A Primer for Carbon-Detected NMR Applications to Intrinsically Disordered Proteins in Solution

MONIQUE BASTIDAS, ERIC B. GIBBS, DEBASHISH SAHU, SCOTT A. SHOWALTER

Department of Chemistry, the Pennsylvania State University, University Park, PA 16802

ABSTRACT: Characterization of intrinsically disordered proteins (IDPs) has grown tremendously over the past two decades. NMR-based structural characterization has been widely embraced by the IDP community, largely because this technique is amenable to highly flexible biomolecules. Particularly, carbon-detect nuclear magnetic resonance (NMR) experiments provide a straight forward and expedient method for completing backbone assignments, thus providing the framework to study the structural and dynamic properties of IDPs. However, these experiments remain unfamiliar to most NMR spectroscopists, thus limiting the breadth of their application. In an effort to remove barriers that may prevent the application of carbon-detected bio-NMR where it has the potential to benefit investigators, here we describe the experimental requirements to collect a robust set of carbon-detected NMR data for complete backbone assignment of IDPs. Specifically, we advocate the use of three-dimensional experiments that exploit magnetization transfer pathways initiated on the aliphatic protons, which produces increased sensitivity and provides a suitable method for IDPs that are only soluble in basic pH conditions (>7.5). The applicability of this strategy to systems featuring a high degree of proline content will also be discussed.

© 2015 Wiley Periodicals, Inc.

Concepts Magn Reson Part A 00: 000–000, 2015.

KEY WORDS: NMR resonance assignment; carbon detection; Intrinsically disordered protein; Pdx1; FCP1

Received 26 February 2015; revised 8 April 2015; accepted 12 April 2015

Correspondence to: Scott A. Showalter; E-mail: sas76@psu.edu

Concepts in Magnetic Resonance Part A, Vol. 00A(00) 00–00 (2015)

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cmr.a.21327

© 2015 Wiley Periodicals, Inc.

INTRODUCTION

NMR spectroscopy is the most powerful method available for determining the conformational and dynamic properties of highly flexible biopolymers in solution. Intrinsically disordered proteins (IDPs) have been embraced by the NMR community in recent years

because their disordered nature pushes the boundaries of structural biology (*1*). Unfortunately, the very same conformational disorder that makes IDPs fascinating to study often leads to poor spectral quality, particularly when traditional proton-detected, triple resonance NMR approaches are utilized. Solving this problem has been a focus of our laboratory for several years. Here, we describe a framework to study structure-function relationships for IDPs that relies exclusively on carbon-detected solution NMR spectroscopy (*2*). We find this strategy to be generally effective and have applied it to an IDP that possesses significant amounts of secondary structure (*3*), as well as an IDP that is nearly indistinguishable from a random coil (*4*).

An excellent general introduction to carbon-detected solution NMR strategies that are applicable to biomolecules was previously published in this journal (*5*). In their review, Bermel et al. expertly address several of the key issues general to carbon-detected measurements on uniformly ^{13}C-enriched biomolecules, placing strong emphasis on the "virtual decoupling" methods that are used to suppress scalar coupling in the direct-detected dimension (*5*). Here, we will present several pulse programs that utilize in-phase/antiphase virtual decoupling and refer the reader to the above mentioned review if this concept is unfamiliar.

Until recently, carbon-detected NMR strategies applicable to proteins in solution were built around so-called "protonless" pulse sequences, in which pulsing on the ^{1}H-nucleus was scrupulously avoided for both excitation and detection (*6*). The choice to remain rigorously "protonless" was motivated by an interest in studying metaloenzymes with paramagnetic centers in their active sites. About 5 years prior to the publication of this article, we and others began to realize that carbon-detected strategies were also a highly efficient means to study IDPs. Without paramagnetic relaxation enhancement of transverse ^{1}H spin-states present as an experimental design factor, we began incorporating polarization transfer from ^{1}H to lower gyromagnetic ratio heteronuclei in the excitation elements of our pulse programs (*7,8*). At present, virtually all carbon-detected biomolecular NMR experiments have been converted to H-start formats that offer enhanced sensitivity. For IDPs, either care must be taken with these experiments to transfer magnetization from aliphatic protons or, if amide-proton polarization is used to initiate the pulse program, the sample must be kept in mildly acidic conditions to avoid overwhelming solvent exchange. Recent advances, including new pulse sequences we introduce here, enable studies of IDPs in basic solution conditions (pH $>$7.5). Overall, our experience suggests that carbon-detected NMR should enable the study of virtually any IDP that is of biological interest, with only very modest constraints on the solution conditions needed for experimental feasibility. This review will summarize the best practices we have established for our laboratory in the hope that interested investigators will be able to efficiently utilize this exciting new strategy for their own disordered protein systems.

EXPERIMENTAL SYSTEMS USED TO PROVIDE CASE STUDIES

Our discussion will be motivated through the presentation of spectra collected on two proteins actively studied in our laboratory, which have very different amino acid compositions, secondary structure biases, and stable solution pH and buffer conditions. The C-terminal tail of the RNA Polymerase II C-terminal domain phosphatase (ctFCP1) is a highly charged polyampholyte that we have shown possesses a partially helical region at its extreme C-terminus (*3*). In contrast, the C-terminal domain of the pancreatic and duodenal homeobox protein 1 (Pdx1-C) is strongly disordered in solution, deficient in acidic or basic amino acid residues, and highly enriched in proline and glycine residues (*4*). Together, these two proteins offer excellent examples of some of the more extreme conditions spectroscopists investigating IDPs may encounter.

PULSE PROGRAM AND SPECTROMETER SELECTION

As a result of several decades of vigorous research, proton-detected biomolecular NMR has become codified in protocols that form the basis for several excellent textbooks. Virtually any classically trained biological NMR spectroscopist will be familiar with the general triple-resonance chemical shift assignment strategy, built around detection through the ^{1}H,^{15}N-HSQC, or the qualitatively similar ^{1}H,^{15}N-TROSY. These experiments are widely popular because they provide nominally one-to-one mapping between resonances and residues, with the exception of prolines, and can be rapidly acquired. Moreover, the use of the highest static magnetic field strength available provides additional benefits, such as improved spectral resolution and the potential to accelerate data collection or reduce sample concentration, due to enhanced sensitivity. Benefits in terms of sample concentration requirements and/or data collection speed are amplified when cryogenically cooled probes are utilized.

In the case of carbon-detected biomolecular NMR, experimental strategies can be built around several two-dimensional (2D)-detection platforms. In uniformly

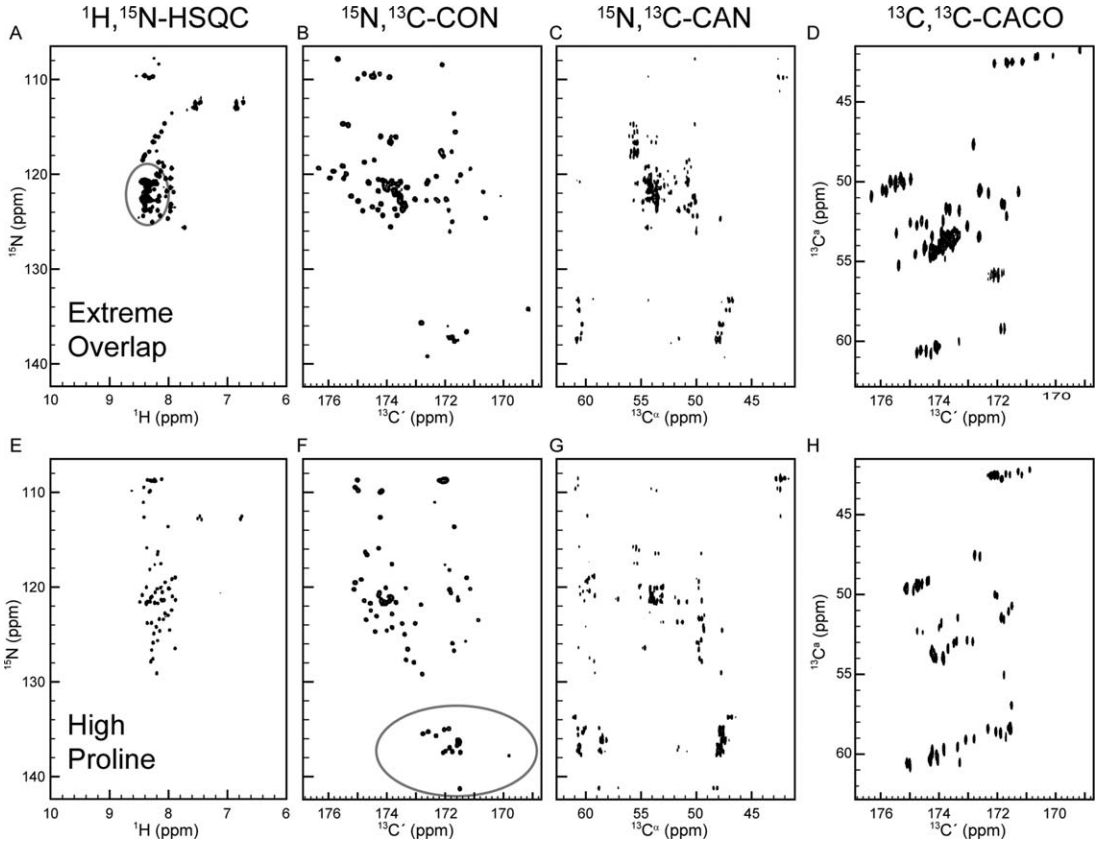

Figure 1 Carbon-detected 2D-NMR experiments are generally effective for generating well-resolved spectra of IDPs. (A–D) Spectra acquired using a 1.0 mM sample of ctFCP1 in 20 mM sodium phosphate, pH 7.0, 100 mM NaCl, 0.02% NaN_3, 10% (v/v) D_2O. (E–H) Spectra acquired using a 1.3 mM sample of Pdx1-C in 50 mM sodium cacodylate, pH 6.5 (adjusted after TCEP addition), 50 mM KCl, 5 mM TCEP, 0.01% NaN_3, 10% (v/v) D_2O. The 2D-experiments displayed are (A,E) ^{1}H,^{15}N-HSQC, (B,F), ^{15}N,^{13}C-CON, (C,G) ^{15}N,^{13}C-CAN, and (D,H) ^{13}C,^{13}C-CACO. All spectra were collected at 298K on a Bruker Avance III spectrometer equipped with a TCI Cryoprobe (^{1}H inner coil), operating at 500 MHz proton resonance frequency (11.7 T).

^{13}C-enriched samples, it is likely that the detection-nucleus chosen will have a robust one-bond scalar coupling to multiple spin-1/2 nuclei that are suitable to generate the indirect spectral dimension. In this section, we briefly introduce three of the most popular carbon-detected 2D experiments, placing particular emphasis on the common variants of the extremely powerful ^{15}N,^{13}C-CON experiment. We close the section with a brief commentary on the performance of CON-based experimental strategies at ultrahigh magnetic field strength.

Selecting the Best 2D Correlation Experiment for the Job

Chemical shift is an inherently local reporter of structure and chemical environment. For cooperatively folded proteins, where well-defined secondary and tertiary structure provides nearly unique chemical environments to each residue of the chain, high spectral resolution in experiments like the ^{1}H,^{15}N-HSQC is often achievable. In contrast, for IDPs, environmental contributions to chemical shifts are effectively averaged out by structural heterogeneity and low sequence complexity, resulting in low signal dispersion, as demonstrated for ctFCP1in Fig. 1(A). However, counter-examples of IDPs that do yield well dispersed signals in a ^{1}H,^{15}N-HSQC are not hard to find, as demonstrated by Pdx1-C in Fig. 1(E). In the case of Pdx1-C, for which 22% of the chain is composed of proline residues, the inability to detect proline precludes attempts at unambiguous resonance assignment by ^{1}H-detect strategies. As such, the acquisition of a well-resolved 2D spectrum does not guarantee project success. While the potential success of NMR for a given system is often gauged by the quality of its ^{1}H,^{15}N-HSQC,

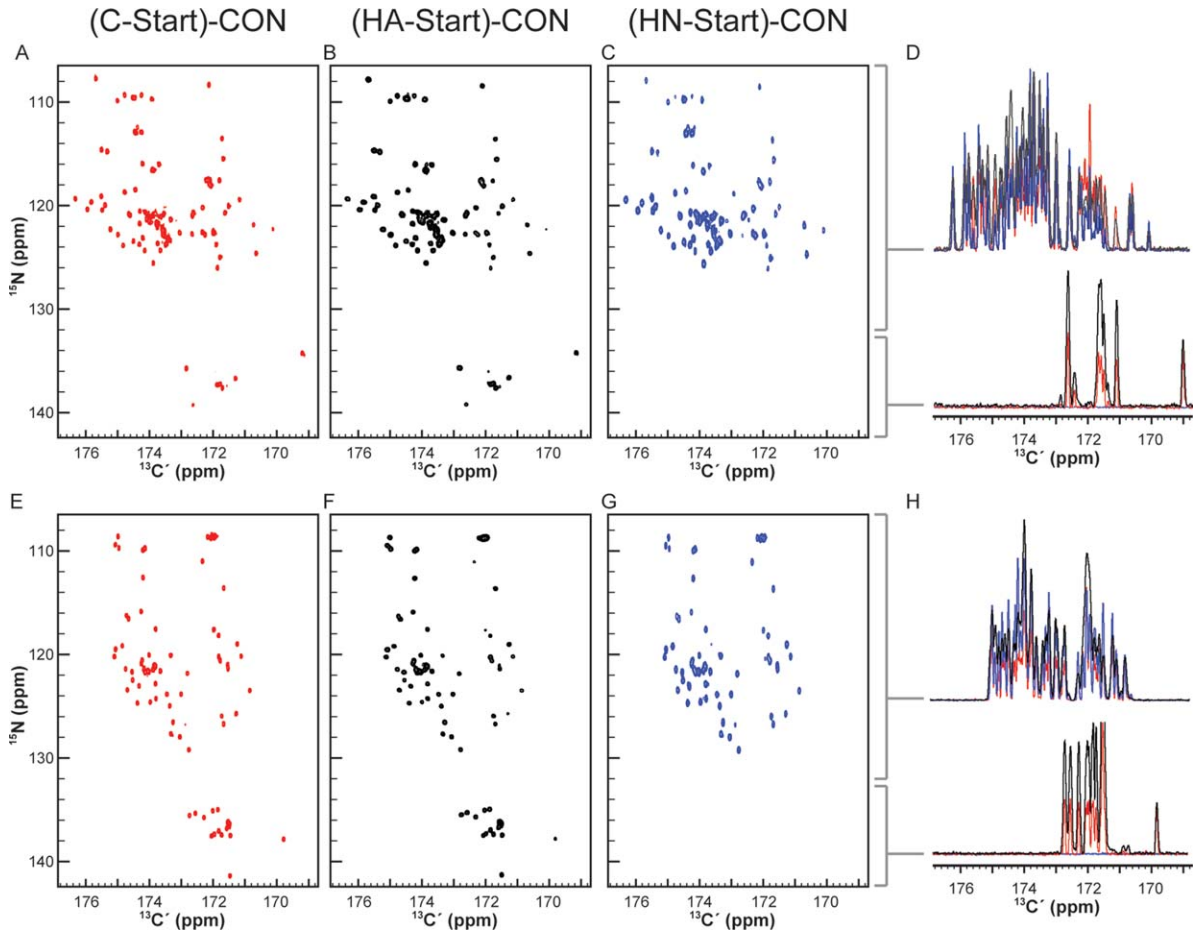

Figure 2 Three common variants of the ^{15}N,^{13}C-CON experiment offer different performance characteristics. Spectra were acquired using (A,E) C-Start CON, (B,F) HA-Start CON, and (C,G) HN-Flip CON. Spectra (A–C) were acquired using a 1.0 mM sample of ctFCP1. Spectra (E–G) were acquired using a 1.3 mM sample of Pdx1-C. Solution conditions were identical to Figure 1. Panels (D,H) display 1D-projections of the nonproline spectral region (105–132 ppm; top) and the proline spectral region (132–142 ppm; bottom). Both the 2D spectra and their 1D projections are colored to display the C-start data in red, the HA-Start data in black, and the HN-flip data in blue. All spectra were collected at 298K on a Bruker Avance III spectrometer equipped with a TCI Cryoprobe (^{1}H inner coil), operating at 500 MHz proton resonance frequency (11.7 T).

identifying an alternative NMR detection platform is preferable to terminating a worthwhile project when proton detection yields poor spectra.

In general, we find that ^{13}C-direct detect methods provide excellent signal dispersion for IDPs, while also directly providing resonances corresponding to proline residues. Both of these features are clearly observed in Fig. 1, which reports carbon-detected spectra of ctFCP1 (panels B–D) and Pdx1-C (panels F–H). Representative pulse sequences for each of these carbon-detected experiments are provided in the Appendix [Figs. A1–A5)]. For example, the ^{13}C,^{13}C-CACO [Figs. 1(D,H)], which relies on the intraresidue Cα-CO correlation, relieves much of the spectral crowding seen in the ^{1}H, ^{15}N-HSQC and yields an analogous one-to-one mapping of the protein backbone, including prolines. However, the ^{13}C,^{13}C-CACO suffers from its own limitations; the Cα dimension is often poorly dispersed, limiting its practical utility. The ^{15}N,^{13}C-CAN provides improved dispersion in both dimensions [Figs. 1(C,G)], while providing both intraresidue and inter-residue correlations for backbone assignment. However, the long refocusing delay periods needed for ^{15}N,^{13}C-CAN make it impractical to build three-dimensional (3D) experiments based on this detection platform, due to intolerable signal intensity loss through spin relaxation. Of the four experiments shown here, the ^{15}N, ^{13}C-CON is our 2D platform of choice [Figs. 1(B,F)]. This

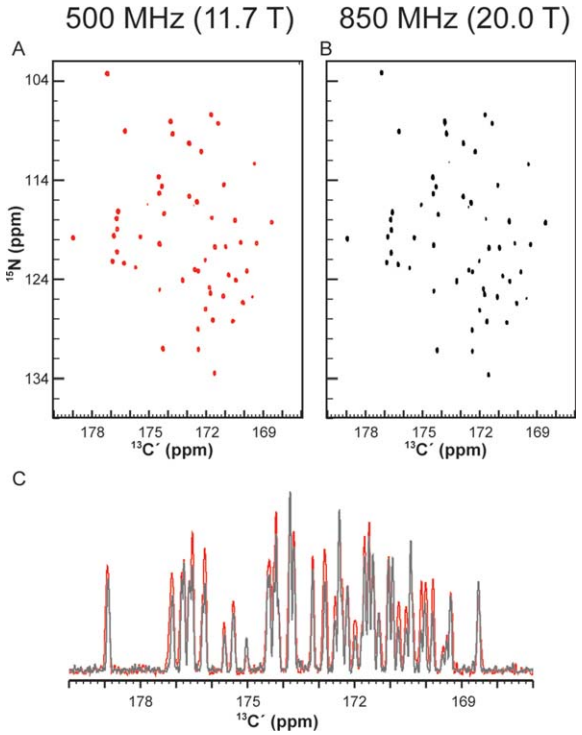

Figure 3 Carbon-detected biomolecular NMR is not practical at ultrahigh field with conventional TCI Cryoprobes (^{1}H inner coil). (A) ^{15}N,^{13}C-(HACA)CON spectrum of 500 μM GB3 collected on a Bruker Avance III spectrometer operating at 500 MHz proton resonance frequency (11.7 T). (B) ^{15}N,^{13}C-(HACA)CON spectrum of 500 μM GB3 collected on a Bruker Avance III spectrometer operating at 850 MHz proton resonance frequency (20.0 T). (C) 1D projections of the spectra in panels A and B show that the signal-to-noise for identically collected ^{15}N,^{13}C-(HACA)CON spectra is very similar for experiments collected at 11.7 and 20.0 T static magnetic field strength, with many resonances displaying greater intensity at 11.7 T.

experiment correlates the amide nitrogen with the carbonyl carbon of the preceding residue and expediently provides the best peak dispersion for IDPs. Importantly, the ^{15}N, ^{13}C-CON serves as a robust detection platform for measuring spin relaxation (*9*), side chain chemical shifts (*7*), and inter-residue correlation experiments for backbone assignment (*4*).

Within the CON suite there are three basic experiments: the (C-Start)-CON [Figs. 2(A,E)], (HA-Start)-CON [Figs. 2(B,F)], and the (HN-Start)-CON [Figs. 2(C,G)]. Representative spectra collected using this suite of ^{15}N, ^{13}C-CON experiments are shown in Fig. 2 for ctFCP1 (panels A–C) and Pdx1-C (panels E–H). All spectra reported in Fig. 2 were collected with 16 scans of signal averaging and 256 increments in the indirect dimension. As can be seen from the one-dimensional (1D) projections in Figs. 2(D,H), the traditional "protonless" carbon-start experiment suffers from reduced signal intensity for a fixed acquisition time, owing to the selection of a low gyromagnetic ratio ^{13}C-nucleus for the initial excitation (compare the red carbon-start projection to the black and blue proton-start projections). Still, for applications to proteins with paramagnetic metal centers (i.e., many metalloenzymes) the benefits of purely "protonless" spectroscopy should not be discounted.

The sensitivity gain associated with proton-start experiments, including both the HA-start [Figs. 2(B,F)] and the HN-start [Figs. 2(C,G)] is pronounced, making these spectra highly attractive for routine applications. For systems possessing a minimal number of proline residues in their sequence, the HN-start CON experiment can be an attractive option and we have used it as the basis for pulse programs to measure, for example, ^{15}N-spin relaxation with good success (*9*). Conversely, the presence of resonances for ^{15}N, ^{13}C pairs including the amide nitrogen of proline residues makes the (HA-Start)-CON the pulse program of choice for most routine applications. Of special note, the tolerance of the HA-start format for basic solution conditions (pH >7.5, discussed in detail below) makes this pulse program especially versatile.

Optimizing Magnetic Field Strength

For classical proton-detected biomolecular NMR experiments, increasing magnetic field strength is often a prudent means to improve spectral resolution, while also enhancing acquisition sensitivity. In contrast, most carbon-detected NMR experiments include obligate polarization transfers that require multiple passages through lengthy delays, such as the $^{1}J_{NC}$ refocusing delay, where the strong carbonyl carbon chemical shift anisotropy (CSA) produces efficient spin relaxation. As a consequence, many carbon-detected spectra, and the ^{15}N,^{13}C-CON especially, perform poorly at ultrahigh magnetic field strengths, compared with expectations based on experience with proton-detected NMR. For example, Fig. 3 shows the ^{15}N,^{13}C-CON spectrum of the globular protein GB3, collected using a TCI cryogenic probe at 11.7 T [Fig. 3(A)] and 20.0 T [Fig. 3(B)]. All other acquisition parameters being equivalent (e.g., number of scans, number of points in the indirect time domain), the sensitivity of the 20.0 T spectrum is not significantly enhanced compared with the 11.7 T spectrum [Fig. 3(C)], although the linewidths in the direct dimension are modestly improved at 20.0 T. In contrast, 3D and higher dimensional spectra used for chemical shift assignment are generally so severely impacted by relaxation losses at ultrahigh field as to yield little or no signal intensity above baseline, even with extensive

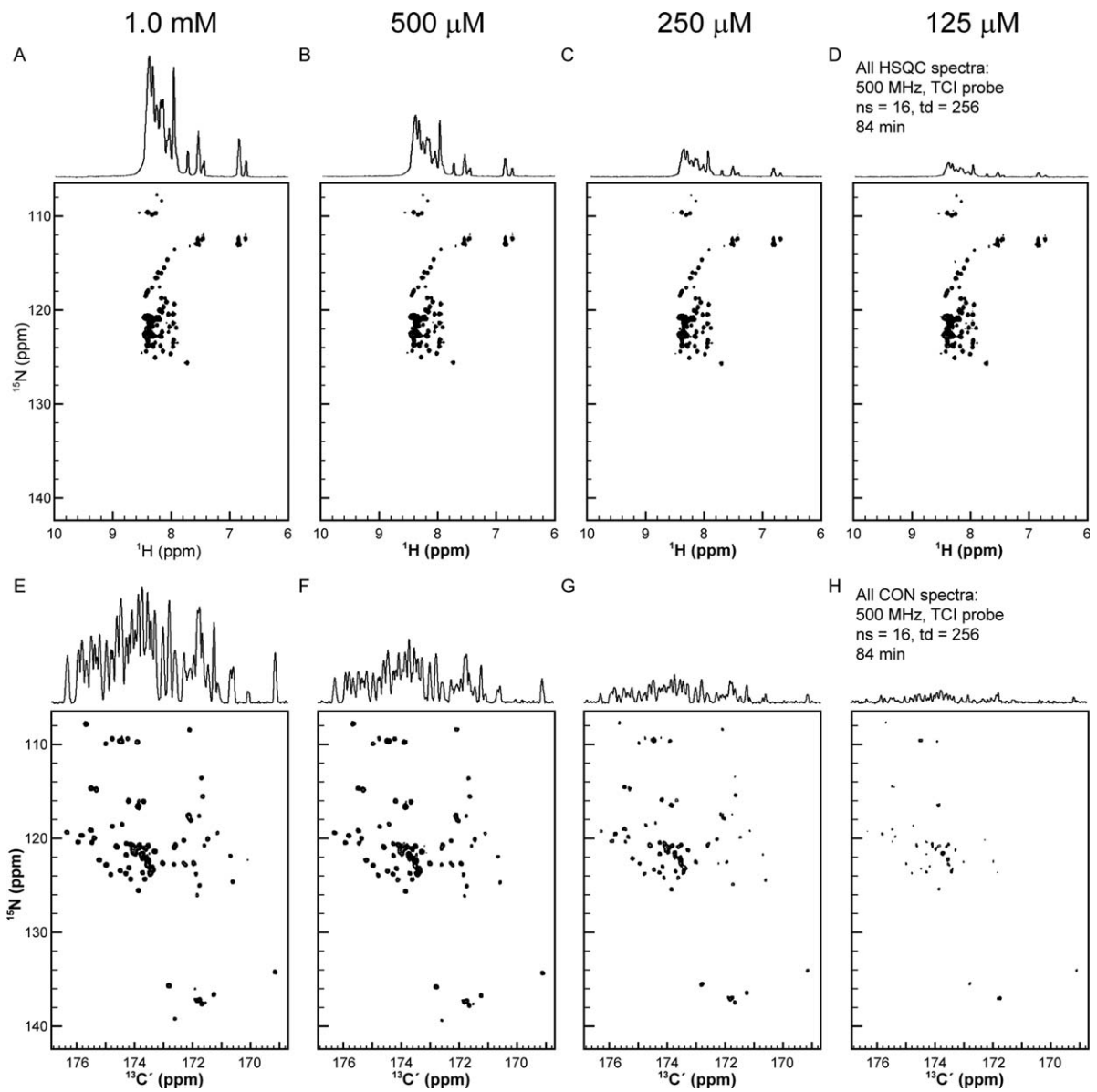

Figure 4 Carbon-detected biomolecular NMR is currently subject to more stringent sample concentration requirements than traditional proton-detected NMR. (A–D) ^{1}H,^{15}N-HSQC spectra and (E–H) ^{15}N,^{13}C-(HACA)CON spectra of ctFCP1 samples present at (A,E) 1.0 mM, (B,F) 500 μM, (C,G) 250 μM, and (D,H) 125 μM. 1D-projections of each spectrum are also provided, scaled to match the baseline noise intensity of the relevant 1.0 mM spectrum. For solution conditions, please see the legend to Fig. 1. Each spectrum was collected with 256 increments in the indirect dimension and 16 scans of signal averaging. All spectra were collected at 298K on a Bruker Avance III spectrometer equipped with a TCI Cryoprobe (^{1}H inner coil), operating at 500 MHz proton resonance frequency (11.7 T).

signal averaging (data not shown). As a general recommendation, moderate magnetic field strengths in the 11.7–16.4 T range are more likely to produce reliable spectra for carbon-detected NMR experiments, particularly where ultrasensitive carbon inner-coil probes are not available.

CONSTRAINTS ON SAMPLE CONDITIONS

A variety of constraints on data quality must be considered when planning the optimal set of NMR experiments to perform and solution conditions to use. With protein samples, solubility is often a limiting factor for

NMR applications, as concentrations in the 0.1–1.0 mM regime are typically needed. All other factors being equal, carbon-detected experiments are generally less sensitive than their proton-detected counterparts, placing an additional burden on the investigator to reach high sample concentrations when carbon-detection is to be applied. Experienced bio-NMR spectroscopists are also familiar with practical constraints on solution conditions, including total salt restrictions with cryogenically cooled probes and pH restrictions when detection is achieved through exchangeable protons. In this section, we will discuss practical guidelines for sample conditions that are likely to yield high-quality carbon-detected bio-NMR spectra. In general, we find that the restrictions on buffer composition and pH are far less stringent than for proton-detected spectroscopy, although the lower limit for sample concentration is more restrictive, absent carbon inner-coil probes.

Protein Concentration Requirements

Proton-detect experiments generally yield high quality spectra with modest sample concentrations (50–200 μM when cryogenic probes are available). However, the importance of protein concentration should not be underestimated when planning a ^{13}C-detect experiment, where the lower sensitivity of ^{13}C-detection will generally demand higher sample concentrations. In our experience, best practice requires ≥500 μM protein concentration in the NMR sample to obtain robust signal for the experiments represented in this work. This trend is easily seen in the ^{1}H,^{15}N-HSQC and ^{15}N,^{13}C-CON spectra of ctFCP1, presented in Fig. 4. Although the ^{1}H,^{15}N-HSQC spectrum is robust all the way down to 125 μM, below 500 μM many resonances are lost in the ^{15}N,^{13}C-CON. While the reduced signal at 250 μM is tolerable for the simple 2D, resonance loss from 3D spectra used for chemical shift assignments is generally too great to proceed below 500 μM (data not shown). Note that concentrations nearer to 1.0 mM may be required for some sensitivity-limited experiments, such as when measuring residual dipolar couplings (RDCs) or when performing 3D experiments with long echo periods.

Buffer Composition

The stability of IDPs at high concentrations (≥500 μM) is crucial if carbon-detected NMR is to be used for their study. As is common for all proteins, many IDPs are only stable under certain conditions (e.g., high salt, acidic or basic pH, presence of glycerol) that are not necessarily compatible with high-resolution NMR measurements, making buffer composition an important control point for experimental design. Our research

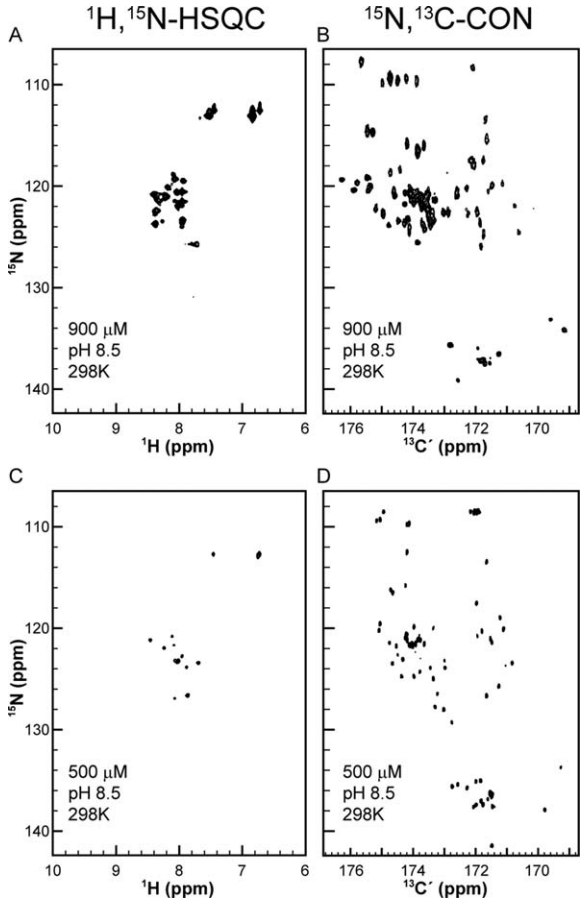

Figure 5 IDPs that are only soluble under basic pH conditions are suitable for investigation by carbon-detected NMR. (A,C) ^{1}H,^{15}N-HSQC spectra and (B,D) ^{15}N,^{13}C-(HACA)CON spectra. (A,B) ctFCP1 spectra were collected on a 900 μM sample in 20 mM Tris, pH 8.5 with 100 mM NaCl. (C,D) Pdx1-C spectra collected on a 500 μM sample in 50 mM Tris, pH 8.5 (adjusted after TCEP addition), with 50 mM KCl, 5 mM TCEP. All spectra were collected at 298K on a Bruker Avance III spectrometer equipped with a TCI Cryoprobe (^{1}H inner coil), operating at 500 MHz proton resonance frequency (11.7 T).

program has been built around several IDPs that require widely varying buffer compositions to reach the concentrations needed. Here, we will survey our findings, which are best summarized by reassuring the interested spectroscopist that concentration requirements are, by far, the most restrictive factor in designing an IDP-based research program around carbon-detected NMR.

Many bio-NMR spectroscopists will be familiar with the long held caution that cryogenic probes create severe restrictions on total salt concentration allowable, because tuning and matching becomes challenging as salt concentration grows above >>500 mM. As many IDPs are highly charged polyampholytes or polyelectrolytes, some require higher

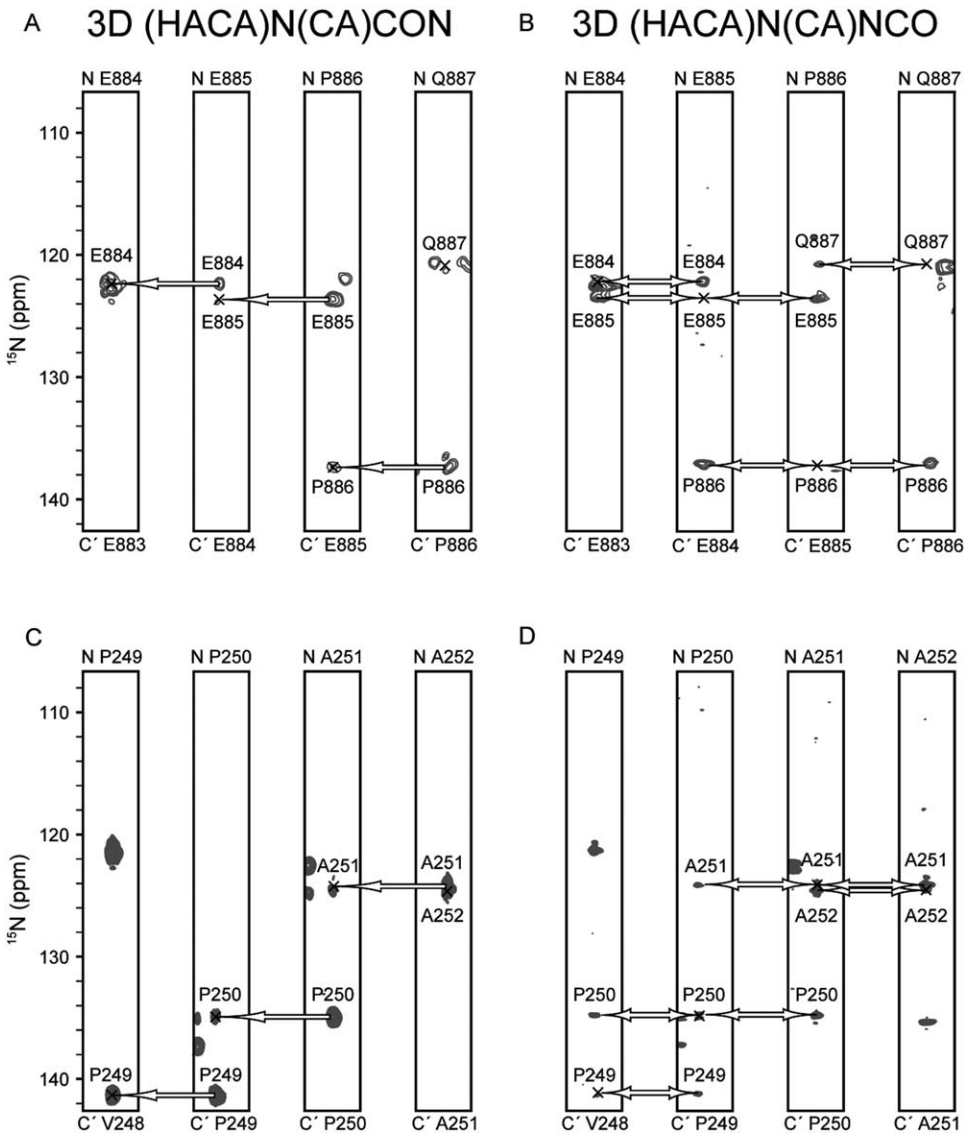

Figure 6 Carbon-detected NMR provides a simple method for backbone assignment of proline-rich IDPs and proteins soluble under basic pH conditions. (A,C) (HACA)N(CA)CON strip plots and (B,D) (HACA)N(CA)NCO strip plots corresponding to 1 mM ctFCP1 (A,B) and 1 mM Pdx1-C (C,D). Solution conditions matched those in Fig. 5. The (HACA)N(CA)CON spectra were collected with 64 × 128 increments in the indirect dimension and 16 scans of signal averaging. The (HACA)N(CA)NCO spectra were collected using 30% nonuniform sampling with 64 × 128 increments in the indirect dimensions and 32 scans of signal averaging. All spectra were collected at 298K on a Bruker Avance III spectrometer equipped with a TCI Cryoprobe (^{1}H inner coil), operating at 500 MHz proton resonance frequency (11.7 T).

than normal salt concentrations to reach stability. While the longer ^{1}H-pulses needed at higher salt do lead to less efficient excitation in H-start CON experiments, detection on carbon dramatically reduces the negative effects of salt on receiver performance. In our experience, carbon-detected experiments such as the (HACA)-CON are readily collected on samples containing total monovalent salt concentrations in excess of 1.0 M. For example, (HACA)-CON spectra of Pdx1-C collected with 1.0 M NaCl added to the sample display a ~20% reduction in signal-to-noise, relative to the conditions in Fig. 3, but otherwise display no adverse effects (data not shown). If salt concentration does become a concern, then reverting to traditional "protonless" carbon-detected strategies presents a viable work-around.

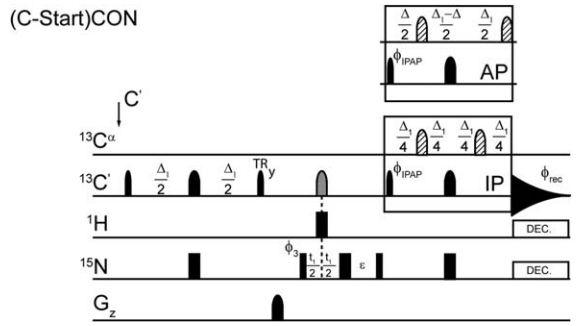

Figure A1 Pulse sequence for the CON-IPAP experiment. The delays are $\Delta = 9.0$ ms, $\Delta_1 = 25$ ms, and $\varepsilon = t_1(0) + \text{pC180}$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_3 = x, -x$; $\varphi_{\text{IPAP}}(\text{IP}) = x$; $\varphi_{\text{IPAP}}(\text{AP}) = -y$; $\varphi_{\text{rec}} = x, -x$. Quadrature detection in the indirect dimension is obtained by States-TPPI incrementation of φ_3.

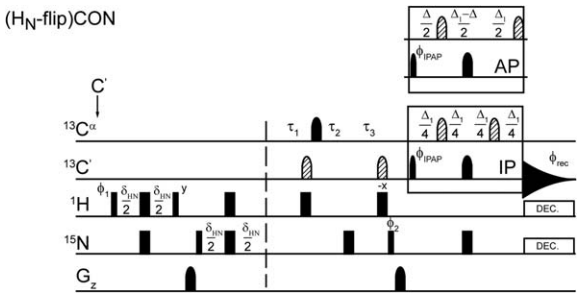

Figure A3 Pulse sequence for the (H_N-flip)CON-IPAP experiment. The delays are $\delta_{\text{NH}} = 4.6$ ms, $\Delta = 9.0$ ms, and $\Delta_1 = 25$ ms. ^{15}N chemical shift evolution is measured during t_1 using a semiconstant time period with delays $\tau_1 = (\Delta_1 + t_1)/2$, $\tau_2 = (1 - \Delta_1/t_{1,\text{max}})t_1/2$, $\tau_3 = (1 - t_1/t_{1,\text{max}})\ \Delta_1/2$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = x, -x$; $\varphi_2 = y, y, -y, -y$; $\varphi_{\text{IPAP}}(\text{IP}) = 4(x), 4(-x)$; $\varphi_{\text{IPAP}}(\text{AP}) = 4(-y), 4(y)$; $\varphi_{\text{rec}} = 2(x, -x, -x, x)$. Quadrature detection in the indirect dimension is obtained by States-TPPI incrementation of φ_2.

Salt content is not the only buffer consideration that matters when designing bio-NMR solution conditions. Solvent-exposed backbone amides are subject to efficient proton-exchange with solvent and the rate of exchange is enhanced in mildly basic conditions (pH 7.5–8.5), compared with the rate of exchange under mildly acidic conditions (pH 5.5–6.5). As IDPs tend to have more solvent-exposed backbones than their cooperatively folded counterparts, the number of amides subject to efficient chemical exchange is dramatically increased, leading to well-documented difficulties applying bio-NMR spectroscopy to IDPs that require basic solution conditions to reach sufficient solubility. The loss of resonances to chemical exchange with solvent is readily observed in Fig. 5, where ^{1}H,^{15}N-HSQC spectra, collected in buffer at pH 8.5 are presented for ctFCP1 [Fig. 5(A)] and Pdx1-C [Fig. 5(C)]. Comparison with the equivalent spectra in Figs. 1(A,E), collected at pH 7.0 and 6.5, respectively, the loss of nearly all resonances from the spectra is readily evident. In contrast, ^{15}N,^{13}C-(HACA)CON spectra of ctFCP1 [Fig. 5(B)] and Pdx1-C [Fig. 5(D)] collected at pH 8.5 are of almost identical spectral quality to those collected at neutral or mildly acidic pH. Therefore, so long as ^{13}C-start or aliphatic ^{1}H-start experiments are selected, carbon-detected NMR strategies should be robust to nearly any conditions of pH needed to ensure the solubility and stability of the NMR construct.

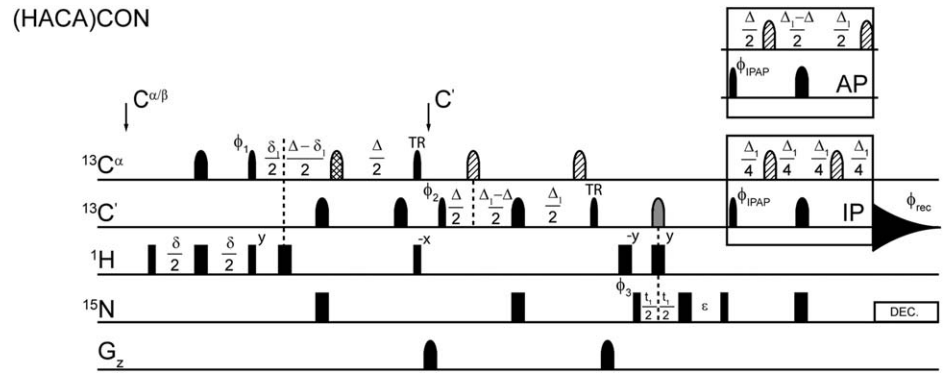

Figure A2 Pulse sequence for the (HACA)CON-IPAP experiment. The delays are $\delta = 3.6$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9.0$ ms, $\Delta_1 = 25$ ms, and $\varepsilon = t_1(0) + \text{pC180}$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = 4(x), 4(-x)$; $\varphi_2 = 2(x), 2(-x)$; $\varphi_3 = x, -x$; $\varphi_{\text{IPAP}}(\text{IP}) = x$; $\varphi_{\text{IPAP}}(\text{AP}) = -y$; $\varphi_{\text{rec}} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection in the indirect dimension is obtained by States-TPPI incrementation of φ_3.

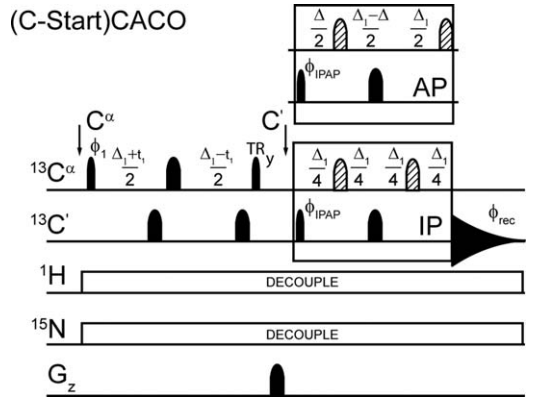

Figure A4 Pulse sequence for the CACO-IPAP experiment. The delays are $\Delta = 9.0$ ms, $\Delta_1 = 25$ ms. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = x, -x$; $\varphi_{IPAP}(IP) = 2(x), 2(-x)$; $\varphi_{IPAP}(AP) = 2(-y),2(y)$; $\varphi_{rec} = x, -x, -x, x$. Quadrature detection in the indirect dimension is obtained by States-TPPI incrementation of φ_1.

Finally, so long as they are not isotope-enriched, additives such as glycerol or biological buffers including Tris, HEPES, and other popular choices should also impose no limitations on signal quality. Carbon-detected NMR is resilient to high salt concentrations and a wide range of pH, although $^1H^N$-start experiments are subject to many of the same limitations as ^1H,^{15}N-HSQC spectroscopy. In summary, the buffer composition needed to achieve a highly concentrated sample for carbon-detected bio-NMR of disordered proteins should rarely be a limiting factor.

CHEMICAL SHIFT ASSIGNMENT STRATEGIES

Unambiguous chemical shift assignment is crucial for interpretation of all NMR spectra, but this step is not always trivial for IDPs. Accordingly, multidimensional carbon-detect approaches have been developed for backbone resonance assignment of IDPs, which are generally more successful than traditional proton-detected strategies for these systems. Several reports of high dimensional (4D and 5D) carbon-detected experiments for chemical shift assignment have appeared in the recent literature (*10,11*), to which we refer the interested reader. In this section, we discuss the 2D and 3D experimental suites generally used in our laboratory, the effects of basic solution pH on some of the available pulse programs, and finally our approach to assigning the backbone resonances of IDPs under basic solution conditions. While learning a new experimental paradigm may seem daunting at first, the reality is that most carbon-detected NMR protocols for chemical shift assignment use the same strategies as their proton-detected counterparts. In most cases, pairs of experiments designed to build nearest-neighbor pairs of spin systems are collected for the purpose of "walking" along the backbone.

An Efficient Method for Proteins Requiring Basic pH

In addition to spectral overlap, another major limitation to using traditional proton-detected experiments for IDPs is pH, particularly where basic pH (>7.5) is required for solubility and stability. Carbon-detect experiments offer a convenient way to overcome reduced spectral quality due to chemical exchange commonly encountered for many IDPs. For example, we have previously reported a strategy that combines $^1H^N$-flip N(CA)CON and N(CA)NCO 3D-spectra with a suite of carbon-detected amino-acid specific 2D spectra in an efficient protocol for chemical shift assignment of IDPs solubilized under mildly acidic conditions (*4*).Modification of our previous protocol by incorporating the (HACA)-start element into the 3D N(CA)CON and N(CA)NCO is straight forward and allows applications to IDPs that are only soluble under basic pH conditions. Pulse sequences incorporating this modification are provided in the Appendix [Figs. A6 and A7] and representative strip-plots for both ctFCP1 and Pdx1-C are provided in Fig. 6. In the 3D (HACA)N(CA)CON spectrum, the amide-nitrogen and carbonyl-carbon chemical shifts of each residue are correlated with the amide-nitrogen chemical shift of the preceding residue, generating the *i to i* – 1 connectivity needed for the walk along the backbone approach to assignment. For highly repetitive or otherwise difficult sequences, the 3D (HACA)N(CA)NCO relieves ambiguity by establishing bidirectional connectivity between each ^{15}N,^{13}C spin-pair in the CON and the amide nitrogen chemical shifts of both its N-terminal ($i - 1$) and C-terminal ($i + 1$) partners in the chain. Using a 30% sparse nonuniform sampling schedule on the 3D (HACA)N(CA)NCO, both of these 3D experiments can be collected in a total of ~5 spectrometer-days, allowing cost-effective acquisition.

Of particular note for proline-rich IDPs, when backbone chemical shift assignments are generated from the (HACA)-start 3D experiments, unambiguous assignment through proline residues becomes straightforward. Even for Pdx1-C, for which 22% of the amino acid residues in our construct are proline, use of the strategy we have discussed here readily yielded nearly complete backbone chemical shift assignment (*4*). It is worth noting that, for ^{15}N,^{13}C spin-pairs involving proline nitrogens, the (HACA)N(CA)NCO

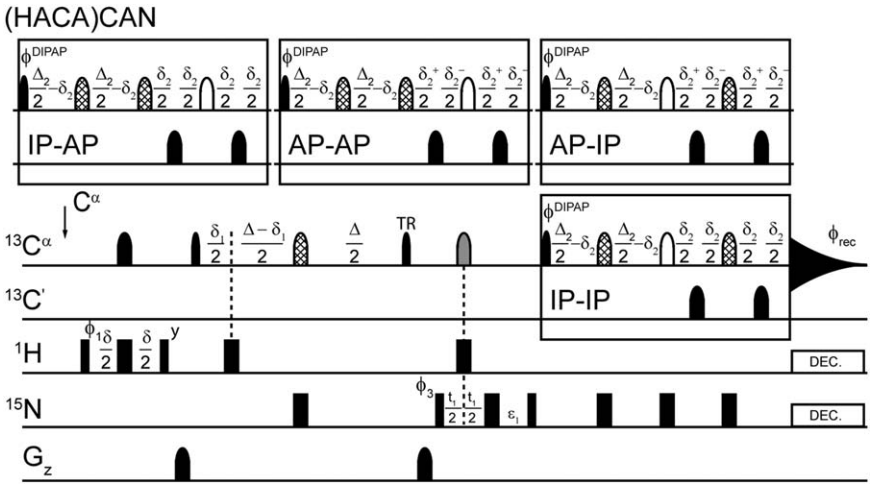

Figure A5 Pulse sequence for the (HACA)CAN-IPAP experiment. The delays are $\delta = 3.6$ ms, $\delta_1 = 2.2$ ms, $\delta_2 = 7.2$ ms, $\delta_2^+ = 5.85$ ms, $\delta_2^- = 1.35$ ms, $\Delta = 9.0$ ms, $\Delta_2 = 24.8$ ms, and $\varepsilon = t_1(0) + \text{pC180}$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = 2(x), 2(-x)$; $\varphi_3 = x, -x$; $\varphi_{\text{DIPAP}}(\text{IP,IP}) = 4(x), 4(-x)$; $\varphi_{\text{DIPAP}}(\text{AP,AP}) = 4(x), 4(-x)$; $\varphi_{\text{DIPAP}}(\text{AP,IP}) = 4(-y), 4(y)$; $\varphi_{\text{DIPAP}}(\text{IP,AP}) = 4(y), 4(-y)$; $\varphi_{\text{rec}} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection in the indirect dimension is obtained by States-TPPI incrementation of φ_3.

auto-correlation peaks are systematically phase-shifted by 90°, relative to the rest of the resonances in the spectrum. As these resonances are generally well resolved and self-evident (they are auto-correlations), this is generally not a problem. In conjunction with carbon-detected amino-acid specific 2D spectroscopy [so-called CAS-NMR (*12*)] these two spectra offer a straightforward strategy for backbone chemical shift assignment of IDPs under nearly any set of reasonable solution conditions. Further acquisition of 3D C_CCCON (*6*) and C_H(CC)CON (*7*) spectra yields complete aliphatic chemical shift assignment as well, providing a robust starting point for structure, dynamics, and functional studies of the target IDP.

CONCLUSIONS

NMR is a powerful tool for probing the structure and function of IDPs. The carbon-detect experiments discussed here and in the cited work provide an efficient method for generating IDP chemical shift assignments as a prerequisite for structure–function studies. In this

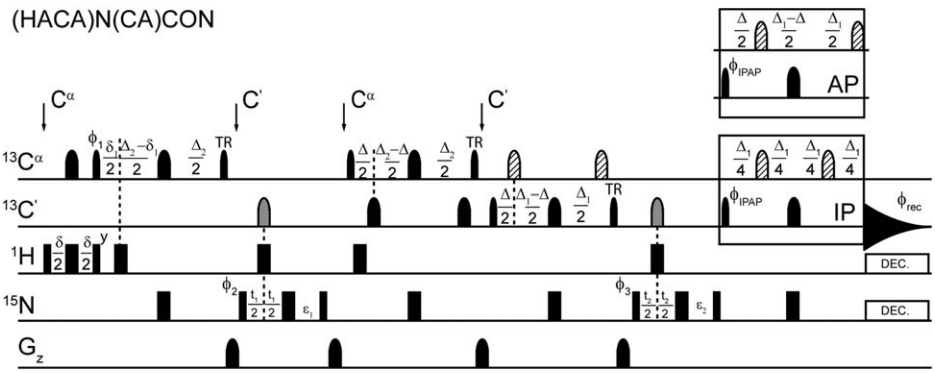

Figure A6 Pulse sequence for the (HACA)N(CA)CON-IPAP experiment. The delays are $\delta = 3.6$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9.0$ ms, $\Delta_1 = 32$ ms, $\Delta_2 = 26$ ms, $\varepsilon_1 = t_1(0) + \text{pC180}$, and $\varepsilon_2 = t_2(0) + \text{pC180}$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = 4(x), 4(-x)$; $\varphi_2 = 2(x), 2(-x)$; $\varphi_3 = x, -x$; $\varphi_{\text{IPAP}}(\text{IP}) = x$; $\varphi_{\text{IPAP}}(\text{AP}) = -y$; $\varphi_{\text{rec}} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection in the t_1 and t_2 dimension is obtained by States-TPPI incrementation of φ_2 and φ_3, respectively.

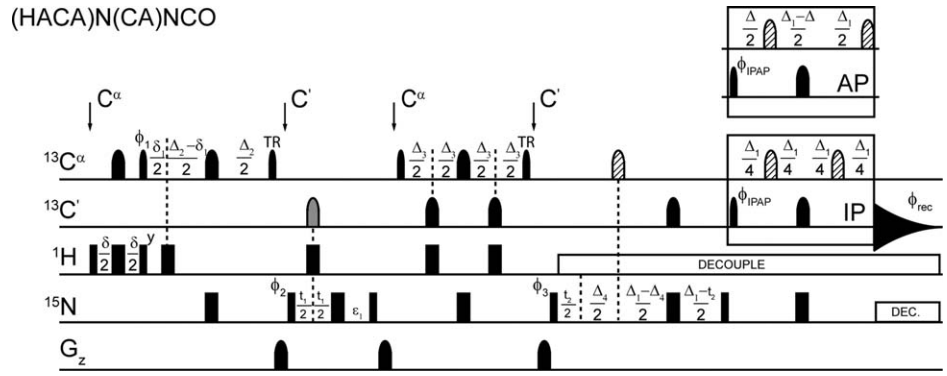

Figure A7 Pulse sequence for the (HACA)N(CA)NCO-IPAP experiment. The delays are $\delta = 3.6$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9.0$ ms, $\Delta_1 = 32$ ms, $\Delta_2 = 26$ ms, $\Delta_3 = 25$ ms, $\Delta_4 = 24.2$ ms, $\varepsilon_1 = t_1(0) + \text{pC180}$, and $\varepsilon_2 = t_2(0) + \text{pC180}$. All pulses are applied with x-phase unless otherwise indicated. The phase cycle is $\varphi_1 = 4(x), 4(-x)$; $\varphi_2 = 2(x), 2(-x)$; $\varphi_3 = x, -x$; $\varphi_{\text{IPAP}}(\text{IP}) = x$; $\varphi_{\text{IPAP}}(\text{AP}) = -y$; $\varphi_{\text{rec}} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection in the t_1 and t_2 dimension is obtained by States-TPPI incrementation of φ_2 and φ_3, respectively.

review, we have discussed best practices for establishing viable experimental conditions and selecting pulse programs to overcome pH-related losses of spectral quality. This protocol should be both efficient and highly robust to the conditions interested investigators are likely to encounter. Our recommendation is that spectroscopists begin their carbon-detected studies of IDPs by optimizing the 2D ^{15}N,^{13}C-CON for their system, where the IDP must be soluble and stable at a minimum concentration of 500 μM; in general, it is less important to worry about pH or total salt concentration. Once these conditions are met, established pulse programs or the new (HACA)-start assignment strategy introduced here may be used to rapidly begin what is sure to be an exciting new project.

ACKNOWLEDGMENTS

This work was supported by an NSF-CAREER award (MCB-0953918) to SAS and an NIH predoctoral fellowship (F31GM101936) to MB. All spectra were collected in the Lloyd Jackman NMR Facility in the Department of Chemistry at the Pennsylvania State University. We thank Dr. Emmanuel Hatzakis for assistance with the instruments.

APPENDIX

Timing diagrams for the seven carbon-detected pulse sequences used to generate the data shown in this paper are presented here as follows: CON-IPAP (Fig. A1), (HACA)CON-IPAP (Fig. A2), (HN-flip)CON-IPAP (Fig. A3), CACO-IPAP (Fig. A4), (HACA)CAN-IPAP (Fig. A5), (HACA)N(CA)CON-IPAP (Fig. A6), and (HACA)N(CA)NCO-IPAP (Fig. A7). The standard c_con_iasq and c_caco_ia pulse sequences from the Topspin 3.2 pulse program library were used to collect all CON-IPAP and CACO-IPAP spectra, respectively. Pulse sequences for the remaining experiments were written in house and are available free of charge through the Penn State Scholar Sphere (https://scholarsphere.psu.edu) or through our laboratory website (http://research.chem.psu.edu/sasgroup/pulseprograms.html).

The figures presented here depict pulses as either narrow rectangles or shapes to represent 90° pulses, or wide rectangles or shapes to represent 180° pulses. In all cases, the pulses are applied at the frequencies indicated on the left of each line. Typical hard ^{1}H and ^{15}N pulse times of 10 and 31 μs, respectively, were used in all experiments. All pulsed field gradients were applied in the Z-direction (G_z) for 1 ms with a sine shape. Solid black pulses on ^{13}C represent standard band-selective Q5 or time-reversed Q5 pulses for excitation and Q3 pulses for inversion, with durations of 384 and 307 μs, respectively, at 11.7 T. Pulses filled with diagonal lines are off-resonance 180° Q3-shaped pulses centered on either ^{13}C′ or ^{13}C$^{\alpha}$ as indicated. Pulses filled with solid white in the (HACA)-CAN-IPAP are off-resonance 180° Q3-shaped pulses centered in the aliphatic region at 49 ppm. Pulses filled in solid gray indicate a band selective adiabatic inversion of both ^{13}C′ and ^{13}C$^{\alpha}$ through a 500-ms CHIRP pulse with 60 Hz sweep and 25% smoothing. Pulses filled with hashed lines are higher-selectivity 180° Q3-shaped pulses applied on resonance with 1,200 μs duration at 11.7 T. The ^{1}H and ^{15}N carriers were placed at 4.7 and 124 ppm, respectively. The ^{13}C carrier was set as indicated by the vertical arrows to either ^{13}C$^{\alpha}$ = 54 ppm or ^{13}C′ = 172 ppm.

Composite pulse decoupling of ^{1}H and ^{15}N was achieved by the use of 3.57 kHz Waltz 65 and 1.25 kHz Garp sequences, respectively. All other relevant acquisition parameters are described in the respective figure legends.

REFERENCES

1. Jensen MR, Ruigrok RW, Blackledge M. 2013. Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 23:426–435.
2. Showalter SA. 2014. Intrinsically disordered proteins: methods for structure and dynamics Studies. EMagRes 3:181–189.
3. Lawrence CW, Bonny A, Showalter SA. 2011. The disordered C-terminus of the RNA polymerase II phosphatase fcp1 is partially helical in the unbound state. Biochem Biophys Res Commun 410:461–465.
4. Sahu D, Bastidas M, Showalter SA. 2014. Generating NMR chemical shift assignments of intrinsically disordered proteins using Carbon-detected NMR methods. Anal Biochem 449:17–25.
5. Bermel W, Felli IC, Kümmerle R, Pierattelli R. 2008. 13C Direct-detection biomolecular NMR. Concepts Magn Reson Part A 32A:183–200.
6. Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R. 2006. ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Prog Nucl Magn Reson Spectrosc 48:25–45.
7. O'Hare B, Benesi AJ, Showalter SA. 2009. Incorporating ^{1}H-chemical shift determination into ^{13}C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358.
8. Bermel W, Bertini I, Felli IC, Pierattelli R. 2009. Speeding up 13C direct detection biomolecular NMR spectroscopy. J Am Chem Soc 131:15339–15345.
9. Lawrence CW, Showalter SA. 2012. Carbon-detected N-15 NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with rap74. J Phys Chem Lett 3:1409–1413.
10. Novacek J, Haba NY, Chill JH, Zidek L, Sklenar V. 2012. 4D non-uniformly sampled HCBCACON and (1)J(NCalpha)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148.
11. Bermel W, Bertini I, Felli IC, Gonnelli L, Kozminski W, Piai A, et al. 2012. Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301.
12. Bermel W, Bertini I, Chill J, Felli IC, Haba N, Kumar MVV, et al. 2012. Exclusively heteronuclear (13) C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). Chembiochem 13:2425–2432.

BIOGRAPHIES

Monique Bastidas graduated with a B.S. degree in Biochemistry in 2010 from California State University, Sacramento, where she worked on the development of an analytical technique to determine the composition of glycans. In 2010, she began her graduate career at Penn State, where her thesis focuses on characterizing structure–function relationships for the intrinsically disordered transcription factor, Pdx1, using biophysical techniques. Broadly, her research interests are related to the study of the structure and interactions of biomolecules.

Eric B. Gibbs received his bachelor's degree (B.A.) in chemistry from Temple University. During this time, he was awarded an NIH T34 fellowship, and worked under Dr. Daniel Strongin and Dr. Ian Blair at the University of Pennsylvania. Eric matriculated to Penn State in 2012 and was awarded an Alfred P. Sloan MPHD award. His research focuses on understanding the effects of post-translational modifications on the structure and function of intrinsically disordered proteins.

Debashish Sahu received his bachelor's degree (B. Tech) in industrial biotechnology from Anna University, India and PhD degree in Molecular Biophysics from University of Texas Medical Branch at Galveston under the supervision of Dr. Junji Iwahara. During his graduate work on the study of NMR study of kinetics of protein translocation on DNA, he was awarded Who's who among students in American Universities and Colleges, Curtis W. Lambert Scholarship, and Robert A. Welch Award for Excellence in Graduate Research in Chemistry. His current research focus is the development and use of carbon detected NMR methods to characterize IDPs.

Scott A. Showalter is an Associate Professor of Chemistry at Penn State, where he has been a member of the faculty since 2008. Previously, Dr. Showalter was an NSF predoctoral fellow in the Molecular Biophysics Ph.D. program at Washington University in St. Louis, MO and an NIH NRSA postdoctoral fellow at the National High Magnetic Field Laboratory in Tallahassee, FL. In 2012, Dr. Showalter was the recipient of the Eastern Analytical Symposium New Faculty Award in NMR Spectroscopy in recognition of his work advancing applications of carbon-detected solution NMR spectroscopy to biomolecules.