

1 **Exclusively heteronuclear NMR experiments for the investigation of intrinsically disordered**
2 **proteins: focusing on proline residues**

3

4 Isabella C. Felli¹, Wolfgang Bermel² and Roberta Pierattelli¹

5

6 ¹CERM Department of Chemistry “Ugo Schiff”, University of Florence, Via Luigi Sacconi 6, 50019 Sesto
7 Fiorentino, Florence, Italy

8 ²Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

9

10 *Corresponce to:*

11 Isabella C. Felli (felli@cerm.unifi.it) and Roberta Pierattelli (roberta.pierattelli@unifi.it)

12

13 **Keywords**

14 ¹³C detection, IDP, NMR

15

16 **Abstract**

17 NMR represents a key spectroscopic technique to contribute to the emerging field of highly flexible,
18 intrinsically disordered proteins (IDPs) or protein regions (IDRs) that lack a stable three-dimensional
19 structure. A set of exclusively heteronuclear NMR experiments tailored for proline residues, highly
20 abundant in IDPs/IDRs, are presented here. They provide a valuable complement to the widely used
21 approach based on amide proton detection, filling the gap introduced by the lack of amide protons in
22 proline residues within polypeptide chains. The novel experiments have very interesting properties for
23 the investigations of IDPs/IDRs of increasing complexity.

24

25

26 Introduction

27 Invisible in X-ray studies of protein crystals, intrinsically disordered regions (IDRs) of complex proteins
28 have been for a long time considered just passive linkers connecting functional globular domains and thus
29 often ignored in structural biology studies. However, in many cases they comprise a significant fraction of
30 the primary sequence of a protein and for this reason they are expected to have a role in protein function
31 (Van Der Lee et al., 2014). The characterization of highly flexible regions of large proteins as well as entire
32 proteins characterized by the lack of a 3D structure, now generally referred to as intrinsically disordered
33 proteins (IDPs), lies well behind that of their folded counterparts and is nowadays pursued by an
34 increasingly large number of studies to fill this knowledge gap. NMR plays a strategic role in this context
35 since it constitutes the major, if not the unique, spectroscopic technique to achieve atomic resolution
36 information on their structural and dynamic properties. However, intrinsic disorder and high flexibility
37 have very relevant effects for NMR investigations such as reduction of chemical shift dispersion as well as
38 efficient exchange processes with the solvent due to the open conformations that, when approaching
39 physiological pH and temperature, broaden amide proton resonances beyond detection. While several
40 elegant experiments were proposed to exploit exchange processes with the solvent (Kurzbach et al., 2017;
41 Olsen et al., 2020; Szekely et al., 2018; Thakur et al., 2013), in general initial NMR investigations of
42 IDPs/IDRs are carried out in conditions in which these critical points are mitigated. Exchange broadening
43 strongly depends on pH and temperature; conditions can be optimized to recover most of the amide
44 proton resonances enabling the acquisition of amide proton detected triple resonance experiments
45 needed for sequence-specific assignment of the resonances. However, in particular for proteins that are
46 largely exposed to the solvent it may be interesting to study them near physiological pH and temperature
47 conditions (Gil et al., 2013). In this context, ^{13}C direct detection NMR developed into a valuable alternative.
48 Although the intrinsic sensitivity of ^{13}C is lower with respect to that of ^1H , ^{13}C nuclear spins are
49 characterized by a large chemical shift dispersion (Dyson, H. Jane; Wright, 2001) and, when coupled to ^{15}N
50 nuclei, provide a well-defined fingerprint of a polypeptide (Bermel et al., 2006a; Hsu et al., 2009; Lopez et
51 al., 2016; Schiavina et al., 2019). These features were exploited to design a suite of 3D experiments based
52 on carbonyl-carbon direct detection for sequential assignment and to measure NMR observables (Felli
53 and Pierattelli, 2014). These experiments, **starting from ^1H polarization**, exploit only heteronuclear
54 chemical shifts in the indirect dimensions to maximize chemical shift dispersion (exclusively heteronuclear
55 experiments) and can be used to study IDPs/IDRs also in conditions in which amide proton resonances are
56 too broad to be detected. In addition, they reveal information about proline residues that lack the amide

57 proton when part of polypeptide chains and cannot be detected in 2D HN correlation experiments even
58 if pH and temperature conditions are optimized to reduce exchange broadening.

59 Proline residues are abundant in IDPs/IDRs and often occur in proline-rich sequences with repetitive units
60 (Theillet et al., 2014). Initial bioinformatics studies on the relative abundance of each amino acid in regions
61 of the protein that could not be observed in X-ray diffraction studies led to the classification of prolines
62 as “disorder promoting” amino acids (Dunker et al., 2008). Nevertheless proline, the only imino acid,
63 features a closed ring in its side chain which confers local rigidity compared to all other amino acids
64 (Williamson, 1994), as also exploited in FRET studies in which proline residues are used as rigid spacers to
65 measure distances (Schuler et al., 2005). These observations clearly show the importance of experimental
66 atomic resolution information on the structural and dynamic properties of proline residues to understand
67 their role in modulating protein function. While abundant information about proline residues in globular
68 protein folds is available either through NMR or X-ray studies (MacArthur and Thornton, 1991), including
69 several examples of cis-trans isomerization of peptide bonds involving proline nitrogen as molecular
70 switches (Lu et al., 2007), their characterization in highly flexible and disordered polypeptides is available
71 only in a handful of cases (Chaves-Arquero et al., 2018; Gibbs et al., 2017; Haba et al., 2013; Hošek et al.,
72 2016; Knoblich et al., 2009; Pérez et al., 2009; Piai et al., 2016)(Chhabra et al., 2018)(Ahuja et al., 2016)
73 and actually early studies on IDPs/IDRs routinely reported assignment statistics only considering all other
74 amino acids (“excluding prolines”).

75 Here we would like to propose an experimental variant of the most widely used ^{13}C detected 3D
76 experiments for sequence-specific assignment of IDPs/IDRs to selectively pick up correlations involving
77 proline nitrogen nuclei and provide key complementary information to that obtained through amide
78 proton detected experiments. They can be collected in a shorter time with respect to standard 3D
79 experiments and provide a valuable addition to the current experimental protocols for the study of IDPs.

80

81 **Materials and methods**

82 Isotopically labelled α -synuclein (^{13}C and ^{15}N) was expressed and purified as previously described (Huang
83 et al., 2005). The NMR sample has 0.6 mM protein concentration in 20 mM phosphate buffer at pH 6.5
84 and 100 mM NaCl in H_2O with 5% D_2O for the lock signal.

85 Isotopically labelled CBP-ID4 (^{13}C and ^{15}N) was expressed and purified as previously described (Piai et al.,
86 2016). The NMR sample has 0.9 mM protein concentration in water buffer containing 20 mM TRIS, 50 mM
87 KCl, at pH 6.9, with 5% D_2O added for the lock signal.

88 NMR experiments were acquired at 288 K (for α -synuclein) and at 283K (for CBP-ID4) with a 16.4 T Bruker
89 AVANCE NEO spectrometer operating at 700.06 MHz ^1H , 176.05 MHz ^{13}C , and 70.97 MHz ^{15}N frequencies,
90 equipped with a 5 mm cryogenically cooled probehead optimized for ^{13}C direct detection (TXO). RF pulses
91 and carrier frequencies typically employed for the investigation of intrinsically disordered proteins were
92 used, except for the modifications introduced to zoom into the proline ^{15}N region. Carrier frequencies
93 were set to 4.7 ppm (^1H), 176.4 ($^{13}\text{C}'$), 53.9 ($^{13}\text{C}^\alpha$), 44.9 ($^{13}\text{C}^{\text{ali}}$); the ^{15}N carrier was set to 137 ppm, in the
94 center of ^{15}N resonances of proline residues. Hard pulses were used for ^1H . Band selective ^{13}C pulses used
95 were Q5 and Q3 (Emsley and Bodenhausen, 1990) of 300 μs and 231 μs for 90° and 180° rotations
96 respectively; a 900 μs Q3 pulse centered at 53.9 ppm was used for selective inversion of C^α . The ^{15}N pulse
97 to invert the ^{15}N proline resonances was a 8000 μs Reburp pulse (Geen and Freeman, 1991); all other ^{15}N
98 pulses were hard pulses. Decoupling was achieved with waltz65 (100 μs , 2.5 kHz) (Zhou et al., 2007) for
99 ^1H and with garp4 (250 μs , 1.0 kHz) (Shaka, A. J.; Barker, P. B.; Freeman, 1985) for ^{15}N . The MOCCA mixing
100 time (Felli et al., 2009; Furrer et al., 2004) in the (HCA)COCON^{Pro} experiment was 350 ms, constituted by
101 repeated $(\Delta-180^\circ-\Delta)_{2n}$ units in which $\Delta=150 \mu\text{s}$ and the 180° pulse was 91.6 μs).

102 The experimental parameters used for the acquisition of the various experiments on α -synuclein and CBP-
103 ID4 are reported in Table 1. Spectra were calibrated using DSS as a reference for ^1H and ^{13}C ; ^{15}N was
104 calibrated indirectly (Markley et al., 1998).

105 **Table 1.** Experimental parameters used.

106

Experiments α -synuclein	Dimension of acquired data			Spectral width (ppm)			NS ^a	d ₁ (s) ^b
	t1	t2	t3	F1	F2	F3		
¹H detected								
¹ H- ¹⁵ N HSQC	800 (¹⁵ N)	2048 (¹ H)		28.1	15.0		2	1.0
¹³C detected								
CON	512 (¹⁵ N)	1024 (¹³ C)		32.0	31.0		2	1.6
CON ^{Pro}	128 (¹⁵ N)	1024 (¹³ C)		5.0	31.0		2	1.6
(H)CBCACON ^{Pro}	128 (¹³ C)	32 (¹⁵ N)	1024 (¹³ C)	60.0	5.0	30.0	4	1.0
(H)CCCON ^{Pro}	128 (¹³ C)	32 (¹⁵ N)	1024 (¹³ C)	70.0	5.0	30.0	4	1.0
(H)CBCANCO ^{Pro}	128 (¹³ C)	16 (¹⁵ N)	1024 (¹³ C)	60.0	5.0	30.0	8	1.0
¹H and ¹³C detected (using multiple receivers)								
CON/HN	600 (¹⁵ N)	1024 (¹³ C)		35.0	31.0		2	1.6
	600 (¹⁵ N)	2048 (¹ H)		35.0	15.0		4	
Experiments CBP-ID4								
Experiments CBP-ID4	Dimension of acquired data			Spectral width (ppm)			NS ^a	d ₁ (s) ^b
	t1	t2	t3	F1	F2	F3		
¹H detected								
¹ H- ¹⁵ N HSQC	800 (¹⁵ N)	2048 (¹ H)		30.0	15.0		2	1.0
¹³C detected								
CON	1024 (¹⁵ N)	1024 (¹³ C)		38.0	30.0		2	2.0
CON ^{Pro}	170 (¹⁵ N)	1024 (¹³ C)		6.5	30.0		2	2.0
(H)CBCACON ^{Pro}	128 (¹³ C)	64 (¹⁵ N)	1024 (¹³ C)	64.5	6.5	30.0	4	1.0
(H)CCCON ^{Pro}	128 (¹³ C)	64 (¹⁵ N)	1024 (¹³ C)	75.7	6.5	30.0	4	1.0
(H)CBCANCO ^{Pro}	128 (¹³ C)	22 (¹⁵ N)	1024 (¹³ C)	64.5	6.5	30.0	16	1.0
(HCA)COCON ^{Pro}	96 (¹³ C)	64 (¹⁵ N)	1024 (¹³ C)	10.8	6.5	30.0	8	1.5
^a number of acquired scans								
^b inter-scan delay								

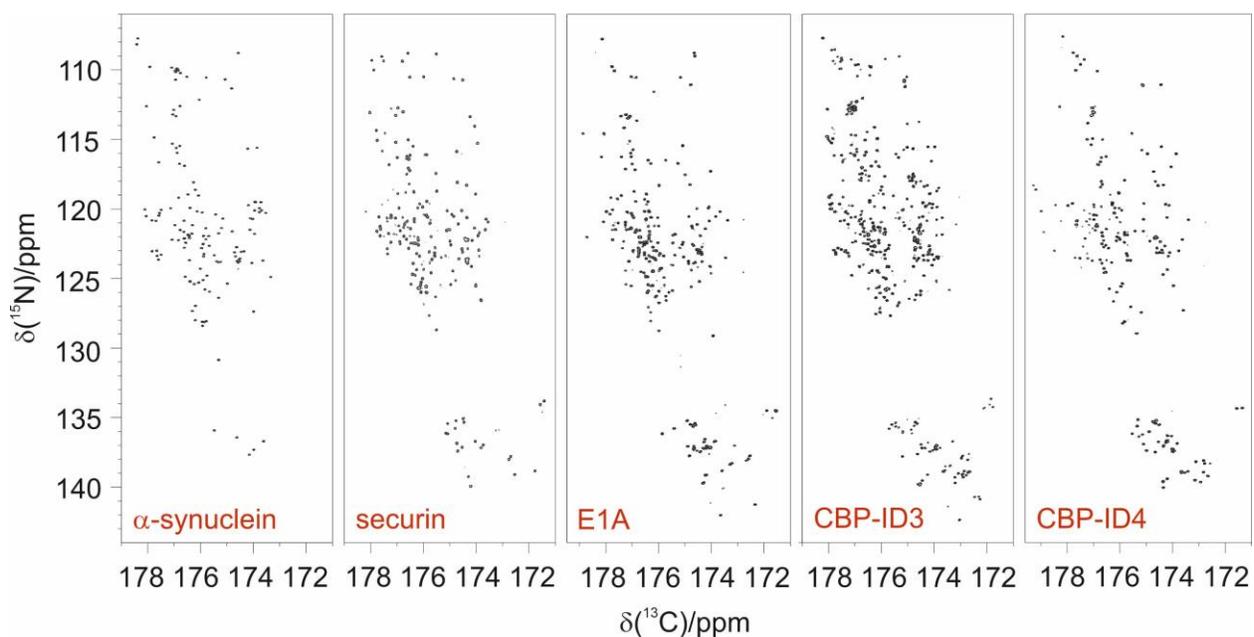
107

108 **Results and discussion**

109 *Advantages of focusing on proline residues*

110 In highly flexible and disordered proteins contributions to signals' chemical shifts deriving from the local
111 environment are averaged out leaving mainly those contributions due to the covalent structure of the
112 polypeptide. Chemical shift ranges predicted for ¹⁵N resonances of imino acids such as proline are quite
113 different from those predicted for amino acids, as expected from the different chemical structure. The 2D
114 CON spectra of several disordered proteins of different size and sequence complexity, reported in Figure
115 1, clearly show that proline residues are quite abundant in IDPs/IDRs and that indeed ¹⁵N resonances of
116 proline residues fall in a well isolated spectral region.

117



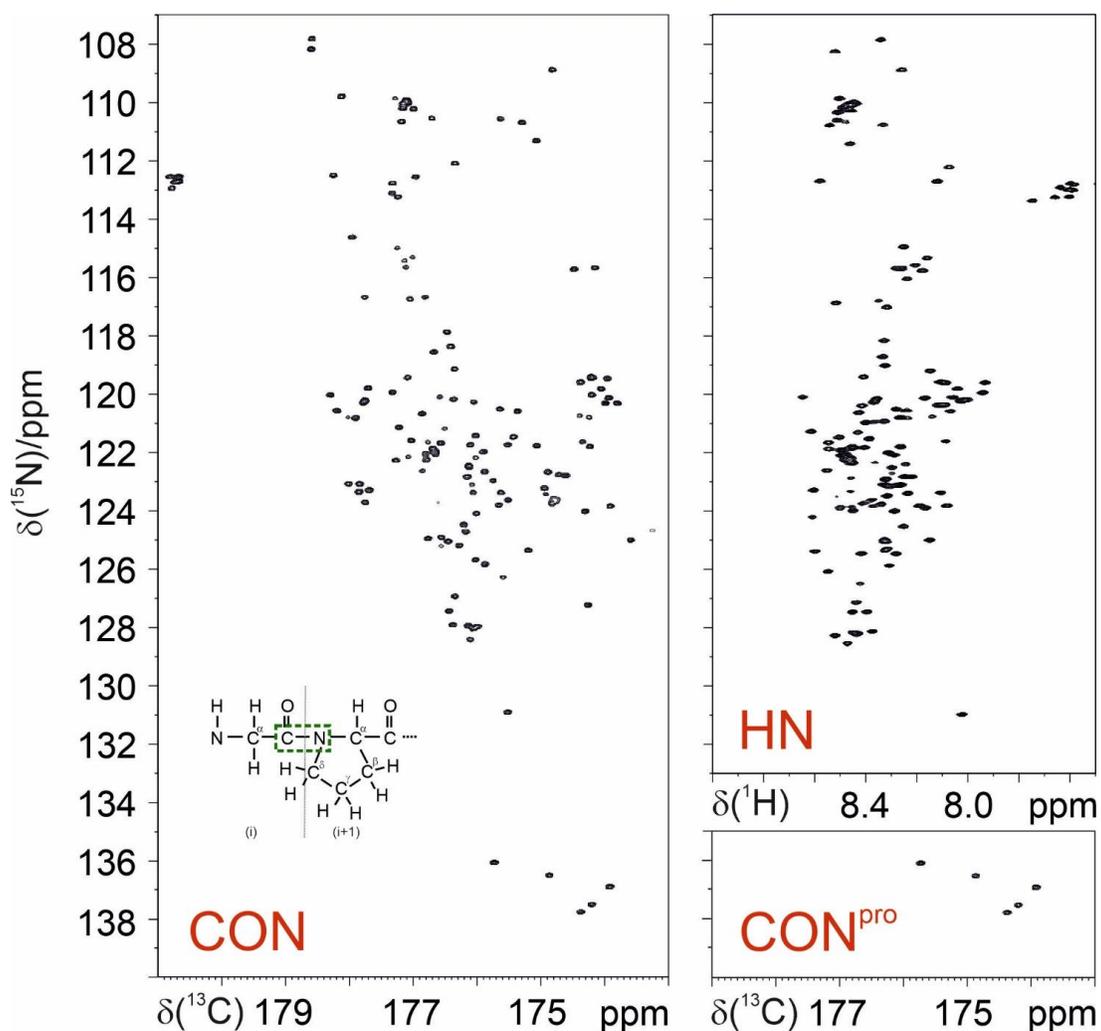
118

119 **Figure 1.** Proline residues are abundant in IDPs/IDRs and their ¹⁵N resonances can be easily detected. They
120 fall in a specific, isolated region of the 2D CON spectrum as illustrated by the examples reported in the
121 figure. From left to right: α -synuclein (140 aa, 4% Pro)(Bermel et al., 2006b); human securin (200 aa, 11%
122 Pro)(Bermel et al., 2009); E1A (243 aa, 16% Pro)(Hošek et al., 2016); CBP-ID3 (407 aa 18% Pro)(Contreras-
123 Martos et al., 2017); CBP-ID4 (207 aa, 22% Pro)(Piai et al., 2016).

124

125 Thus, ^{15}N resonances of proline residues in IDPs/IDRs can be selectively irradiated enabling us to focus on
 126 this spectral region. This can be achieved through the use of band-selective ^{15}N pulses as shown for the
 127 simple case of the CON experiment (Murrall et al., 2018): the selective CON spectrum in the proline region
 128 (CON^{Pro} , Figure 2) provides the complementary information that is missing in 2D HN correlation
 129 experiments, even when pH and temperature are optimized to enhance the detectability of amide protons
 130 (Figure 2).

131



132

133

134 **Figure 2.** Comparison of the 2D CON (left) and 2D HN (right, top) spectra recorded on α -synuclein. The
 135 CON^{Pro} spectrum (right, bottom), reported below the HN panel, clearly illustrates how this experiment
 136 provides the missing information with respect to that available in the HN-detected spectrum. In the inset
 137 of the 2D CON spectrum a scheme of a $\text{Gly}_{i-1}\text{-Pro}_i$ dipeptide highlights the nuclei that give rise to the $\text{C}^1_{i-1}\text{N}_i$
 138 correlations detected in CON spectra (circled in green).

139 The same strategy exploiting band-selective ^{15}N pulses can be used to design experimental variants of
140 triple resonance ^{13}C detected experiments to focus on the ^{15}N proline region and enable us to selectively
141 detect the desired correlations. When implementing this idea into these experiments, such as the 3D
142 (H)CBCACON (Bermel et al., 2009), ^{15}N pulses could all be substituted with band-selective ones. However,
143 instead of substituting all ^{15}N pulses, it is sufficient to introduce a 180° band-selective ^{15}N pulse in one of
144 the $\text{C}'_{i-1}\text{-N}_i$ coherence transfer steps to introduce the desired selectivity in the proline region.

145 As an example the pulse sequence of the 3D (H)CBCACON^{Pro} experiment is shown in Figure 3. The inclusion
146 of the ^{15}N band-selective pulse in the $\text{C}'\text{-N}$ coherence transfer step is used to generate the $\text{C}'_{i-1}\text{-N}_i$ antiphase
147 coherence ($2\text{C}'_y\text{N}_z$) involving the ^{15}N nuclear spin of proline residues (i); for all other amino acid types the
148 evolution of the $\text{C}'_{i-1}\text{-N}_i$ scalar coupling ($^1J_{\text{C}'_{i-1}\text{N}_i}$) is refocused by the 180° band-selective pulse on the
149 carbonyl carbon nuclei only. To achieve the desired selectivity on the ^{15}N proline resonances with respect
150 to those of all other amino acids an 8 ms Reburp pulse (Geen and Freeman, 1991) was used here; this
151 pulse may appear quite long, but it can be accommodated well in the $\text{C}'\text{-N}$ coherence transfer block that
152 requires about 32 ms ($1/2^1J_{\text{C}'_{i-1}\text{N}_i}$). Considering a 8-10 ppm spectral width necessary to cover the ^{15}N
153 proline-region in the indirect dimension (Figure 1), the implementation of this ^{15}N band-selective pulse
154 allows us to reduce the spectral width by a factor of about 4 with respect to that needed to cover the
155 whole spectral region in which backbone ^{15}N nuclear spins resonate, that is about 36-40 ppm. This means
156 that the same resolution can be achieved in a fraction of the time since $1/4$ (or less) of the FIDs should be
157 collected, provided sensitivity is not a limiting factor. Thus, it becomes feasible to acquire spectra with
158 very high resolution, extending the acquisition time in all the indirect dimensions to contrast the reduced
159 chemical shift dispersion typical of IDPs. Non-uniform sampling strategies (Hoch et al., 2014; Kazimierczuk
160 et al., 2010, 2011; Robson et al., 2019) can of course be implemented to reduce acquisition times; also in
161 this case reducing the spectral complexity (the number of cross-peaks is reduced when focusing on proline
162 ^{15}N resonances only) is expected to contribute to reducing experimental times. Out of the full 3D spectrum
163 only a small portion, the one containing the information that is completely missing in amide proton
164 detected experiments, can thus be acquired with the necessary resolution to provide site-specific atomic
165 information.

166 Since C' detected experiments all exploit the $\text{C}'_{i-1}\text{-N}_i$ correlation, which is an inter-residue correlation
167 linking the nitrogen of an amino acid (i) to the carbonyl carbon of the previous one (i-1) across the peptide
168 bond (Figure 2 inset), focusing on proline residues can facilitate the identification of specific $\text{X}_{i-1}\text{-Pro}_i$ pairs
169 through inspection of C^α and C^β chemical shifts of the amino acid preceding proline residues. Such

170 information can be achieved through the 3D (H)CBCACON^{Pro} experiment and can be very useful to identify
171 specific pairs such as Gly/Pro, Ala/Pro, Ser/Pro and Thr/Pro. Acquisition of the 3D (H)CCCON^{Pro}
172 experiment, in parallel to the 3D (H)CBCACON^{Pro}, provides information on aliphatic ¹³C nuclear chemical
173 shifts of the whole side chain. This contributes to narrowing down the possibilities in all cases in which it
174 is not possible to identify the type of amino acid preceding the proline considering **only** their ¹³C^α and ¹³C^β
175 chemical shifts. This is the case for example of Arg/Lys/Gln or Phe/Leu.

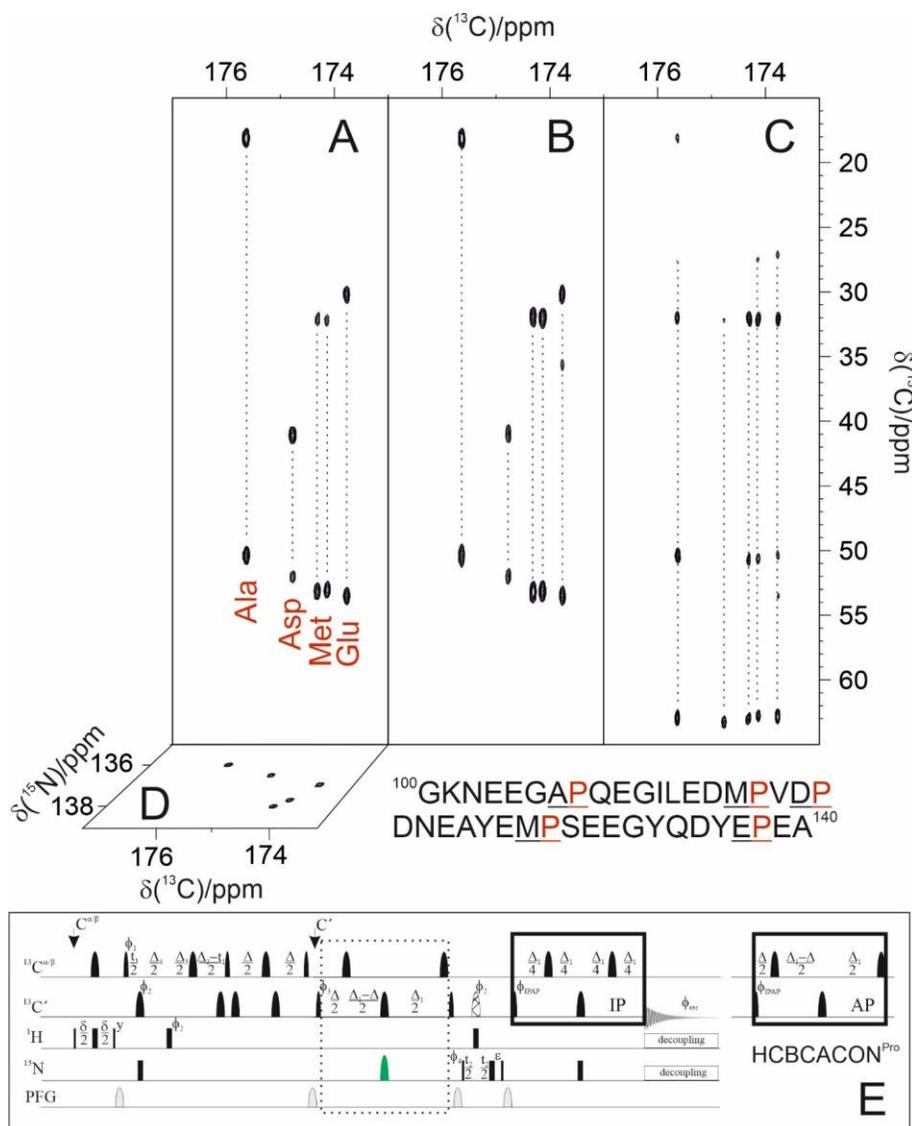
176 Similarly, the insertion of the ¹⁵N band-selective pulse in the proline region in the 3D (H)CBCANCO (Bermel
177 et al., 2006a) enables us to detect the ¹³C resonances of the whole proline ring, providing the
178 complementary information for sequence-specific assignment. The closed proline ring introduces an
179 additional heteronuclear scalar coupling (¹J_{Ni-Cδi}) that also provides the correlations with ¹³C^δ and ¹³C^γ, in
180 parallel to ¹³C^α and ¹³C^β chemical shifts. Indeed the band-selective pulses used for the ¹³C aliphatic region
181 cover also ¹³C^γ and ¹³C^δ resonances (not only ¹³C^α and ¹³C^β ones). In addition analysis of the observed
182 chemical shifts for proline side chains can be correlated to the local conformation, in particular to the
183 cis/trans isomers of the peptide bond involving proline nitrogen nuclei (Schubert et al., 2002; Shen and
184 Bax, 2010). Finally, additional information for sequence-specific assignment can be achieved by exploiting
185 the same approach for the COCON experiment in its ¹³C-start (Bermel et al., 2006b; Felli et al., 2009) as
186 well and in its ¹H-start variants (Mateos et al., 2020).

187

188 *Assignment strategy*

189 To illustrate the approach, experiments were acquired on the well-known IDP α-synuclein. Even if this
190 protein only contains a little number of proline residues (5/140), they are all clustered in a small portion
191 of it (108-138) and thus constitute about 15% of the amino acids in this region. Furthermore, this terminus
192 has a very peculiar amino acidic composition (36% Asp/Glu, 9% Tyr) and it was shown to be the part of
193 the protein that is involved in sensing calcium concentration jumps associated with the transmission of
194 nerve signals (Binolfi et al., 2006; Lautenschläger et al., 2018; Nielsen et al., 2001). Proline residues,
195 embedded in two motifs (DPD and EPE), were shown to facilitate the interaction of carboxylate side chains
196 of Asp and Glu with calcium, even in a flexible and disordered state (Pontoriero et al., 2020).

197 Focusing on the proline ¹⁵N region in case of α-synuclein greatly simplifies the spectral complexity enabling
198 us to illustrate the sequence-specific assignment of the resonances just by visual inspection of the first
199 planes of the 3D spectra described here (Figure 3).



200

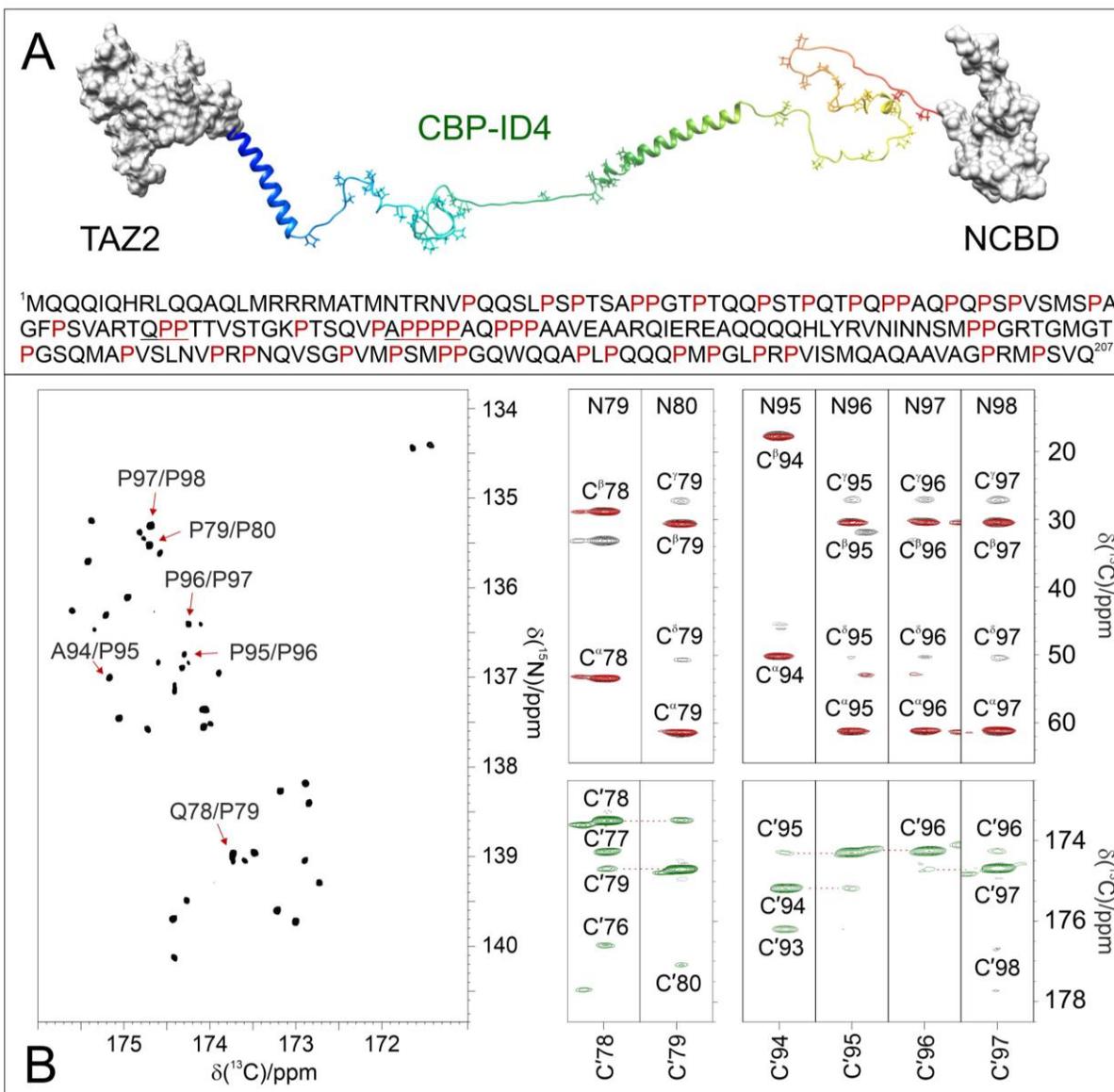
201 **Figure 3.** The implementation of the proposed strategy on α -synuclein renders NMR spectra so
 202 informative that proline resonances can be assigned just by visual inspection of the figure. To this end the
 203 first ^{13}C - ^{13}C planes of the 3D (H)CBCACON^{Pro} (A), 3D (H)CCCON^{Pro} (B), 3D (H)CBCANCO^{Pro} (C) are shown; the
 204 ^{13}C - ^{15}N plane of the 3D (H)CBCACON^{Pro} is also shown (D). The portion of the primary sequence of α -
 205 synuclein hosting its five proline residues is also reported (bottom right). The pulse sequence to acquire
 206 the 3D (H)CBCACON^{Pro} experiment (E) is reported as an example of the implementation of the proposed
 207 approach (dotted box). The delays are: $\epsilon = t_2(0)$, $\delta = 3.6$ ms, $\Delta = 9$ ms, $\Delta_1 = 25$ ms, $\Delta_2 = 8$ ms, $\Delta_3 = \Delta_2 - \Delta_4 =$
 208 5.8 ms, $\Delta_4 = 2.2$ ms. The phase cycle is: $\phi_1 = x, -x$; $\phi_2 = 8(x), 8(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = 2(x), 2(-x)$; $\phi_{\text{IPAP}}(\text{IP}) =$
 209 x ; $\phi_{\text{IPAP}}(\text{AP}) = -y$; $\phi_{\text{rec}} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection was obtained by incrementing phase ϕ_3
 210 (t_1) and ϕ_4 (t_2) in a States-TPPI manner. **The IPAP approach was implemented for homonuclear decoupling**
 211 **in the direct acquisition dimension to suppress the large one bond scalar coupling constants ($^1J_{\text{C}\alpha\text{-C}}$)**(Felli
 212 **and Pierattelli, 2015); alternative approaches can be implemented that exploit band-selective**
 213 **homonuclear decoupling** (Alik et al., 2020; Ying et al., 2014) or processing algorithms that thus only require
 214 **the in-phase spectra** (Karunanithy and Hansen, 2021; Shimba et al., 2003).

215 Indeed, the C'-N projections of the 3D spectra (^{13}C - ^{15}N planes) show that the cross-peaks are well resolved
216 in both dimensions; selection of the ^{15}N proline region enables us to differentiate the signals through the
217 carbonyl carbon chemical shifts of the preceding amino acid. Therefore inspection of the first ^{13}C - ^{13}C plane
218 of the 3D (H)CBCACON^{Pro} experiment (Figure 3A) shows the distinctive $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shift patterns
219 of the residues preceding proline that, by comparison with the primary sequence of the protein, already
220 suggests us the identity of three residue pairs: Ala-Pro, Asp-Pro, Glu-Pro. These can thus be assigned to
221 Ala 107-Pro 108, Asp 119-Pro 120, Glu 138-Pro 139. Comparison with the first ^{13}C - ^{13}C plane of the 3D
222 (H)CCCON^{Pro} (Figure 3B) confirms that an extra cross peak can be detected for the Glu-Pro pair, as
223 expected for amino acids which have a side chain with more than two aliphatic carbon atoms. The
224 remaining signals derive from the two Met-Pro pairs, in agreement with the observed chemical shifts.
225 They can be assigned in a sequence specific manner by comparison of these spectra with the
226 complementary ones based on amide proton detection. The final panel shows the first ^{13}C - ^{13}C plane of
227 the 3D (H)CBCANCO^{Pro} (Figure 3C). This experiment reveals the correlations of the ^{15}N with $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$
228 within each amino acid. In the case of proline, the closed ring introduces additional scalar couplings that
229 are responsible for two additional cross-peaks, the ones of the ^{15}N with $^{13}\text{C}^\gamma$ and $^{13}\text{C}^\delta$, as clearly observed
230 in Figure 3. Chemical shifts show only minor differences between the resonances but still significant to
231 discriminate the different residues provided spectra are acquired with high resolution.

232

233 *A challenging case*

234 A compelling example of complexity is provided by the ID4 flexible linker of CREB-binding protein (CBP), a
235 large transcription co-regulator (Dyson and Wright, 2016). CBP-ID4 connects two well-characterized
236 globular domains (TAZ2, 92 amino acids and NCB2, 59 amino acids)(De Guzman et al., 2000; Kjaergaard
237 et al., 2010) and is constituted by 207 amino acids out of which 45 are proline residues, including several
238 repeated PP motifs (Piai et al., 2016) (Figure 4A). The 2D CON^{Pro} spectrum of CBP-ID4, reported in Figure
239 4B (left panel), shows the C'_{i-1} - N_i correlations of proline residues of ID4. Interestingly, despite the small
240 spectral region, a high number of resolved resonances is observed. The initial count of cross-peaks in this
241 spectrum reveals 42 out of the 45 expected correlations, highlighting the potential of this experimental
242 strategy for the investigation of IDRs/IDPs of increasing complexity. The excellent chemical shift dispersion
243 of the inter-residue C'_{i-1} - N_i correlations is certainly one of the most important aspects to reduce cross peak
244 overlap. Resolution is further enhanced in this region by the narrow linewidths of proline ^{15}N resonances
245 due to the lack of the dipolar contribution of an amide proton to the transverse relaxation.



246

247

248 **Figure 4.** The case of CBP-ID4 (residues 1851–2057 of human CBP). (A) One of the possible conformations
 249 of the CBP-ID4 fragment (blue-to-red ribbons) and the two flanking domains TAZ2 and NCBD clearly
 250 demonstrate that in this region of CBP the intrinsically disordered part is highly prevalent with respect to
 251 ordered ones. The amino acid sequence of CBP-ID4 is also reported. (B) The 2D CON^{Pro} spectrum shown
 252 on the left, which at first sight could seem like a 2D spectrum of a small globular protein, reports the
 253 proline fingerprint of this complex IDR. Several strips extracted from the 3D (H)CBCACON^{Pro} (black
 254 contours), 3D (H)CCCON^{Pro} (red contours), 3D (HCA)COCON^{Pro} (green contours) are shown to illustrate the
 255 information available for sequence specific resonance assignment of this proline-rich fragment of CBP-
 256 ID4.

257 These two features contribute to establishing this spectral region as a key one for the assignment of a
258 large IDR. Indeed, when passing from 2D to 3D experiments, long acquisition times in the ^{15}N dimension
259 are possible and enable us to provide the extra contribution to resolution enhancement needed to focus
260 on complex IDRs and to collect additional information on the proline residues and their neighbouring
261 amino acids. As an example of the quality of the spectra, Figure 4B reports several strips extracted from
262 the pro-selective 3D experiments that were essential for the investigation of a particularly proline-rich
263 region of ID4, the one in between two partially populated α -helices (Piai et al., 2016). This is composed by
264 27 prolines (out of 76 amino acids) which constitute 35% of the amino acids in this region including several
265 proline-rich motifs (PXP, PXXP, PP as well as PPP and PPPP). Figure 4B shows the strips, extracted from
266 the 3D proline-selective experiments, that were used to assign resonances in the two proline-rich regions,
267 78-80 (QPP) and 94-98 (APPP). The strips extracted from the 3D (H)CBCACON^{Pro} and 3D (H)CCCON^{Pro}
268 (Figure 4B, right panels, black and red contours respectively) are very useful to identify the X_{i-1} -Pro_{*i*} pairs
269 that match in this case with a Gln-Pro and an Ala-Pro pair as well as several Pro-Pro ones. The 3D
270 (H)CBCANCO^{Pro} completes the picture providing information about ^{13}C resonances of each proline ring (C^α ,
271 C^β , C^γ , C^δ , not shown for sake of clarity of the figure). However in regions with a high abundance of proline
272 residues additional information is needed for their sequence-specific assignment. To this end the 3D
273 (HCA)COCON^{Pro} (Figure 4B, right panels, green contours) is very useful, as demonstrated for these two
274 proline-rich fragments. This experiment, which includes an isotropic mixing element in the carbonyl region
275 (MOCCA in this case (Felli et al., 2009; Furrer et al., 2004)), enables us to detect correlations of a carbonyl
276 with the neighbouring ones through the small $^3J_{C'C}$ scalar couplings. In case of proline residues the most
277 intense cross-peak is generally observed for the preceding amino acid (C'_{i-2}). However additional peaks
278 are detected also with neighbouring ones and support the sequence-specific assignment process.

279 It is interesting to note, once the sequence-specific assignment becomes available, that C'_{i-1} -N_{*i*} correlations
280 fall in distinctive spectral regions of the CON^{Pro} 2D spectrum, as already pointed out for selected residue
281 pairs such as Gly-Pro, Ser-Pro, Thr-Pro, Val-Pro (Murrall et al., 2018). For example, inspection of Figure 1
282 allows us to identify Gly-Pro pairs in all the CON spectra of different proteins from their characteristic
283 chemical shifts (in the top-right portion of the proline region). An additional contribution towards smaller
284 ^{15}N chemical shifts can also be identified in cases in which more than one proline follows a specific amino
285 acid-type such as for Ala-Pro, Met-Pro and Gln-Pro cross peaks that are shifted to lower ^{15}N chemical shifts
286 when an additional proline follows in the primary sequence. These effects are likely to originate from a
287 combination of effects deriving from the covalent structure (primary sequence in this case) as well as from
288 local conformations. Needless to say that the experimental investigation of these aspects in more detail

289 constitutes an important point to describe the structural and dynamic properties at atomic resolution of
290 the proline-rich parts of highly flexible IDRs. The proposed experiments are thus expected to become of
291 general applicability for studies of IDPs/IDRs in solution.

292 The data generated on proline-rich sequences are of course very relevant to populate **databases such as**
293 **the Biological Magnetic Resonance Data Bank (BMRB, <https://bmr.io/>)** with more information on proline
294 residues in highly flexible protein regions. This in turn will generate more accurate reference data in
295 chemical shift databases to determine local structural propensities through the comparison of
296 experimental shifts with reference ones (Camilloni et al., 2012; Tamiola et al., 2010) improving our
297 understanding of the importance of transient secondary structure elements in determining protein
298 function. On this respect, CBP itself provides another enlightening example with CBP-ID3 (residues 674-
299 1079 of CBP), which features a high number of proline residues (75 out of 406 residues) representing 18%
300 of its primary sequence. The distribution in this case is along the entire sequence but less frequent toward
301 the end, where a β -strand conformation propensity is sampled (Contreras-Martos et al., 2017). Also in
302 this case the distribution of proline residues is important in shaping the conformational space accessible
303 to the polypeptide, facilitating the interaction with protein's partners. Determination of additional
304 observables, such as the $^3J_{CC}$ through the 3D (HCA)COCON^{Pro} experiment or of different ones through
305 modified experimental variants of these experiments are expected to contribute to the characterization
306 of novel motifs in IDRs/IDPs.

307 The experimental strategy proposed here focuses on a remarkably small spectral region which however
308 turns out to be one of the most interesting ones, in particular in the perspective of studying IDPs/IDRs of
309 increasing complexity, somehow reminiscent of other strategies that have been proposed in which only
310 selected residue types are investigated to access information on challenging systems (such as for example
311 the studies of large systems enabled by Methyl-TROSY spectroscopy (Kay, 2011; Schütz and Sprangers,
312 2020). Interestingly the analysis of the NMR spectra presented here enables one to classify the observed
313 cross peaks into residue types, also in absence of sequence-specific assignment. This might provide
314 interesting information for complex IDPs/IDRs in which one is interested in the investigation of the
315 contribution of specific residue types, such as to monitor the occurrence of post-translational
316 modifications or even other phenomena that are more difficult to investigate like liquid-liquid phase
317 separation.

318 **Conclusions and perspectives**

319 Detection and assignment of proline-rich regions of highly flexible intrinsically disordered proteins allows
320 us to have a glimpse on the ways in which proline residues encode specific properties in IDRs/IDPs by
321 simply tuning their distribution along the primary sequence. NMR spectroscopy is particularly well suited
322 for the task, since proline residues have attractive features from the NMR point of view, starting from the
323 peculiar chemical shifts of ^{15}N nuclear spins. In addition, the lack of the attached amide proton implies
324 that one of the major contributions to relaxation of ^{15}N spins is absent and thus proline nitrogen signals
325 have small linewidths. These characteristics make them a very useful starting point for sequential
326 assignment purposes and structure characterization. Furthermore, they provide a set of NMR signals with
327 promising properties to enable high-resolution studies of increasingly large IDPs/IDRs.

328 Several approaches either based on H^{N} or on H^{α} direct detection have been proposed to bypass the
329 problem introduced in sequence-specific assignment by the lack of amide protons typical of proline
330 residues (Hellman et al., 2014; Kanelis et al., 2000; Karjalainen et al., 2020; Löhr et al., 2000; Mäntylähti
331 et al., 2010; Tossavainen et al., 2020; Wong et al., 2018). While these can result useful for systems with
332 moderate complexity (H^{α} detection) or for systems that feature isolated proline residues in the primary
333 sequence ($\text{H}^{\text{N}}/\text{H}^{\alpha}$), they are not as efficient for complex IDRs/IDPs in which high resolution is mandatory
334 and in which often consecutive proline residues are encountered. **Detection of ^{15}N -based experiments can
335 also provide direct observation of proline nuclei (Chhabra et al., 2018) but, despite the excellent resolution
336 achievable with IDPs, these experiments still suffer from sensitivity limitations.**

337 **Concluding, the experiments proposed here** are crucial to assign intrinsically disordered protein regions
338 presenting many repeated motifs including proline residues as well as poly-proline segments **reducing
339 spectral complexity and experimental time without compromise in resolution.** Since NMR probeheads
340 with high ^{13}C sensitivity have become widely available, it is expected that this set of experiments will be
341 applied as an easy-to-use tool also to complement H^{N} -based assignment.

342

343 **Acknowledgments**

344 The support of the CERM/CIRMMP center of Instruct-ERIC is gratefully acknowledged. This work has been
345 funded in part by Fondazione CR Firenze and by the Italian Ministry of University and Research (FOE
346 funding). Maria Grazia Murrari, Letizia Pontoriero and Marco Schiavina are acknowledged for their
347 contribution in the early stages of the project.

348 **References**

- 349 Ahuja, P., Cantrelle, F. X., Huvent, I., Hanouille, X., Lopez, J., Smet, C., Wieruszeski, J. M., Landrieu, I. and
350 Lippens, G.: Proline conformation in a functional Tau fragment, *J. Mol. Biol.*, 428(1), 79–91,
351 doi:10.1016/j.jmb.2015.11.023, 2016.
- 352 Alik, A., Bouguechtouli, C., Julien, M., Bermel, W., Ghouil, R., Zinn-Justin, S. and Theillet, F. X.: Sensitivity-
353 Enhanced ¹³C-NMR spectroscopy for monitoring multisite phosphorylation at physiological temperature
354 and pH, *Angew. Chemie - Int. Ed.*, 1–6, doi:10.1002/anie.202002288, 2020.
- 355 Bermel, W., Bertini, I., Felli, I. C., Kümmerle, R. and Pierattelli, R.: Novel ¹³C direct detection experiments,
356 including extension to the third dimension, to perform the complete assignment of proteins, *J. Magn.*
357 *Reson.*, 178(1), 56–64, doi:10.1016/j.jmr.2005.08.011, 2006a.
- 358 Bermel, W., Bertini, I., Felli, I. C., Lee, Y.-M., Luchinat, C. and Pierattelli, R.: Protonless NMR experiments
359 for sequence-specific assignment of backbone nuclei in unfolded proteins, *J. Am. Chem. Soc.*, 128(12),
360 3918–3919, doi:10.1021/ja0582206, 2006b.
- 361 Bermel, W., Bertini, I., Csizmok, V., Felli, I. C., Pierattelli, R. and Tompa, P.: H-start for exclusively
362 heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins, *J. Magn. Reson.*, 198(2),
363 275–281, doi:10.1016/j.jmr.2009.02.012, 2009.
- 364 Binolfi, A., Rasia, R. M., Bertoncini, C. W., Ceolin, M., Zweckstetter, M., Griesinger, C., Jovin, T. M. and
365 Fernández, C. O.: Interaction of α -synuclein with divalent metal ions reveals key differences: A link
366 between structure, binding specificity and fibrillation enhancement, *J. Am. Chem. Soc.*, 128(30), 9893–
367 9901, doi:10.1021/ja0618649, 2006.
- 368 Camilloni, C., De Simone, A., Vranken, W. F. and Vendruscolo, M.: Determination of secondary structure
369 populations in disordered states of proteins using nuclear magnetic resonance chemical shifts,
370 *Biochemistry*, 51(11), 2224–2231, doi:10.1021/bi3001825, 2012.
- 371 Chaves-Arquero, B., Pantoja-Uceda, D., Roque, A., Ponte, I., Suau, P. and Jiménez, M. A.: A CON-based
372 NMR assignment strategy for pro-rich intrinsically disordered proteins with low signal dispersion: the C-
373 terminal domain of histone H1.0 as a case study, *J. Biomol. NMR*, 72(3–4), 139–148,
374 doi:10.1007/s10858-018-0213-2, 2018.
- 375 Chhabra, S., Fischer, P., Takeuchi, K., Dubey, A., Ziarek, J. J., Boeszoermyeni, A., Mathieu, D., Bermel, W.,
376 Davey, N. E., Wagner, G. and Arthanari, H.: ¹⁵N detection harnesses the slow relaxation property of
377 nitrogen: Delivering enhanced resolution for intrinsically disordered proteins, *Proc. Natl. Acad. Sci.*,
378 115(8), E1710–E1719, doi:10.1073/pnas.1717560115, 2018.
- 379 Contreras-Martos, S., Piai, A., Kosol, S., Varadi, M., Bekesi, A., Lebrun, P., Volkov, A. N., Gevaert, K.,
380 Pierattelli, R., Felli, I. C. and Tompa, P.: Linking functions: An additional role for an intrinsically
381 disordered linker domain in the transcriptional coactivator CBP, *Sci. Rep.*, 7(1), 4676,
382 doi:10.1038/s41598-017-04611-x, 2017.
- 383 Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z. and
384 Uversky, V. N.: The unfoldomics decade: An update on intrinsically disordered proteins, *BMC Genomics*,
385 9(SUPPL. 2), 1–26, doi:10.1186/1471-2164-9-S2-S1, 2008.
- 386 Dyson, H. Jane; Wright, P.: Nuclear magnetic resonance methods for elucidation of structure and
387 dynamics in disordered states, *Methods Enzymol.*, 339, 258–270, 2001.

388 Dyson, H. J. and Wright, P. E.: Role of intrinsic protein disorder in the function and interactions of the
389 transcriptional coactivators CREB-binding Protein (CBP) and p300, *J. Biol. Chem.*, 291(13), 6714–6722,
390 doi:10.1074/jbc.R115.692020, 2016.

391 Emsley, L. and Bodenhausen, G.: Gaussian pulse cascades: New analytical functions for rectangular
392 selective inversion and in-phase excitation in NMR, *Chem. Phys. Lett.*, 165(6), 469–476,
393 doi:10.1016/0009-2614(90)87025-M, 1990.

394 Felli, I. C. and Pierattelli, R.: Novel methods based on ¹³C detection to study intrinsically disordered
395 proteins, *J. Magn. Reson.*, 241(1), 115–125, doi:10.1016/j.jmr.2013.10.020, 2014.

396 Felli, I. C. and Pierattelli, R.: Spin-state-selective methods in solution- and solid-state biomolecular ¹³C
397 NMR, *Prog. Nucl. Magn. Reson. Spectrosc.*, 84–85, 1–13, doi:10.1016/j.pnmrs.2014.10.001, 2015.

398 Felli, I. C., Pierattelli, R., Glaser, S. J. and Luy, B.: Relaxation-optimised Hartmann-Hahn transfer using a
399 specifically Tailored MOCCA-XY16 mixing sequence for carbonyl-carbonyl correlation spectroscopy in ¹³C
400 direct detection NMR experiments, *J. Biomol. NMR*, 43(3), doi:10.1007/s10858-009-9302-6, 2009.

401 Furrer, J., Kramer, F., Marino, J. P., Glaser, S. J. and Luy, B.: Homonuclear Hartmann-Hahn transfer with
402 reduced relaxation losses by use of the MOCCA-XY16 multiple pulse sequence, *J. Magn. Reson.*, 166(1),
403 39–46, doi:10.1016/j.jmr.2003.09.013, 2004.

404 Geen, H. and Freeman, R.: Band-selective radiofrequency pulses, *J. Magn. Reson.*, 93(1), 93–141,
405 doi:10.1016/0022-2364(91)90034-Q, 1991.

406 Gibbs, E. B., Lu, F., Portz, B., Fisher, M. J., Medellin, B. P., Laremore, T. N., Zhang, Y. J., Gilmour, D. S. and
407 Showalter, S. A.: Phosphorylation induces sequence-specific conformational switches in the RNA
408 polymerase II C-terminal domain, *Nat. Commun.*, 8(May), 1–11, doi:10.1038/ncomms15233, 2017.

409 Gil, S., Hošek, T., Solyom, Z., Kümmerle, R., Brutscher, B., Pierattelli, R. and Felli, I. C.: NMR spectroscopic
410 studies of intrinsically disordered proteins at near-physiological conditions, *Angew. Chemie - Int. Ed.*,
411 52(45), 11808–11812, doi:10.1002/anie.201304272, 2013.

412 De Guzman, R. N., Liu, H. Y., Martinez-Yamout, M., Dyson, H. J. and Wright, P. E.: Solution structure of
413 the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP, *J. Mol. Biol.*, 303(2), 243–253,
414 doi:10.1006/jmbi.2000.4141, 2000.

415 Haba, N. Y., Gross, R., Novacek, J., Shaked, H., Zidek, L., Barda-Saad, M. and Chill, J. H.: NMR determines
416 transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein,
417 *Biophys. J.*, 105(2), 481–493, doi:10.1016/j.bpj.2013.05.046, 2013.

418 Hellman, M., Piirainen, H., Jaakola, V. P. and Permi, P.: Bridge over troubled proline: Assignment of
419 intrinsically disordered proteins using (HCA)CON(CAN)H and (HCA)N(CA)CO(N)H experiments
420 concomitantly with HNCO and i(HCA)CO(CA)NH, *J. Biomol. NMR*, 58(1), 49–60, doi:10.1007/s10858-013-
421 9804-0, 2014.

422 Hoch, J. C., Maciejewski, M. W., Mobli, M., Schuyler, A. D. and Stern, A. S.: Nonuniform sampling and
423 maximum entropy reconstruction in multidimensional NMR, *Acc. Chem. Res.*, 47(2), 708–717,
424 doi:10.1021/ar400244v, 2014.

425 Hošek, T., Calçada, E. O., Nogueira, M. O., Salvi, M., Pagani, T. D., Felli, I. C. and Pierattelli, R.: Structural
426 and dynamic characterization of the molecular hub early region 1A (E1A) from human adenovirus,
427 *Chem. - A Eur. J.*, 22(37), doi:10.1002/chem.201602510, 2016.

428 Hsu, S.-T. D., Bertocini, C. W. and Dobson, C. M.: Use of protonless NMR spectroscopy to alleviate the
429 loss of information resulting from exchange-broadening, *J. Am. Chem. Soc.*, 131(21), 7222–7223,
430 doi:10.1021/ja902307q, 2009.

431 Huang, C., Ren, G., Zhou, H. and Wang, C.: A new method for purification of recombinant human α -
432 synuclein in *Escherichia coli*, *Protein Expr. Purif.*, 42(1), 173–177, doi:10.1016/j.pep.2005.02.014, 2005.

433 Kanelis, V., Donaldson, L., Muhandiram, D. R., Rotin, D., Forman-Kay, J. D. and Kay, L. E.: Sequential
434 assignment of proline-rich regions in proteins: Application to modular binding domain complexes, *J.*
435 *Biomol. NMR*, 16(3), 253–259, doi:10.1023/A:1008355012528, 2000.

436 Karjalainen, M., Tossavainen, H., Hellman, M. and Permi, P.: HACANCOi: a new $H\alpha$ -detected experiment
437 for backbone resonance assignment of intrinsically disordered proteins, *J. Biomol. NMR*, 74(12), 741–
438 752, doi:10.1007/s10858-020-00347-5, 2020.

439 Karunanithy, G. and Hansen, D. F.: FID-Net: A versatile deep neural network architecture for NMR
440 spectral reconstruction and virtual decoupling, *J. Biomol. NMR*, 75(4), 179–191, doi:10.1007/s10858-
441 021-00366-w, 2021.

442 Kay, L. E.: Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view, *J.*
443 *Magn. Reson.*, 210(2), 159–170, doi:10.1016/j.jmr.2011.03.008, 2011.

444 Kazimierczuk, K., Stanek, J., Zawadzka-Kazimierczuk, A. and Koźmiński, W.: Random sampling in
445 multidimensional NMR spectroscopy, *Prog. Nucl. Magn. Reson. Spectrosc.*, 57(4), 420–434,
446 doi:10.1016/j.pnmrs.2010.07.002, 2010.

447 Kazimierczuk, K., Misiak, M., Stanek, J., Zawadzka-Kazimierczuk, A. and Koźmiński, W.: Generalized
448 fourier transform for non-uniform sampled data, in *Topics in current chemistry*, vol. 312, pp. 79–124.,
449 2011.

450 Kjaergaard, M., Teilum, K. and Poulsen, F. M.: Conformational selection in the molten globule state of
451 the nuclear coactivator binding domain of CBP, *Proc. Natl. Acad. Sci. U. S. A.*, 107(28), 12535–12540,
452 doi:10.1073/pnas.1001693107, 2010.

453 Knoblich, K., Whittaker, S., Ludwig, C., Michiels, P., Jiang, T., Schaffhausen, B. and Günther, U.: Backbone
454 assignment of the N-terminal polyomavirus large T antigen, *Biomol. NMR Assign.*, 3(1), 119–123,
455 doi:10.1007/s12104-009-9155-7, 2009.

456 Kurzbach, D., Canet, E., Flamm, A. G., Jhajharia, A., Weber, E. M. M., Konrat, R. and Bodenhausen, G.:
457 Investigation of intrinsically disordered proteins through exchange with hyperpolarized water, *Angew.*
458 *Chemie - Int. Ed.*, 56(1), 389–392, doi:10.1002/anie.201608903, 2017.

459 Lautenschläger, J., Stephens, A. D., Fusco, G., Ströhl, F., Curry, N., Zacharopoulou, M., Michel, C. H.,
460 Laine, R., Nespovitaya, N., Fantham, M., Pinotsi, D., Zago, W., Fraser, P., Tandon, A., St George-Hyslop,
461 P., Rees, E., Phillips, J. J., De Simone, A., Kaminski, C. F. and Schierle, G. S. K.: C-terminal calcium binding
462 of α -synuclein modulates synaptic vesicle interaction, *Nat. Commun.*, 9(1), doi:10.1038/s41467-018-
463 03111-4, 2018.

464 Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M.,
465 Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P.,
466 Uversky, V. N., Wright, P. E. and Babu, M. M.: Classification of intrinsically disordered regions and
467 proteins, *Chem. Rev.*, 114(13), 6589–6631, doi:10.1021/cr400525m, 2014.

468 Löhr, F., Pfeiffer, S., Lin, Y. J., Hartleib, J., Klimmek, O. and Rüterjans, H.: HNCAN pulse sequences for
469 sequential backbone resonance assignment across proline residues in perdeuterated proteins, *J. Biomol.*
470 *NMR*, 18(4), 337–346, doi:10.1023/A:1026737732576, 2000.

471 Lopez, J., Schneider, R., Cantrelle, F. X., Huvent, I. and Lippens, G.: Studying intrinsically disordered
472 Proteins under true in vivo conditions by combined cross-polarization and carbonyl-detection NMR
473 Spectroscopy, *Angew. Chemie - Int. Ed.*, 55(26), 7418–7422, doi:10.1002/anie.201601850, 2016.

474 Lu, K. P., Finn, G., Lee, T. H. and Nicholson, L. K.: Prolyl cis-trans isomerization as a molecular timer, *Nat.*
475 *Chem. Biol.*, 3(10), 619–629, doi:10.1038/nchembio.2007.35, 2007.

476 MacArthur, M. W. and Thornton, J. M.: Influence of proline residues on protein conformation, *J. Mol.*
477 *Biol.*, 218(2), 397–412, doi:10.1016/0022-2836(91)90721-H, 1991.

478 Mäntylähti, S., Aitio, O., Hellman, M. and Permi, P.: HA-detected experiments for the backbone
479 assignment of intrinsically disordered proteins, *J. Biomol. NMR*, 47(3), 171–181, doi:10.1007/s10858-
480 010-9421-0, 2010.

481 Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E. and Wüthrich, K.:
482 Recommendations for the presentation of NMR structures of proteins and nucleic acids, *J. Biomol. NMR*,
483 12(1), 1–23, doi:10.1023/A:1008290618449, 1998.

484 Mateos, B., Conrad-Billroth, C., Schiavina, M., Beier, A., Kontaxis, G., Konrat, R., Felli, I. C. and Pierattelli,
485 R.: The ambivalent role of proline residues in an intrinsically disordered protein: From disorder
486 promoters to compaction facilitators, *J. Mol. Biol.*, 432(9), doi:10.1016/j.jmb.2019.11.015, 2020.

487 Murralli, M. G., Piai, A., Bermel, W., Felli, I. C. and Pierattelli, R.: Proline fingerprint in intrinsically
488 disordered proteins, *ChemBioChem*, 19(15), 1625–1629, doi:10.1002/cbic.201800172, 2018.

489 Nielsen, M. S., Vorum, H., Lindersson, E. and Jensen, P. H.: Ca²⁺ Binding to α -synuclein regulates ligand
490 binding and oligomerization, *J. Biol. Chem.*, 276(25), 22680–22684, doi:10.1074/jbc.M101181200, 2001.

491 Olsen, G. L., Szekely, O., Mateos, B., Kadeřávek, P., Ferrage, F., Konrat, R., Pierattelli, R., Felli, I. C.,
492 Bodenhausen, G., Kurzbach, D. and Frydman, L.: Sensitivity-enhanced three-dimensional and carbon-
493 detected two-dimensional NMR of proteins using hyperpolarized water, *J. Biomol. NMR*, 74(2–3), 161–
494 171, doi:10.1007/s10858-020-00301-5, 2020.

495 Pérez, Y., Gairí, M., Pons, M. and Bernadó, P.: Structural characterization of the natively unfolded N-
496 terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain.,
497 *J. Mol. Biol.*, 391(1), 136–148, doi:10.1016/j.jmb.2009.06.018, 2009.

498 Piai, A., Calçada, E. O., Tarenzi, T., Grande, A. Del, Varadi, M., Tompa, P., Felli, I. C. and Pierattelli, R.: Just
499 a flexible linker? the structural and dynamic properties of CBP-ID4 revealed by NMR spectroscopy,
500 *Biophys. J.*, 110(2), 372–381, doi:10.1016/j.bpj.2015.11.3516, 2016.

501 Pontoriero, L., Schiavina, M., Murralli, M. G., Pierattelli, R. and Felli, I. C.: Monitoring the Interaction of α -
502 synuclein with calcium ions through exclusively heteronuclear Nuclear Magnetic Resonance
503 experiments, *Angew. Chemie*, 132(42), 18696–18704, doi:10.1002/ange.202008079, 2020.

504 Robson, S., Arthanari, H., Hyberts, S. G. and Wagner, G.: Nonuniform sampling for NMR spectroscopy, in
505 *Methods Enzymol.*, pp. 263–291., 2019.

506 Schiavina, M., Murralli, M. G., Pontoriero, L., Sainati, V., Kümmerle, R., Bermel, W., Pierattelli, R. and

507 Felli, I. C.: Taking simultaneous snapshots of intrinsically disordered proteins in action, *Biophys. J.*,
508 117(1), doi:10.1016/j.bpj.2019.05.017, 2019.

509 Schubert, M., Labudde, D., Oschkinat, H. and Schmieder, P.: A software tool for the prediction of Xaa-Pro
510 peptide bond conformations in proteins based on ^{13}C chemical shift statistics, *J. Biomol. NMR*, 24(2),
511 149–154, doi:10.1023/A:1020997118364, 2002.

512 Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. and Eaton, W. A.: Polyproline and the
513 “spectroscopic ruler” revisited with single-molecule fluorescence., *Proc. Natl. Acad. Sci. U. S. A.*, 102(8),
514 2754–9, doi:10.1073/pnas.0408164102, 2005.

515 Schütz, S. and Sprangers, R.: Methyl TROSY spectroscopy: A versatile NMR approach to study challenging
516 biological systems, *Prog. Nucl. Magn. Reson. Spectrosc.*, 116, 56–84, doi:10.1016/j.pnmrs.2019.09.004,
517 2020.

518 Shaka, A. J.; Barker, P. B.; Freeman, R.: Computer-optimized decoupling scheme for wideband
519 applications and low-level operation, *J. Magn. Reson.*, 64(547), 552, 1985.

520 Shen, Y. and Bax, A.: Prediction of Xaa-Pro peptide bond conformation from sequence and chemical
521 shifts, *J. Biomol. NMR*, 46(3), 199–204, doi:10.1007/s10858-009-9395-y, 2010.

522 Shimba, N., Stern, A. S., Craik, C. S., Hoch, J. C. and Dötsch, V.: Elimination of $^{13}\text{C}\alpha$ splitting in protein
523 NMR spectra by deconvolution with maximum entropy reconstruction, *J. Am. Chem. Soc.*, 125(9), 2382–
524 2383, doi:10.1021/ja027973e, 2003.

525 Szekely, O., Olsen, G. L., Felli, I. C. and Frydman, L.: High-resolution 2D NMR of disordered proteins
526 enhanced by hyperpolarized water, , doi:10.1021/acs.analchem.8b00585, 2018.

527 Tamiola, K., Acar, B. and Mulder, F. A. A.: Sequence-specific random coil chemical shifts of intrinsically
528 disordered proteins, *J. Am. Chem. Soc.*, 132(51), 18000–18003, doi:10.1021/ja105656t, 2010.

529 Thakur, A., Chandra, K., Dubey, A., D’Silva, P. and Atreya, H. S.: Rapid characterization of hydrogen
530 exchange in proteins, *Angew. Chemie - Int. Ed.*, 52(9), 2440–2443, doi:10.1002/anie.201206828, 2013.

531 Theillet, F., Kalmar, L., Tompa, P., Han, K., Selenko, P., Dunker, A. K., Daughdrill, G. W. and Uversky, V. N.:
532 The alphabet of intrinsic disorder, *Intrinsically Disord. Proteins*, 1(April), e24360, doi:10.4161/idp.24360,
533 2014.

534 Tossavainen, H., Salovaara, S., Hellman, M., Ihalin, R. and Permi, P.: Dispersion from $\text{C}\alpha$ or NH: 4D
535 experiments for backbone resonance assignment of intrinsically disordered proteins, *J. Biomol. NMR*,
536 74(2–3), 147–159, doi:10.1007/s10858-020-00299-w, 2020.

537 Williamson, M. P.: The structure and function of proline-rich regions in proteins, *Biochem. J.*, 297(2),
538 249–260, doi:10.1042/bj2970249, 1994.

539 Wong, L. E., Maier, J., Wienands, J., Becker, S. and Griesinger, C.: Sensitivity-enhanced four-dimensional
540 amide–amide correlation NMR experiments for sequential assignment of proline-rich disordered
541 proteins, *J. Am. Chem. Soc.*, 140(10), 3518–3522, doi:10.1021/jacs.8b00215, 2018.

542 Ying, J., Li, F., Lee, J. H. and Bax, A.: $^{13}\text{C}\alpha$ decoupling during direct observation of carbonyl resonances in
543 solution NMR of isotopically enriched proteins, *J. Biomol. NMR*, 60(1), 15–21, doi:10.1007/s10858-014-
544 9853-z, 2014.

545 Zhou, Z., Kümmerle, R., Qiu, X., Redwine, D., Cong, R., Taha, A., Baugh, D. and Winniford, B.: A new

546 decoupling method for accurate quantification of polyethylene copolymer composition and triad
547 sequence distribution with ^{13}C NMR, *J. Magn. Reson.*, 187(2), 225–233, doi:10.1016/j.jmr.2007.05.005,
548 2007.

549