

The Use of ^{13}C Direct-Detect NMR to Characterize Flexible and Disordered Proteins

Erik C. Cook*, Grace A. Usher[†], Scott A. Showalter^{*,†,1}

*Department of Chemistry, The Pennsylvania State University, University Park, PA, United States

[†]Department of Biochemistry and Molecular Biology, Center for Eukaryotic Gene Regulation, The Pennsylvania State University, University Park, PA, United States

¹Corresponding author: e-mail address: sas76@psu.edu

Contents

1. Introduction	82
2. Preparation of ^{15}N - and ^{13}C -Enriched Proteins	83
2.1 Plasmid Construction and Protein Expression	83
2.2 Purification of Pdx1-C	85
3. Resonance Assignment Through ^{13}C Direct-Detect Spectra	86
3.1 2D NMR	87
3.2 3D NMR	89
4. Paramagnetic Relaxation Enhancement	91
4.1 PRE Protocol	92
5. Residual Dipolar Couplings	94
5.1 RDC Protocol	95
6. Spin Relaxation Measurement	96
6.1 Protocol for ^{15}N Relaxation Measurements	97
7. Summary	98
Acknowledgments	98
References	98

Abstract

NMR spectroscopy remains the only experimental technique that provides (near) atomistic structural information for intrinsically disordered proteins (IDPs), but their sequence and structure characteristics still pose major challenges for high-resolution spectroscopy. Carbon-13 direct-detect NMR spectroscopy can overcome poor spectral dispersion and other difficulties associated with traditional ^1H -detected NMR of nonaggregating disordered proteins. This chapter presents spectroscopic protocols suitable for complete characterization of IDPs that rely exclusively on ^{13}C direct-detect experiments. The protocols described span initial characterization and chemical shift assignment; structure constraint through residual dipolar coupling and paramagnetic

relaxation enhancement measurements; and assessment of intramolecular dynamics through ^{15}N spin relaxation. The experiments described empower investigators to establish molecular mechanisms and structure–function relationships for IDPs and other proteins characterized by high internal flexibility.



1. INTRODUCTION

As this volume makes clear, intrinsically disordered proteins (IDPs) are the subject of intensive study, owing both to their critical role in physiological interactions such as in signal transduction or transcription, and to their pathological interactions in neurodegenerative and other diseases (Uversky, 2018). Despite clear interest in understanding the structure–function relationship of IDPs, their very nature—high conformational disorder and frequent strong bias in amino acid composition—poses a major barrier to their comprehensive investigation. However, tackling these challenges has provided novel insights into the structural, thermodynamic, and kinetic properties that adapt IDPs for their known functions (Gibbs & Showalter, 2015). Recently, we reviewed the broad and major breakthroughs that NMR has contributed to this field (Gibbs, Cook, & Showalter, 2017). While using NMR to study IDPs is established as a successful strategy, application of ^{13}C direct-detect experiments, which remain rare in biomolecular NMR, has emerged as an especially potent strategy that merits widespread adoption.

Early structural characterization of folded proteins by NMR relied almost exclusively on proton-detected experiments because the ^1H nucleus possesses a high gyromagnetic ratio. Experiments built on the $^1\text{H}, ^{15}\text{N}$ -HSQC detection platform are generally successful for folded systems, where resonances are well dispersed; in contrast, resonance overlap caused by poor spectral dispersion in the proton dimension is characteristic of IDPs and often precludes quantitative analysis. Motivated by the need to overcome the spectral dispersion problem, ^{13}C direct-detect experiments have been developed for the characterization of IDPs. The excellent spectral quality achieved through ^{13}C direct-detect spectroscopy has allowed us to interrogate IDP structure (Lawrence, Bonny, & Showalter, 2011), interactions with binding partners (Lawrence & Showalter, 2012), and posttranslational modification (Cordek et al., 2014). Others have recently adopted this strategy, leading to novel insights into transcription regulation (Martin et al., 2016) and liquid–liquid phase separation (Dao et al., 2018).

Given the success we and others have experienced with ^{13}C direct-detect NMR of IDPs, it is the objective of this chapter to empower NMR spectroscopists who wish to leverage this technology. To motivate our discussion, this chapter will present data acquired on the intrinsically disordered C-terminal domain of the pancreatic and duodenal homeobox protein I (Pdx1-C throughout). Pdx1 is a transcription factor that regulates the genes coding for insulin, glucagon, and the machinery necessary for endocrine pancreas-controlled glucose homeostasis (Andrali, Sampley, Vanderford, & Ozcan, 2008). The chapter introduces the ^{13}C , ^{15}N -CON as an ideal substitute for the ^1H , ^{15}N -HSQC detection platform, yielding multidimensional experiments for chemical shift determination, structure constraint, and the exploration of intramolecular dynamics. The collection of experimental protocols discussed will provide a robust foundation for comprehensive structure–function studies of IDPs and their interactions.



2. PREPARATION OF ^{15}N - AND ^{13}C -ENRICHED PROTEINS

Successful investigation of protein structure–function relationships through NMR spectroscopy depends on the reproducible generation of recombinant protein that is both stable and homogeneous at room temperature. By necessity, ^{13}C direct-detect experiments require the additional costs of ^{13}C enrichment for all samples, which elevates the need to develop an efficient strategy for optimized protein expression. Importantly, when utilizing “off-the-shelf” inverse configuration cryogenic probes, ^{13}C direct-detect NMR requires protein concentrations ≥ 1.0 mM for most applications. In contrast with traditional ^1H detection, these methods are tolerant to high salt, basic pH, and the presence of solubility-enhancing cosolutes (Bastidas, Gibbs, Sahu, & Showalter, 2015). We assume the reader is familiar with strategies for optimization of protein expression. While there are additional challenges associated with IDP expression and purification, appropriate solution conditions can be achieved for most nonaggregating disordered proteins.

2.1 Plasmid Construction and Protein Expression

Example spectra in this chapter were recorded on the 79 C-terminal amino acids from the human Pdx1 protein (Pdx1-C), which has a high proline abundance (18 of 79 residues) and that otherwise is characterized as a weak polyampholyte, with 0.22 fraction of charged residues and 0.01 net charge per residue (Das & Pappu, 2013). A gene encoding Pdx1-C with an

N-terminal tryptophan for ease of purification was purchased from GeneArt and subcloned into pET49b(+) (Novagen) using the *Xma*I and *Bam*HI restriction sites. This plasmid expresses a GST-6xHis-Pdx1-C fusion protein with a 3C protease site between the fusion tags and the Pdx1-C construct; cleavage leaves a nonnative N-terminal gly-pro-gly-trp, followed by native Pdx1-C, as shown in Fig. 1A. Pdx1-C was expressed in *fluA2*⁻ BL-21

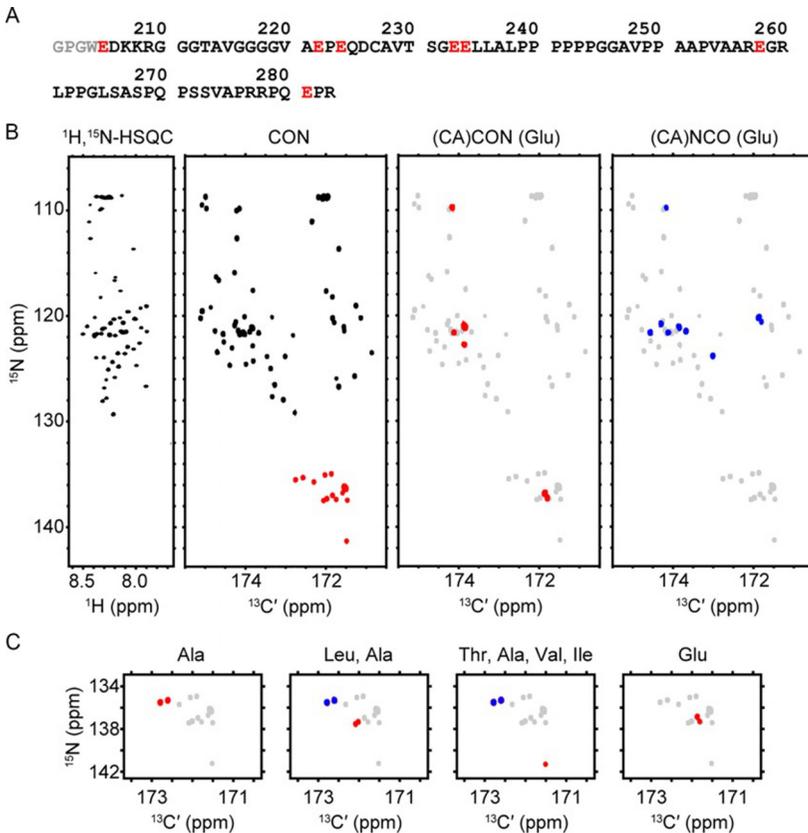


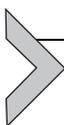
Fig. 1 Representative 2D spectra of Pdx1-C. (A) The amino acid sequence is shown with the cloning artifact in *gray* and glutamate residues in *red*. (B) *From the left*, the ¹H, ¹⁵N-HSQC; ¹³C, ¹⁵N-CON, with Pro *red*; Glu-selective ¹³C, ¹⁵N-(HACA)CON in *red*, with the nonselective spectrum in *gray*; and Glu-selective ¹³C, ¹⁵N-(HACA)NCO in *blue*, with the nonselective spectrum in *gray*. (C) Focus on the Pro-resonances in representative CAS NMR ¹³C, ¹⁵N-(HACA)CON spectra, with the nonselective spectrum in *gray*. *From the left*: Ala-selected resonances in *red*; Leu/Ala selection in *red*, with Ala-only selection superposed in *blue*; Thr, Ala, Val, Ile selection in *red*, with Ala selection superimposed in *blue*; Glu selection in *red*.

(DE3) *Escherichia coli* using quadruple density Luria broth (LB) starter cultures as described by [Marley, Lu, and Bracken \(2001\)](#), which improves yield over traditional M9-only expression.

2.2 Purification of Pdx1-C

1. Prepare 4×100 mL and 4×1 L of LB, each with $50 \mu\text{g}/\text{mL}$ kanamycin. Inoculate the 4×100 mL of LB with *E. coli* transformed with the Pdx1-C coding plasmid and incubate overnight at 37°C with agitation. The next day, use each 100 mL culture to inoculate one 1 L culture and grow until the OD_{600} reaches 0.6. Harvest the cells by centrifugation at $3300 \times g$ and wash residual LB from the cells with 1 L of sterilized M9 salts, lacking glucose and ammonium chloride. Pellet again by centrifugation and resuspend the cell mass in 1 L of M9 supplemented with D-glucose- $^{13}\text{C}_6$ and ^{15}N ammonium chloride. Measure the starting OD_{600} and continue incubation at 37°C until the OD_{600} has increased by ~ 0.1 U; at this time, add $500 \mu\text{L}$ of 1 M IPTG and incubate for an additional 3–4 h at 37°C .
2. Harvest the cells by centrifugation at $3300 \times g$ and discard supernatant. All subsequent steps should be performed at 4°C . Resuspend the cells in lysis buffer (50 mM sodium phosphate, pH 8.0, 500 mM sodium chloride, 20 mM imidazole, and 5 mM β -mercaptoethanol), supplemented with 1 mg/mL lysozyme, 1 mM PMSF, and $1 \times$ protease inhibitor cocktail (EMD Millipore). Incubate cells on ice for 30 min, lyse through sonication, and centrifuge the lysate at $14,000 \times g$.
3. Apply clarified supernatant to 5 mL of Ni^{2+} -loaded immobilized metal affinity chromatography sepharose resin (G Biosciences), preequilibrated with lysis buffer. Wash resin with 40 mL of lysis buffer, supplemented with 0.1% Triton X-100. Rinse resin with 20 mL lysis buffer and elute protein with 20 mL of lysis buffer, supplemented with 200 mM imidazole.
4. Add 3C protease to the eluent and dialyze overnight with stirring at 4°C against 1.6 L of 50 mM sodium phosphate, pH 8.0, 500 mM sodium chloride, and 5 mM β -mercaptoethanol in a 1000 MWCO dialysis tube.
5. Transfer the dialyzed solution to 5 mL of Ni^{2+} -loaded immobilized metal affinity chromatography sepharose resin, preequilibrated with the dialysate. Wash the resin with 15 mL of dialysate. Collect the flow through in both cases; these fractions should contain tag-free Pdx1-C.

6. Finish purification through size exclusion chromatography (SEC) using an S200 1.5 cm \times 75 cm size exclusion column. Preequilibrate the column with SEC buffer (50 mM sodium phosphate, pH 6.5, 50 mM sodium chloride, and 5 mM dithiothreitol (DTT)). Buffer exchange Pdx1-C into SEC buffer and concentrate the protein solution to $<500\ \mu\text{L}$ prior to loading it onto the column. Set the flow rate to 1 mL/minute and collect 1 mL fractions.
7. Pool all fractions of pure Pdx1-C and concentrate using a 3000 MWCO centrifugal filter (MilliporeSigma) to a final volume of $500\ \mu\text{L}$. Add $1\times$ protease inhibitor cocktail, 0.01% (w/v) sodium azide, and 10% (v/v) D_2O , then transfer to an NMR tube.



3. RESONANCE ASSIGNMENT THROUGH ^{13}C DIRECT-DETECT SPECTRA

Chemical shift assignment is usually a means to an end, and is often the most substantial barrier to addressing biological function through NMR spectroscopy. In terms of applications to IDPs, we find that overcoming the assignment challenge is one of the greatest strengths of the ^{13}C direct-detect strategy presented in this chapter. Carbon-detected spectra generally offer superior resonance dispersion and are conducive to the assignment and study of proline-rich constructs. In the case of Pdx1-C, the ^1H , ^{15}N -HSQC spectrum displayed in Fig. 1B demonstrates acceptable resolution, in the sense that most resonances are resolved (nearly) to baseline, but their occupancy of a very narrow chemical shift range is still evident. The contrast between the ^1H , ^{15}N -HSQC and the ^{13}C , ^{15}N -CON experiment (Bermel, Bertini, Felli, Kummerle, & Pierattelli, 2006), also shown in Fig. 1B, is dramatic; detection on $^{13}\text{C}'$ imparts sharp linewidths and excellent chemical shift dispersion to the Pdx1-C spectrum. Pdx1-C features 18 proline residues in its sequence, which are absent in the ^1H , ^{15}N -HSQC, but readily identified in the ^{13}C , ^{15}N -CON by their characteristic ^{15}N chemical shift between 132 and 144 ppm (colored red in Fig. 1B for emphasis).

Given the benefits of ^{13}C direct-detect experiments, it is important to acknowledge their major limitation compared with conventional ^1H -detect NMR: the gyromagnetic ratio of ^{13}C is modest, which results in reduced sensitivity per scan, particularly when using an inverse probe configuration. Fortunately, experimental design can mitigate this disadvantage in most cases. While ^{13}C direct-detect spectroscopy was first developed as a

Table 1 ^{13}C Direct-Detect NMR Experiments for Chemical Shift Assignment

Experiment	Nuclei Observed	Reference
<i>2D spectra</i>		
CON, (HACA)CON	C_{i-1}' , N_i	Bermel et al. (2008)
CAS (CA)CON	C_{i-1}' , N_i where ^{13}C is from selected amino acid	Sahu, Bastidas, and Showalter (2014)
CAS (CA)NCO	C_{i-1}' , N_i where ^{13}C is from selected amino acid C_i' , N_i where ^{15}N is from the selected amino acid	Sahu et al. (2014)
<i>3D spectra</i>		
(HACA)N(CA)CON	C_{i-1}' , N_i , N_{i-1}	Bastidas et al. (2015)
(HACA)N(CA)NCO	C_{i-1}' , N_i , N_{i-1} , N_{i+1}	Bastidas et al. (2015)
CCCON	C_{i-1}' , N_i , C_{ali} , $i-1$	Bermel et al. (2008)
H(CC)CON	C_{i-1}' , N_i , H_{ali} , $i-1$	O'Hare, Benesi, and Showalter (2009)

“protonless” strategy (Bermel, Felli, Kümmerle, & Pierattelli, 2008), we favor utilizing proton-start experiments that enhance per-scan signal to noise. For example, the (HACA)-CON is initiated by magnetization transfer from the $^1\text{H}_\alpha$ nucleus, yielding improved per-scan signal intensity. Finally, the experiments presented here are amenable to nonuniform sampling strategies, which can yield dramatic and time-efficient improvements in spectral quality (Brutscher et al., 2015). Application of these principles led to complete chemical shift assignment of Pdx1-C (BMRB 19596), using only 2 weeks of instrument time on an 11.7T Bruker Avance III NMR spectrometer, equipped with a TCI Cryoprobe. Table 1 shows an overview of the experiments required. The remainder of this section provides rationale and protocols for the specific experiments employed.

3.1 2D NMR

The ^{13}C , ^{15}N -CON serves as the foundation for ^{13}C direct-detect chemical shift assignment, analogous to the ^1H , ^{15}N -HSQC in conventional ^1H -detected spectroscopy. In addition, carbon-detect amino acid-specific (CAS NMR) variants of the ^{13}C , ^{15}N -CON spectrum provide a powerful complement to 3D-based assignments (Bermel et al., 2012), as we have

described for Pdx1-C (Sahu et al., 2014). The CAS NMR strategy utilizes pairs of amino acid-specific spectra, collected in (CA)CON and (CA)NCO format. The general approach is analogous to the proton-detected triple-resonance NMR assignment strategy: the (CA)CON is filtered such that the only resonances retained are for peptide planes in which the $^{13}\text{C}'$ corresponds to the amino acid matching the selection filter; the (CA)NCO is bidirectional, building dipeptide linkages in which the selected amino acids are present either to the N-terminal or C-terminal side of the peptide plane. By convention, the resonances in CON-format spectra are labeled with the residue number of the amino acid that contributes the ^{15}N to the peptide plane; in this scheme, general references are made to the ^{15}N of residue i and the $^{13}\text{C}'$ of residue $i - 1$. This concept is demonstrated in Fig. 1B, which displays glutamate-selective (CA)CON and (CA)NCO spectra. In the (CA)CON spectrum, only one resonance appears in the ^{15}N chemical shift range associated with glycine (≤ 114 ppm); thus G259 was unambiguously assigned. The set of resonances corresponding to peptide planes where the ^{15}N belongs to a glutamate are identified as the peaks appearing in the (CA)NCO, but not the (CA)CON. The utility of this technique is especially clear in the context of the resonances that arise from peptide bonds where proline contributes the ^{15}N atom. In Fig. 1C, we see how to combine the A-, TAVI-, LA-, and E-specific spectra to accelerate assignment of the proline resonances, which is advantageous for a proline-rich system like Pdx1-C.

3.1.1 Experimental Considerations for ^{13}C Direct-Detect 2D Spectra

1. Prepare an NMR sample, perhaps as described in Section 2.2, and collect preliminary ^1H , ^{15}N -HSQC, ^{13}C , ^{15}N -CON, and ^{13}C , ^{15}N -(HACA) CON spectra. Collect both the carbon-start and the proton-start versions of the CON to determine which yields a better spectrum for the protein of interest. Typically, the proton-start spectrum is preferred, due to enhanced sensitivity, but we have experienced some situations in which internal dynamics or solution conditions (most notably high salt) make “protonless” spectra preferable. Suggested acquisition parameters are summarized in Table 2.
2. Homonuclear decoupling during acquisition is challenging, so “virtual decoupling” to remove $^{13}\text{C}'$ - $^{13}\text{C}_\alpha$ scalar coupling has been adopted as the standard method to simplify the presentation of ^{13}C direct-detect spectra. All of the pulse programs we have published and distribute

Table 2 Recommended Parameters for 2D NMR Experiments

	HSQC	CON	(HACA)CON	PRE	RDC	Relaxation
Detection	¹ H	¹³ C				
F2 TD (direct)	2048	1024	1024	1024	1024	1024
F1 TD (indirect)	256	256	256	512	512	512
Number of scans	≤8	16	8	32	32	32

record spectra in an in-phase/antiphase (IPAP) format that allows for script-based virtual decoupling. For example, executing the standard Bruker AU script “splitcomb” shifts and recombines the IPAP spectra for final presentation of the data.

- Depending on sample conditions, CAS NMR may require up to 1 day for each (CA)CON, (CA)NCO pair. These spectra tend to be weaker than the nonselective ¹³C, ¹⁵N-CON, especially the (CA)NCO, so a minimum of 32 scans may be needed for optimal signal averaging.
- The (CA)NCO often yields weak $i - 1$ resonances, even when the self-identifying resonance is strong. This phenomenon is evident in Fig. 1, where some of the (CA)CON resonances do not appear in the glutamate-selective (CA)NCO. The absence of an $i - 1$ resonance is not a cause for concern, so long as it was identified in the paired (CA)CON spectrum.
- Many CAS NMR spectra are not selective for a single amino acid type, but rather for a subset of chemically similar amino acids. Ambiguity can be alleviated by collecting an exhaustive set of CAS NMR spectra. For example, the alanine-selective spectrum can be used to distinguish the leucine and alanine-associated resonances in the “LA” selective spectrum.

3.2 3D NMR

Most spectroscopists familiar with triple-resonance assignment strategies are accustomed to building backbone connectivity through the ¹³C_α and ¹³C_β nuclei, which is also feasible with ¹³C direct-detect spectra (Motackova et al., 2010). However, aliphatic carbon chemical shifts of IDPs generally do not deviate much from random coil values, which can produce assignment complications for repetitive sequences. Instead, we find great advantage in 3D spectra built from the (CA)CON and (CA)NCO that correlate

the backbone amide ^{15}N chemical shifts of pairs or triplets of amino acids. The (HACA)N(CA)CON records the ^{15}N of residue i as an autocorrelation peak with a cross-correlation to the ^{15}N chemical shift for residue $i - 1$. The (HACA)N(CA)NCO provides complementary information: the indirect ^{15}N dimension records cross-correlation peaks to the ^{15}N resonance of residues $i - 1$ and $i + 1$. Use of these experiments is illustrated in Fig. 2 for residues 255–259 from Pdx1-C, where the corresponding resonances in the 2D ^{13}C , ^{15}N -CON are colored in red as a visual reference. The 3D CCCON (Bermel, Bertini, Felli, Piccioli, & Pierattelli, 2006) and H(CC)CON (O'Hare et al., 2009) can be used to complete aliphatic ^{13}C and ^1H chemical shift assignment, respectively, as demonstrated in Fig. 3 for the same Pdx1-C residues that were used to demonstrate backbone assignment.

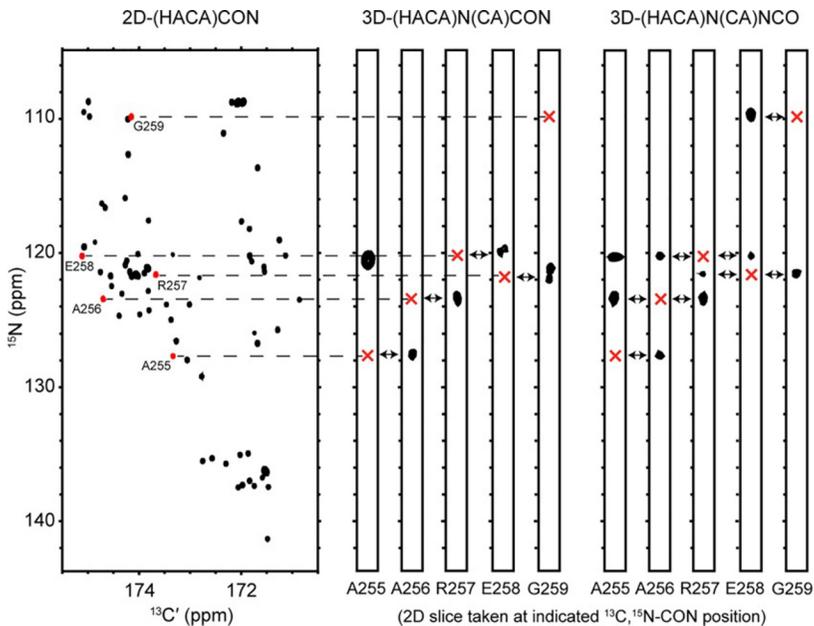


Fig. 2 Demonstrated use of the (HACA)N(CA)CON and (HACA)N(CA)NCO to build backbone sequence connectivity, using residues A255–G259 of Pdx1-C. The ^{13}C , ^{15}N -CON on the *left* has been colored such that the five assigned residues are in *red*. For a representative segment of five residues, a *dashed line* is displayed that connects the labeled 2D ^{13}C , ^{15}N -CON resonance to the autocorrelation peak (displayed as a *red X*) in the 2D planes extracted from the 3D spectrum to highlight ^{15}N correlations in the third dimension. Cross-peak connectivity is indicated through *double-headed arrows* between the strips.

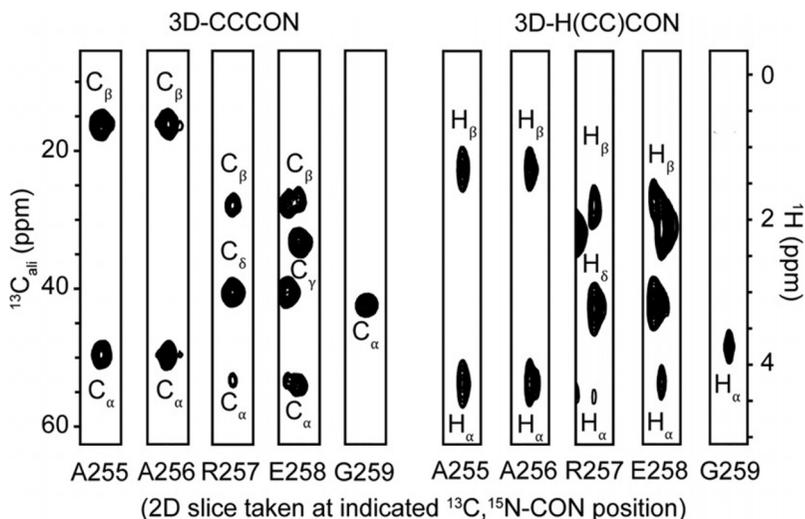
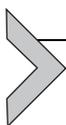


Fig. 3 The CCCON and H(CC)CON experiments identify the aliphatic carbon and proton chemical shifts, respectively. 2D projections show resonances associated with residues A255–G259.

3.2.1 Experimental Considerations for ^{13}C Direct-Detect 3D Spectra

1. There are both H_N -start (Bermel et al., 2012) and H_α -start (Bastidas et al., 2015) versions of the N(CA)CON and N(CA)NCO. In general, we favor the H_α -start versions, which offer better performance for proline-rich proteins.
2. The (HACA)N(CA)NCO is auxiliary, but can dramatically improve backbone coverage for difficult cases due to its ability to report on residues $i - 1$ and $i + 1$. This spectrum displays low per-scan signal to noise, so where time and sample longevity permit, additional signal averaging can be beneficial.
3. The CCCON and H(CC)CON offer facile assignment of all nuclei required for secondary structure determination. If collected in $^1\text{H}_\text{ali}$ -start formats, both offer high sensitivity.



4. PARAMAGNETIC RELAXATION ENHANCEMENT

The dynamic nature of disordered protein ensembles presents a significant barrier to studying their structure–function relationships unless methods are employed that can report on low-population or short-lived states that may be critical for biological function. Paramagnetic relaxation

enhancement (PRE) techniques offer an avenue to investigate transient intra- and intermolecular IDP interactions and, for this reason, have been widely adopted (Schneider et al., 2012). The PRE is similar in mechanism to the nuclear Overhauser effect (NOE), except the dipolar coupling exploited for the measurement exists between an unpaired electron and nearby nuclei. Due to the high gyromagnetic ratio of the electron, PRE can report on contacts that are short lived, or even on persistent colocalization out to the 20–30 Å length scale (Clore & Iwahara, 2009). Semiquantitative readout is achieved through recording paired paramagnetic and diamagnetic spectra, such as the 2D ^{13}C , ^{15}N -CON spectra discussed in this chapter. The ratio of the paramagnetic and diamagnetic resonance intensity is proportional to the distance separating the spin probe and the nuclei that contribute the resonance peak.

Most IDPs do not contain native paramagnetic centers, so covalent modification with a spin probe is required. The most common method is to attach a nitroxide spin label, such as *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl)methyl methanesulfonylthioate (MTSL) to a cysteine sidechain (Berliner, Grunwald, Hankovszky, & Hideg, 1982). Protein engineering to ensure that the system studied presents one cysteine for modification may be required; in any case, perform control experiments to verify that mutagenesis and MTSL conjugation have not altered the structure and biological activity of the IDP investigated.

In the case of Pdx1-C, the native cysteine at position 227 of the human sequence provides an excellent position for MTSL conjugation, leading to the PRE data shown in Fig. 4. Traditionally, PRE has been reported through the ^1H , ^{15}N -HSQC, as displayed on the left, but it is clear that spectral crowding can impair quantitative peak height determination when IDPs are investigated using this spectrum. In contrast, the improved dispersion in the ^{13}C , ^{15}N -CON spectrum provides excellent quantitative detection for IDPs. Notably, comparison of the ^{13}C -start and ^1H -start CON (middle and right panels, respectively) reveals several resonances that lose intensity only in the ^1H -start format (examples highlighted with blue boxes), which suggests that the excitation nucleus selected scales the sensitivity of CON-detected measurements to the relaxation effects of the paramagnetic center.

4.1 PRE Protocol

1. Purify the target protein using the method from Section 2.2 or equivalent, concentrate it to $>500\ \mu\text{M}$, and incubate for 30 min with 10 *M* equivalents of DTT to ensure near complete disulfide bond reduction.

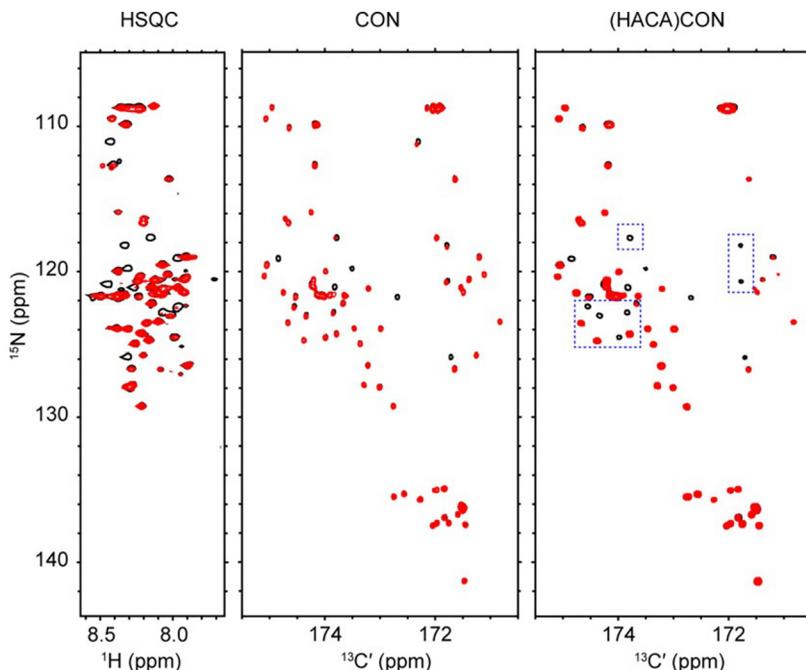


Fig. 4 Pdx1-C spectra for PRE measurement through ^1H , ^{15}N -HSQC, ^{13}C , ^{15}N -CON, and ^{13}C , ^{15}N -(HACA)CON. Spectra acquired on the paramagnetic sample are shown in *red* as an overlay with the diamagnetic spectra in *black*. Under paramagnetic conditions, several resonances lose significant intensity in the ^{13}C , ^{15}N -(HACA)CON, but not the ^{13}C , ^{15}N -CON due to the difference in excitation nucleus. Three clusters of resonances that demonstrate this phenomenon are highlighted through enclosure in *dashed blue boxes*.

2. Remove the unreacted DTT by passing the protein over a PD-10 desalting column equilibrated with buffer that is free of reducing agent. Immediately combine the reduced protein with 20M equivalents of MTSL dissolved in methanol. Use a blackout tube or aluminum foil to protect the sample, as MTSL is light sensitive. Allow the labeling reaction to proceed at 4°C for 1 h and then buffer exchange the MTSL-labeled protein into NMR buffer without reducing agent.
3. It is recommended that an aliquot of MTSL-labeled protein be reserved for analysis of the labeling efficiency through mass spectrometry or other means.
4. Collect a ^{13}C , ^{15}N -CON spectrum of the paramagnetic sample using the experimental parameters suggested in [Table 2](#). If sample lifetime permits, it is ideal to collect both the ^{13}C -start and the $^1\text{H}_{\alpha}$ -start version of the CON, as they tend to report the PRE differently.

5. Quench the nitroxide radical by adding 5 M equivalents of sodium ascorbate to the sample. Repeat collection of the experiments recorded in Step 4 using the now diamagnetic sample. All spectroscopic parameters should be identical in order to ensure that the only changes in resonance intensity arise from differences in paramagnetism.
6. Calculate the PRE for each residue as the ratio of the peak intensities in the paramagnetic and diamagnetic forms ($PRE = I_{\text{para}}/I_{\text{dia}}$). Overlapping peaks should be omitted from analysis. As a best practice, perform measurements in duplicate and from independent biological replicates if it is possible to do so.



5. RESIDUAL DIPOLAR COUPLINGS

By virtue of their enhanced structural dynamics, most disordered proteins are not well suited for structural constraint by NOEs. For this reason, residual dipolar coupling (RDC) has emerged as a successful technique for generating IDP ensemble constraints. In an isotropic medium, the dipolar coupling between nuclear spins averages to zero; however, with slight solution anisotropy, averaging is incomplete and a “residual” dipolar coupling is preserved without loss of the sharp solution line shape. When dipolar coupling exists between nuclei that are covalently bonded, through up to two bonds, their internuclear distance is essentially fixed. Thus, the RDC serves as a proxy for internuclear vector orientation relative to a global coordinate system. The origin and applications of the RDC have been reviewed (Salmon & Blackledge, 2015), so the general theory will not be explained further here.

As a supplement to existing ^1H -detected RDC experiments, we have developed $^{13}\text{C}, ^{15}\text{N}$ -CON variants that encode the NH, $\text{C}'\text{C}_\alpha$, $\text{C}'\text{N}$, and $\text{H}_\alpha\text{C}_\alpha$ RDC. In all cases, RDCs are recorded through a pair of spectra collected using the “in-phase (IP)/antiphase (AP)” technique in which the scalar coupling between the two nuclei under investigation is not refocused during the indirect chemical shift evolution period, resulting in the pair of spectra like those depicted on the left of Fig. 5. In the IP spectrum, both components of the doublet are phased equivalently, whereas the AP doublet features a 180° phase offset between its two components. Addition and subtraction of the IP and AP spectra leads to the spectra depicted on the right of Fig. 5, with the advantage that the scalar coupling constant can be determined from the peak offset without adding spectral crowding. Comparison of the splitting in the isotropic and aligned sample yields the RDC value for each spin pair through subtraction.

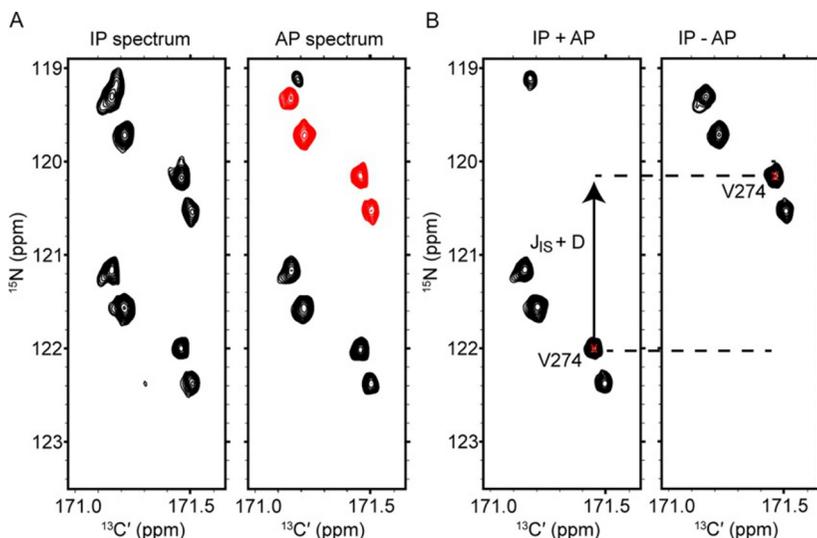


Fig. 5 RDC experiments are collected in an IPAP format as demonstrated for V274. (A) $^1J_{\text{NH}}$ is preserved to produce a doublet that is either in-phase or antiphase (negative phase is indicated through *red contours*). (B) Algebraic sum and difference spectra yield two spectral representations that preserve either the upfield or downfield component of the doublet, which facilitates quantitation of peak positions and determination of the scalar coupling constant.

5.1 RDC Protocol

1. Establish solution conditions that lead to good alignment without spectral distortion or perturbation of the IDP conformational ensemble. For general strategies, we refer the reader to established protocols (Chen & Tjandra, 2012). This protocol reflects studies of Pdx1-C, for which residual alignment was introduced with a compressed, neutral, 7% polyacrylamide gel (Sass, Musco, Stahl, Wingfield, & Grzesiek, 2000).
2. Collection of RDCs is achieved through variants of the ^{13}C , ^{15}N -CON that preserve scalar coupling in the indirect dimension, such as the (HN) CON-IPAP, which preserves $^1J_{\text{NH}}$ in the ^{15}N dimension. Refer to Table 2 for suitable experimental parameters.
3. For alignment by gel compression, prepare the unaligned sample in a Shigemi-style tube. Collect the (HN)CON-IPAP and any other desired spectra needed to record, e.g., $^1J_{\text{C}'\text{N}}$, $^1J_{\text{H}\alpha\text{C}\alpha}$, or $^1J_{\text{C}'\text{C}\alpha}$.
4. Add the dehydrated gel to the Shigemi-style tube containing the sample. Set the plunger to the height needed for the desired compression ratio and secure it with parafilm. Gentle agitation of the tube is recommended for at least 20 min to prevent the gel from sticking to the wall of the tube

as it expands. It is best practice to allow overnight equilibration; if possible, gel hydration should occur in the bore of the magnet, such that equilibration of the aligned state with respect to the magnetic field is simultaneously established.

5. Collect the (HN)CON-IPAP spectrum, as well as any other desired spectra, on the aligned sample with the same experimental parameters as the unaligned sample.
6. The spectra acquired utilize the IPAP method in two different ways that must be sequentially resolved in processing. First, use the “splitcomb” AU script within the Bruker Topspin software, or equivalent, to achieve virtual decoupling in the direct dimension of the CON. After virtual decoupling, split the in-phase and antiphase spectra, using the “split” AU script or equivalent. The IP and AP spectra may now be Fourier transformed and processed as usual.
7. Algebraic addition and subtraction of the IP and AP spectra through the “add2d” command in Topspin, or equivalent, yields convenient subspectra from which the resonance frequency of the upfield and downfield components of the doublet pattern can be measured. Splitting in the unaligned sample yields $^1J_{\text{NH}}$, while splitting in the aligned spectrum yields $^1J_{\text{NH}} + ^1D_{\text{NH}}$; subtraction of the two splitting values produces the final RDC measurement.



6. SPIN RELAXATION MEASUREMENT

Measurement of ^{15}N spin relaxation is a routine method used to assess protein backbone flexibility in both well-folded and intrinsically disordered systems. Adaptation of the ^{13}C , ^{15}N -CON to report ^{15}N R_1 and R_2 relaxation rates has been reported (Lawrence & Showalter, 2012). While less common to employ, $^{13}\text{C}'$ spin relaxation is also accessible in a ^{13}C direct-detect format (Bermel, Bertini, Felli, Peruzzini, & Pierattelli, 2010; Pasat, Zintsmaster, & Peng, 2008), as well as various cross-relaxation pathways (Bertini, Felli, Gonnelli, Kumar, & Pierattelli, 2011). In contrast with spin relaxation of cooperatively folded and globular systems, the ^{15}N R_2 of IDPs tends toward unexpectedly low values (i.e., long T_2 relaxation times) that are close in magnitude to the recorded R_1 values. In other words, the spin relaxation of IDPs often looks similar to that of small molecules, even for large systems. The molecular basis for this phenomenon is explained through polymer physics considerations (Pappu, Wang, Vitalis, & Crick, 2008), which lead to the conclusion that the thermal decorrelation length is short for IDPs compared to the overall chain length. Collection in a

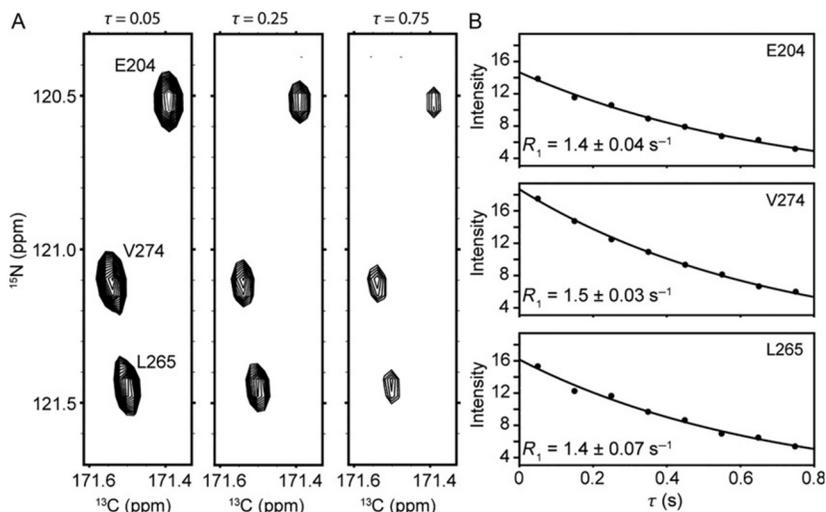


Fig. 6 ^{15}N spin relaxation can be recorded through intensity changes in $^{13}\text{C}, ^{15}\text{N}$ -CON spectra. (A) The signal intensity for resonances E204, L265, and V274 are shown as a function of longitudinal relaxation delay time. (B) The intensities from a full set of $^{13}\text{C}, ^{15}\text{N}$ -CON spectra are fit to a single exponential decay, yielding R_1 , here, or R_2 for equivalent data.

^{13}C direct-detect format is directly analogous to standard ^1H -detected experiments. For example, residues E204, V274, and L265 from Pdx1-C display monoexponential relaxation, as demonstrated in Fig. 6.

6.1 Protocol for ^{15}N Relaxation Measurements

1. Construct the delay list for the R_1 and R_2 experiments. For R_1 , a minimum delay of 50 ms with linear spacing to a final delay of ≥ 750 ms is recommended. Using Carr–Purcell–Meiboom–Gill (CPMG) for R_2 measurement constrains the relaxation delays to multiples of the CPMG supercycle (typically 16 ms) and delays ≥ 304 ms may be required.
2. Refer to Table 2 for suitable indirect dimension size and signal averaging.
3. It is preferable to collect each relaxation set as a “pseudo-3D,” in which the data are interleaved, which will minimize the influence of systematic artifacts on the data. In this case, analysis requires splitting the data into serial 2D spectra by using the AU script “split-relax,” or equivalent. All subsequent processing is standard.
4. For each resonance in the $^{13}\text{C}, ^{15}\text{N}$ -CON, record the peak intensity as a function of relaxation delay time. Data should fit to a monoexponential decay model.



7. SUMMARY

This chapter has surveyed a purely ^{13}C direct-detect approach to biomolecular NMR that is designed to provide comprehensive and quantitative insight into the average structures and conformational dynamics of the solution ensembles adopted by IDPs and other highly flexible systems. The strategy described requires a probe with very high ^{13}C -sensitivity, which typically means a cryogenically cooled probe, and proteins that are soluble to high concentration ($\geq 1\text{ mM}$). For most nonaggregating IDPs, the investigator should be able to achieve the needed concentration; the availability of suitable probes is also becoming more common. For ease of use, we have made Bruker-formatted pulse programs for all experiments described in this chapter available through the Penn State ScholarSphere through the stable URL: <https://scholarsphere.psu.edu/collections/5712mv15x>, or a keyword search on the ScholarSphere home page. Many of the described pulse sequences are already incorporated into contemporary distributions of the Bruker Topspin pulse program library. As a result of pulse program and hardware availability, and facilitated by the protocols described here, ^{13}C direct-detect NMR is poised to become a leading tool in the disordered protein community.

ACKNOWLEDGMENTS

We thank the staff of the Lloyd Jackman NMR Facility in the Penn State Chemistry Department for their consistent support of our research program. We thank the Penn State ScholarSphere team (<http://scholarsphere.psu.edu>) for establishing permanent and stable web hosting of our pulse programs and other aspects of the research contributing to this chapter. This work was funded in part by NSF award MCB-1515974 to S.A.S.

REFERENCES

- Andrali, S. S., Sampley, M. L., Vanderford, N. L., & Ozcan, S. (2008). Glucose regulation of insulin gene expression in pancreatic beta-cells. *The Biochemical Journal*, *415*, 1–10.
- Bastidas, M., Gibbs, E. B., Sahu, D., & Showalter, S. A. (2015). A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts in Magnetic Resonance Part A*, *44*, 54–66.
- Berliner, L. J., Grunwald, J., Hankovszky, H. O., & Hideg, K. (1982). A novel reversible thiol-specific spin label: Papain active site labeling and inhibition. *Analytical Biochemistry*, *119*, 450–455.
- Bermel, W., Bertini, I., Chill, J., Felli, I. C., Haba, N., Kumar, M. V. V., et al. (2012). Exclusively heteronuclear (^{13}C) C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). *ChemBiochem*, *13*, 2425–2432.
- Bermel, W., Bertini, I., Felli, I. C., Gonnelli, L., Kozminski, W., Piai, A., et al. (2012). Speeding up sequence specific assignment of IDPs. *Journal of Biomolecular NMR*, *53*, 293–301.

- Bermel, W., Bertini, I., Felli, I. C., Kummerle, R., & Pierattelli, R. (2006). Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *Journal of Magnetic Resonance*, *178*, 56–64.
- Bermel, W., Bertini, I., Felli, I. C., Peruzzini, R., & Pierattelli, R. (2010). Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins. *Chemphyschem*, *11*, 689–695.
- Bermel, W., Bertini, I., Felli, I. C., Piccioli, M., & Pierattelli, R. (2006). ^{13}C -detected protonless NMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *48*, 25–45.
- Bermel, W., Felli, I. C., Kümmerle, R., & Pierattelli, R. (2008). ^{13}C direct-detection biomolecular NMR. *Concepts in Magnetic Resonance Part A*, *32A*, 183–200.
- Bertini, I., Felli, I. C., Gonnelli, L., Kumar, M. V. V., & Pierattelli, R. (2011). High-resolution characterization of intrinsic disorder in proteins: Expanding the suite of (^{13}C)-detected NMR spectroscopy experiments to determine key observables. *Chembiochem*, *12*, 2347–2352.
- Brutscher, B., Felli, I. C., Gil-Caballero, S., Hosek, T., Kummerle, R., Piai, A., et al. (2015). NMR methods for the study of intrinsically disordered proteins structure, dynamics, and interactions: General overview and practical guidelines. *Advances in Experimental Medicine and Biology*, *870*, 49–122.
- Chen, K., & Tjandra, N. (2012). The use of residual dipolar coupling in studying proteins by NMR. *Topics in Current Chemistry*, *326*, 47–67.
- Clore, G. M., & Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews*, *109*, 4108–4139.
- Cordek, D. G., Croom-Perez, T. J., Hwang, J., Hargittai, M. R., Subba-Reddy, C. V., Han, Q., et al. (2014). Expanding the proteome of an RNA virus by phosphorylation of an intrinsically disordered viral protein. *The Journal of Biological Chemistry*, *289*, 24397–24416.
- Dao, T. P., Kolaitis, R. M., Kim, H. J., O'Donovan, K., Martyniak, B., Colicino, E., et al. (2018). Ubiquitin modulates liquid–liquid phase separation of UBQLN2 via disruption of multivalent interactions. *Molecular Cell*, *69*, 965–978.e6.
- Das, R. K., & Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 13392–13397.
- Gibbs, E. B., Cook, E. C., & Showalter, S. A. (2017). Application of NMR to studies of intrinsically disordered proteins. *Archives of Biochemistry and Biophysics*, *628*, 57–70.
- Gibbs, E. B., & Showalter, S. A. (2015). Quantitative biophysical characterization of intrinsically disordered proteins. *Biochemistry*, *54*, 1314–1326.
- Lawrence, C. W., Bonny, A., & Showalter, S. A. (2011). The disordered C-terminus of the RNA polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochemical and Biophysical Research Communications*, *410*, 461–465.
- Lawrence, C. W., & Showalter, S. A. (2012). Carbon-detected N-15 NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. *Journal of Physical Chemistry Letters*, *3*, 1409–1413.
- Marley, J., Lu, M., & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, *20*, 71–75.
- Martin, E. W., Holehouse, A. S., Grace, C. R., Hughes, A., Pappu, R. V., & Mittag, T. (2016). Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *Journal of the American Chemical Society*, *138*, 15323–15335.
- Motackova, V., Novacek, J., Zawadzka-Kazimierczuk, A., Kazimierczuk, K., Zidek, L., Sanderova, H., et al. (2010). Strategy for complete NMR assignment of disordered

- proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *Journal of Biomolecular NMR*, *48*, 169–177.
- O'Hare, B., Benesi, A. J., & Showalter, S. A. (2009). Incorporating ^1H -chemical shift determination into ^{13}C -direct detected spectroscopy of intrinsically disordered proteins in solution. *Journal of Magnetic Resonance*, *200*, 354–358.
- Pappu, R. V., Wang, X., Vitalis, A., & Crick, S. L. (2008). A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Archives of Biochemistry and Biophysics*, *469*, 132–141.
- Pasat, G., Zintsmaster, J. S., & Peng, J. W. (2008). Direct ^{13}C -detection for carbonyl relaxation studies of protein dynamics. *Journal of Magnetic Resonance*, *193*, 226–232.
- Sahu, D., Bastidas, M., & Showalter, S. A. (2014). Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Analytical Biochemistry*, *449*, 17–25.
- Salmon, L., & Blackledge, M. (2015). Investigating protein conformational energy landscapes and atomic resolution dynamics from NMR dipolar couplings: A review. *Reports on Progress in Physics*, *78*, 126601.
- Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T., & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *Journal of Biomolecular NMR*, *18*, 303–309.
- Schneider, R., Huang, J. R., Yao, M., Communie, G., Ozenne, V., Mollica, L., et al. (2012). Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Molecular BioSystems*, *8*, 58–68.
- Uversky, V. N. (2018). Intrinsic disorder, protein-protein interactions, and disease. *Advances in Protein Chemistry and Structural Biology*, *110*, 85–121.