. As the magnetization transfer cascade originates from the H^α spin, and the transfer route includes sampling of ^{15}N frequencies of prolines as well, the experiments facilitate assignment of isolated proline residues (XP or PX dipeptide stretches, where P is a proline and X stands for any non-proline residue) together with other residues in IDPs with high sensitivity. Analogously to H^α -detected experiments, under identical conditions (vide supra), the coherence transfer efficiency for both prolines and non-prolines in the 4D (HACA)CON(CA)NH experiment reaches 0.19, if $2T_{\text{C}}$ is set to 28 ms (and 0.15 with $2T_{\text{C}} = 6.2$ ms) and in the (HACA)N(CA)CONH experiment $I \sim 0.20$. Hence, both of these experiments are optimal also for the proline assignment in XPX moieties.

In the ^1H , ^{15}N HSQC only 41% of the residues displayed resolved peaks, whereas from the two 4D spectra, accurate amide ^1H , ^{15}N shifts were obtained for 85% of BilRI residues. Still overlapping peaks were found for residues ^{73}Lys - ^{80}Met , ^{113}Lys - ^{120}Met and ^{153}Lys - ^{154}Ala in the $^{71/111/151}\text{EVKDAI/VSEKM}^{80/120/160}$ stretches as well as for residues ^{103}Asp - ^{105}Ala , ^{143}Asp - ^{145}Ala in the middle of the $^{99/139}\text{KDAKDAAADKM}^{109/149}$ stretches. Despite the overlap, with the help of the high peak resolution in the CON, it was possible to walk sequentially the whole amino acid sequence with the exception of the first three residues, which are not part of the native BilRI sequence. These residues are missing also from the ^1H , ^{15}N HSQC, likely due to fast chemical exchange of their amide proton with the solvent protons. Examples of the resolving power of the additional ^1H , ^{15}N HSQC dimension are presented in Fig. 6 and Suppl. Figs. S3, S4. The sequential walk over a proline residue is depicted in Suppl. Fig. S5.

It is interesting to note that throughout the identical stretches $^{71/111}\text{EVKDAIVSEKMDAM}^{83/123}$ and $^{99/139}\text{KDAKDAAADKM}^{109/149}$ the corresponding peaks are resolved in the CON spectrum. Naturally, the chemical shift differences become small in the middle of the repeats, $\sim 0.03\text{--}0.04$ ppm for both C' and N^{H} , but the peaks remain, however, distinct. $^{76/116}\text{Ile}$ $\text{C}^\alpha\text{-H}^\alpha$ is resolved into separate peaks in a high-resolution CT ^1H , ^{13}C -HSQC spectrum (t_1 acquisition time 51 ms) as well. To have a different chemical shift, the corresponding atoms in the stretches should experience a mutually different structural environment. Future studies will uncover whether this arises from nearest-neighbor effects

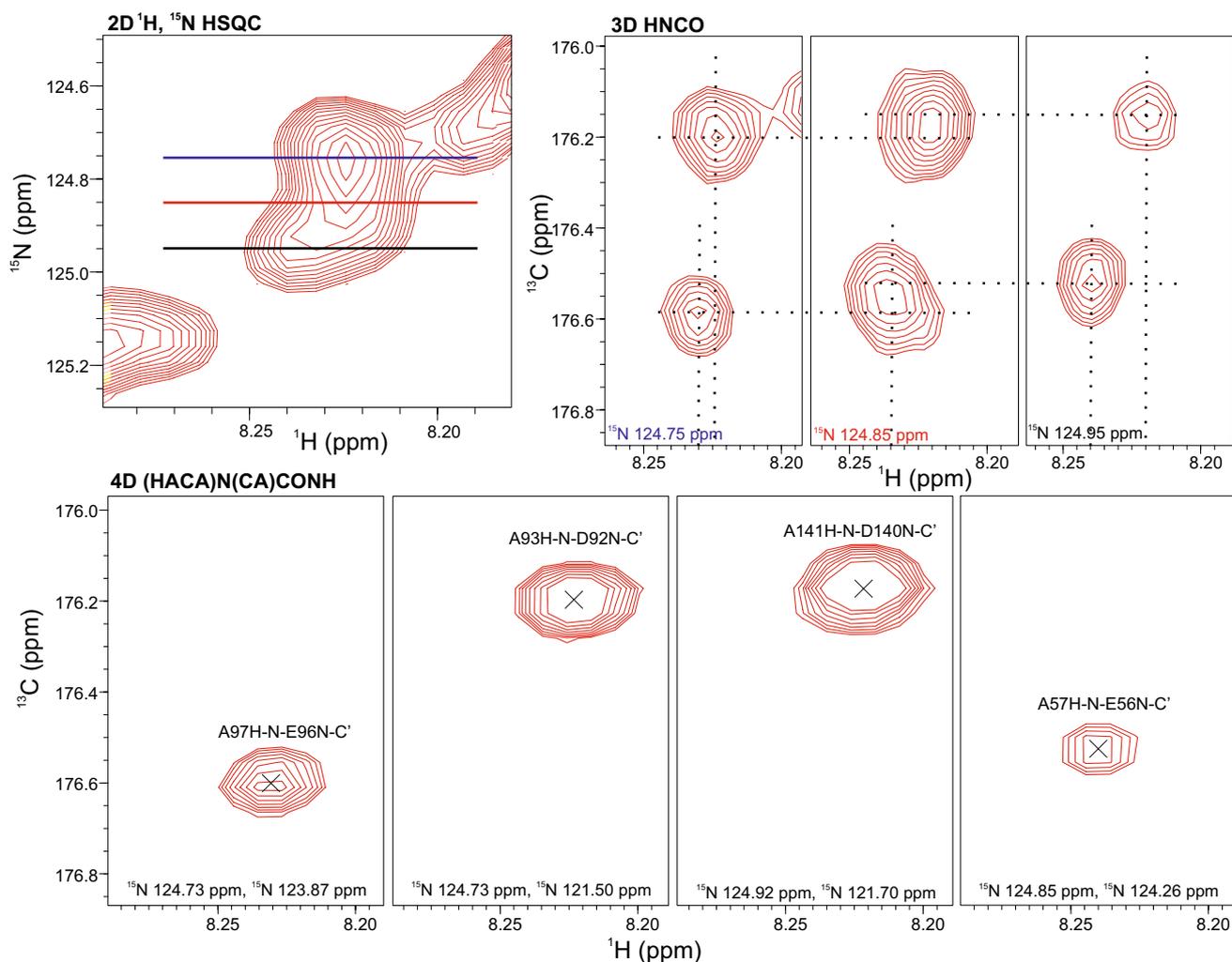


Fig. 6 Resolving power of the $i-1$ amide nitrogen in the fourth dimension. Compared are 3D HNCO and 4D (HACA)N(CA)CONH planes from the ^1H , ^{15}N -HSQC region shown in the upper left corner. From 2D ^1H - $^{13}\text{C}'$ planes retrieved from the 3D HNCO at ^{15}N frequencies marked with lines of different colors it is difficult to determine the number of peaks present and their chemical shifts in this overlapping

peak area of the ^1H , ^{15}N -HSQC. Tentative peak positions are shown with broken lines. In contrast, in the 4D (HACA)N(CA)CONH spectrum four nicely resolved peaks are found. The fourth dimension, i.e. the second chemical shift marked in the 4D planes effectively spreads peaks overlapping in the 2D and 3D spectra

being effective over five residues or from distinct long-range interactions within the protein.

During the past several years, several different NMR experiments and assignment schemes have been developed for studying IDPs. Based on the detected spin, these can be categorized to three different classes i.e. H^{N} -, $^{13}\text{C}'$ - and H^{α} -detection approaches (for review, see e.g., Brutscher et al. 2015). The H^{N} -detected experiments offer the highest sensitivity under acidic conditions and can be combined with the BEST and TROSY implementations (Solyom et al. 2013; Brutscher et al. 2015). However, they show limited performance under alkali conditions and/or with IDPs having high proline content, which can be partially compensated using an H^{α} -start, H^{N} -detected approach as shown here

and also previously demonstrated (Mäntylähti et al. 2009; Hellman et al. 2014; Yoshimura et al. 2015). For instance, Yoshimura et al. (2015) proposed $^{13}\text{C}'$ - $^{13}\text{C}'$ TOCSY transfer in HNCOCNH-type experiments which show connectivities for i , $i \pm 1$, $i \pm 2$ residues i.e. enabling connection of i and $i \pm 2$ residues over a single proline at $i \pm 1$ position. The (HACA)CON(CA)NH and (HACA)N(CA)CONH experiments proposed here link the residues flanking a single proline through its ^{15}N chemical shift.

Unlike H^{N} -detected experiments, the ^{13}C -detection is almost non-susceptible to amide proton exchange with water at high pH. The ^{13}C -detection based experiments for the assignment of IDPs generally apply the 'CON' strategy for connecting the neighboring residues in low-complexity

regions of IDPs, which is also suitable for the assignment of prolines (Pantoja-Uceda and Santoro 2014; Sahu et al. 2014; Brutscher et al. 2015; Chaves-Arquero et al. 2018). Theoretical coherence transfer efficiency in the ^{13}C -detected experiment, hacaCONcaNCO (Pantoja-Uceda et al. 2014) attains 0.19, which is comparable to the sensitivity of the (HACA)CON(CA)NH experiment proposed here. However, given that $\gamma_{\text{H}}/\gamma_{\text{C}}$ equals 4 and the attainable sensitivity has $\gamma^{3/2}$ dependence on the detected spin, the ^{13}C -detection results in an inherent sensitivity loss by a factor of 8 with respect to the ^1H -detection. This can partially be redeemed using a ^{13}C -detection optimized probehead and $^{13}\text{C}'$ -detection, which offers higher resolution in the direct detection dimension due to absence of homonuclear couplings. If such a probehead with an inner coil for ^{13}C is not available, this sets limit to the feasible sample concentration (> 0.3–0.5 mM) which might become an issue with some IDPs. On the other hand, homonuclear coupling between H^{N} and H^{α} can be removed in ^1H -detected experiments using the homonuclear BASH decoupling scheme (Ying et al. 2014).

Finally, the H^{α} -detected experiments offer higher sensitivity with respect to the ^{13}C -detection, enable assignment of consecutive prolines and are not prone to linebroadening due to chemical exchange with the solvent (Mäntylähti et al. 2010, 2011). However, given that H^{α} resonances may partially be underneath the water signal unless sample is prepared in 99.99% D_2O , all experiments proposed here employ the sensitivity-enhanced, echo/antiecho gradient selection in conjunction with the water flip back approach. This enabled us to carry out all the measurements in 95% H_2O and accomplish a nearly complete assignment of H^{α} resonances of the 165-residue BilRI. When working with more dilute samples (< 0.25 mM), dissolving protein into D_2O might be advantageous.

In all, with the proposed suite of 4D experiments, a complete assignment of backbone H^{N} , N^{H} , C' chemical shifts was obtained for all but the three N-terminal residues, which are not part of the BilRI native sequence. H^{α} , C^{α} and C^{β} chemical shifts were obtained for 87, 76, and 66% of the residues, respectively. These experiments offer alternative approach for the assignment of IDPs. Hence, they should not be considered superior per se to the existing methods, but as demonstrated here, they have features that may facilitate accomplishing the assignment of challenging IDPs such as BilRI.

Conclusions

In this paper, we have presented a suite of H^{α} -detected and H^{α} -start, H^{N} -detected 4D experiments that by exploiting chemical shift dispersion brought by an additional C^{α} or N dimension, facilitate the unambiguous backbone resonance

assignment of IDPs in comparison to their established 3D counterparts. The H^{α} -detected experiments provide versatile strategy for the assignment of proline-rich amino acid sequences and/or proteins at alkali pH, prone to fast chemical exchange with the solvent. The proposed 4D H^{α} -detected experiments provide information with the most vital chemical shifts i.e. H^{α} , $^{13}\text{C}^{\alpha}$, and $^{13}\text{C}'$ for conformational restraining in IDPs. The H^{α} -start, H^{N} -detected (HACA)CON(CA)NH and (HACA)N(CA)CONH experiments offer an alternative assignment protocol for IDPs, by providing connectivities over three sequential residues with the H^{N} -detection that allows for connecting single proline-containing segments of the amino acid sequence. The feasibility of the proposed experiments was experimentally verified on with BilRI, a 165-residue protein, containing three highly similar 40-residue repeats and 12 K-D-A triplet sequences, which represents an extremely challenging case of resonance assignment of IDPs.

Acknowledgements Open access funding provided by University of Jyväskylä (JYU). This work is supported by the grants from the Academy of Finland (Number 288235 to PP, 265609, 272960 and 303781 to RI).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Åberg CH, Kelk P, Johansson A (2015) *Aggregatibacter actinomycetemcomitans*: virulence of its leukotoxin and association with aggressive periodontitis. *Virulence* 6:188–195. <https://doi.org/10.4161/21505594.2014.982428>
- Ahlstrand T, Tuominen H, Beklen A, Torittu A, Oscarsson J, Sormunen R, Pöllänen MT, Permi P, Ihalin R (2017) A novel intrinsically disordered outer membrane lipoprotein of *Aggregatibacter actinomycetemcomitans* binds various cytokines and plays a role in biofilm response to interleukin-1 β and interleukin-8. *Virulence* 8:115–134. <https://doi.org/10.1080/21505594.2016.1216294>
- Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006a) ^{13}C -detected protonless NMR spectroscopy of proteins in solution. *Prog Nucl Magn Reson Spectrosc* 48:25–45. <https://doi.org/10.1016/j.pnmrs.2005.09.002>
- Bermel W, Bertini I, Felli IC, Kümmerle R, Pierrattelli R (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. <https://doi.org/10.1016/j.jmr.2005.08.011>

- Bermel W, Bertini I, Felli IC, Pierattelli R (2009) Speeding up (13)C direct detection biomolecular NMR spectroscopy. *J Am Chem Soc* 131:15339–15345. <https://doi.org/10.1021/ja9058525>
- Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J (2012) Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 53:293–301. <https://doi.org/10.1007/s10858-012-9639-0>
- Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. *J Magn Reson Ser A* 102:293–301. <https://doi.org/10.1006/jmra.1993.1107>
- Borchers WM, Daughdrill GW (2018) Using NMR chemical shifts to determine residue-specific secondary structure populations for intrinsically disordered proteins. *Meth Enzymol* 611:101–136. <https://doi.org/10.1016/bs.mie.2018.09.011>
- Brutscher B (2002) Intraresidue HNCA and COHNCA experiments for protein backbone resonance assignment. *J Magn Reson* 156:155–159. <https://doi.org/10.1006/jmre.2002.2546>
- Brutscher B, Felli IC, Gil-Caballero S, Hosek T, Kummerle R, Piai A, Pietarelli R, Solyom Z (2015) NMR methods for the study of intrinsically disordered proteins structure, dynamics, and interactions: general overview and practical guidelines. *Adv Exp Med Biol* 870:49–122. https://doi.org/10.1007/978-3-319-20164-1_3
- Chaves-Arquero B, Pantoja-Uceda D, Roque A, Ponte J, Suau P, Jimenez MA (2018) A CON-based NMR assignment strategy for pro-rich intrinsically disordered proteins with low signal dispersion: the C-terminal domain of histone H1.0 as a case study. *J Biomol NMR* 72:139–148. <https://doi.org/10.1007/s10858-018-0213-2>
- Delaglio F, Torchia DA, Bax A (1991) Measurement of ^{15}N - ^{13}C J couplings in staphylococcal nuclease. *J Biomol NMR* 1:439–446. <https://doi.org/10.1007/bf02192865>
- Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59. [https://doi.org/10.1016/S1093-3263\(00\)00138-8](https://doi.org/10.1016/S1093-3263(00)00138-8)
- Dyson HJ (2016) Making sense of intrinsically disordered proteins. *Biophys J* 110:1013–1016. <https://doi.org/10.1016/j.bpj.2016.01.030>
- 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. [https://doi.org/10.1016/0022-2364\(92\)90242-Y](https://doi.org/10.1016/0022-2364(92)90242-Y)
- Fine DH, Kaplan JB, Kachlany SC, Schreiner HC (2006) How we got attached to *Actinobacillus actinomycetemcomitans*: A model for infectious diseases. *Periodontol* 42:114–157. <https://doi.org/10.1111/j.1600-0757.2006.00189.x>
- Fine DH, Patil AG, Velusamy SK (2019) *Aggregatibacter actinomycetemcomitans* (Aa) under the radar: myths and misunderstandings of aa and its role in aggressive periodontitis. *Front Immunol* 10:728. <https://doi.org/10.3389/fimmu.2019.00728>
- Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. *J Biomol NMR* 35:27–37. <https://doi.org/10.1007/s10858-006-0030-x>
- Hellman M, Piirainen H, Jaakola V, 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. <https://doi.org/10.1007/s10858-013-9804-0>
- Hyvärinen K, Mäntylä P, Buhlin K, Paju S, Nieminen MS, Sinisalo J, Pussinen PJ (2012) A common periodontal pathogen has an adverse association with both acute and stable coronary artery disease. *Atherosclerosis* 223:478–484. <https://doi.org/10.1016/j.atherosclerosis.2012.05.021>
- Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663–10665
- Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2013) High-dimensional NMR spectra for structural studies of biomolecules. *ChemPhysChem* 14:3015–3025. <https://doi.org/10.1002/cphc.201300277>
- Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* 241:74–85. <https://doi.org/10.1016/j.jmr.2013.11.011>
- Kozarov EV, Dorn BR, Shelburne CE, Dunn WA Jr, Progulsk-Fox A (2005) Human atherosclerotic plaque contains viable invasive *Actinobacillus actinomycetemcomitans* and *Porphyromonas gingivalis*. *Arterioscler Thromb Vasc Biol* 25:e17–e18. <https://doi.org/10.1161/01.ATV.0000155018.67835.1a>
- Mäntylähti S, Tossavainen H, Hellman M, Permi P (2009) An intra-residual i(HCA)CO(CA)NH experiment for the assignment of main-chain resonances in ^{15}N , ^{13}C labeled proteins. *J Biomol NMR* 45:301–310. <https://doi.org/10.1007/s10858-009-9373-4>
- Mäntylähti S, Aitio O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR* 47:171–181. <https://doi.org/10.1007/s10858-010-9421-0>
- Mäntylähti 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. <https://doi.org/10.1007/s10858-011-9470-z>
- Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR-spectra without phase cycling – application to the study of hydrogen-exchange in proteins. *J Magn Reson* 85:393–399. [https://doi.org/10.1016/0022-2364\(89\)90152-2](https://doi.org/10.1016/0022-2364(89)90152-2)
- Motáčková V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Zídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J Biomol NMR* 48:169–177. <https://doi.org/10.1007/s10858-010-9447-3>
- Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Zídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D ^{13}C -detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* 50:1–11. <https://doi.org/10.1007/s10858-011-9496-2>
- Paino A, Tuominen H, Jääskeläinen M, Alanko J, Nuutila J, Asikainen SE, Pelliniemi LJ, Pöllänen MT, Chen C, Ihalin R (2011) Trimeric form of intracellular ATP synthase subunit β of *Aggregatibacter actinomycetemcomitans* binds human interleukin-1 β . *PLoS ONE* 6:e18929. <https://doi.org/10.1371/journal.pone.0018929>
- Paino A, Lohermaa E, Sormunen R, Tuominen H, Korhonen J, Pöllänen MT, Ihalin R (2012) Interleukin-1 β is internalised by viable *Aggregatibacter actinomycetemcomitans* biofilm and localises to the outer edges of nucleoids. *Cytokine* 60:565–574. <https://doi.org/10.1016/j.cyto.2012.07.024>
- Paino A, Ahlstrand T, Nuutila J, Navickaite I, Lahti M, Tuominen H, Välimaa H, Lamminmäki U, Pöllänen MT, Ihalin R (2013) Identification of a novel bacterial outer membrane interleukin-1B-binding protein from *Aggregatibacter actinomycetemcomitans*. *PLoS ONE* 8:e70509. <https://doi.org/10.1371/journal.pone.0070509>

- Pantoja-Uceda D, Santoro J (2014) New ^{13}C -detected experiments for the assignment of intrinsically disordered proteins. *J Biomol NMR* 59:43–50. <https://doi.org/10.1007/s10858-014-9827-1>
- Permi P (2002) Intraresidual HNCA: An experiment for correlating only intraresidual backbone resonances. *J Biomol NMR* 23:201–209. <https://doi.org/10.1023/A:1019819514298>
- Permi P, Annala A (2004) Coherence transfer in proteins. *Prog Nucl Magn Reson Spectrosc* 44:97–137. <https://doi.org/10.1016/j.pnmrs.2003.12.001>
- Permi P, Hellman M (2012) Alpha proton detection based backbone assignment of intrinsically disordered proteins. *Methods Mol Biol* 895:211–226. https://doi.org/10.1007/978-1-61779-927-3_15
- Sahu D, Bastidas M, Showalter S (2014) Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detect NMR methods. *Anal Biochem* 449:17–25. <https://doi.org/10.1016/j.ab.2013.12.005>
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34:93–202. [https://doi.org/10.1016/S0079-6565\(98\)00025-9](https://doi.org/10.1016/S0079-6565(98)00025-9)
- Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, Sørensen OW, Griesinger C (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* 4:301–306. <https://doi.org/10.1007/BF00175254>
- Shaka AJ, Keeler J (1987) Broadband spin decoupling in isotropic liquids. *Prog Nucl Magn Reson Spectrosc* 19:47–129. [https://doi.org/10.1016/0079-6565\(87\)80008-0](https://doi.org/10.1016/0079-6565(87)80008-0)
- Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson* 64:547–552. [https://doi.org/10.1016/0022-2364\(85\)90122-2](https://doi.org/10.1016/0022-2364(85)90122-2)
- 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. <https://doi.org/10.1007/s10858-013-9715-0>
- Takeuchi H, Hirano T, Whitmore SE, Morisaki I, Amano A, Lamont RJ (2013) The serine phosphatase SerB of *Porphyromonas gingivalis* suppresses IL-8 production by dephosphorylation of NF- κ B RelA/p65. *PLoS Pathog* 9:e1003326. <https://doi.org/10.1371/journal.ppat.1003326>
- Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427. [https://doi.org/10.1002/1097-0134\(2000115\)41:3%3c415:AID-PROT130%3e3.3.CO;2-Z](https://doi.org/10.1002/1097-0134(2000115)41:3%3c415:AID-PROT130%3e3.3.CO;2-Z)
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. <https://doi.org/10.1002/prot.20449>
- Yao J, Dyson HJ, Wright PE (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett* 419:285–289. [https://doi.org/10.1016/s0014-5793\(97\)01474-9](https://doi.org/10.1016/s0014-5793(97)01474-9)
- Ying J, Roche J, Bax A (2014) Homonuclear decoupling for enhancing resolution and sensitivity in NOE and RDC measurements of peptides and proteins. *J Magn Reson* 241:97–102. <https://doi.org/10.1016/j.jmr.2013.11.006>
- Yoshimura Y, Kulminkaya N, Mulder FAA (2015) Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone ^{15}N and $^{13}\text{C}'$ chemical shifts of multiple contiguous residues in highly resolved 3D spectra. *J Biomol NMR* 61:109–121. <https://doi.org/10.1007/s10858-014-9890-7>
- Zhou Z, Kümmerle R, Qiu X, Redwine D, Cong R, Taha A, Baugh D, Winniford B (2007) A new decoupling method for accurate quantification of polyethylene copolymer composition and triad sequence distribution with ^{13}C NMR. *J Magn Reson* 187:225–233. <https://doi.org/10.1016/j.jmr.2007.05.005>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.