Analytical Biochemistry 449 (2014) 17–25

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods

Debashish Sahu, Monique Bastidas, Scott A. Showalter *

Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 26 November 2013
Accepted 2 December 2013
Available online 10 December 2013

Keywords:
Intrinsically disordered proteins
Carbon detect NMR
Amino acid-filtered NMR
CAS–NMR
Pdx1

ABSTRACT

There is an extraordinary need to describe the structures of intrinsically disordered proteins (IDPs) due to their role in various biological processes involved in signaling and transcription. However, general study of IDPs by NMR spectroscopy is limited by the poor ^{1}H amide chemical shift dispersion typically observed in their spectra. Recently, ^{13}C direct-detected NMR spectroscopy has been recognized as enabling broad structural study of IDPs. Most notably, multidimensional experiments based on the ^{15}N,^{13}C CON spectrum make complete chemical shift assignment feasible. Here we document a collection of NMR-based tools that efficiently lead to chemical shift assignment of IDPs, motivated by a case study of the C-terminal disordered region from the human pancreatic transcription factor Pdx1. Our strategy builds on the combination of two three-dimensional (3D) experiments, (H_N-flip)N(CA)CON and 3D (H_N-flip)N(CA)NCO, that enable daisy chain connections to be built along the IDP backbone, facilitated by acquisition of amino acid-specific ^{15}N,^{13}C CON-detected experiments. Assignments are completed through carbon-detected, total correlation spectroscopy (TOCSY)-based side chain chemical shift measurement. Conducting our study required producing valuable modifications to many previously published pulse sequences, motivating us to announce the creation of a database of our pulse programs, which we make freely available through our website.

Intrinsically disordered proteins (IDPs)[1] partially or completely lack a cooperatively folded structure under native conditions. Even so, intrinsically disordered regions are abundant in proteins executing various biological processes requiring protein–protein interactions such as molecular signaling and transcription [1]. This prevalent functional role for intrinsic disorder led Wright and Dyson to recognize that protein function derives from native structure regardless of whether or not folding occurs [2]. Therefore, even in the absence of spatial and temporal order, native structure must always be considered significant. Recent NMR-derived structure models of disordered regions from the tumor suppressor protein p53 [3] and the cyclin-dependent kinase inhibitor Sic1 [4] have provided critical insight into the functional role of protein disorder for these systems. Success stories like these are not common, however, because most IDPs yield solution NMR spectra characterized by severe spectral overlap that precludes high-resolution study. Recent work done in our laboratory demonstrates that the problem with studying IDPs is not inherent to the proteins but rather rests on the decision to employ NMR methods that have been optimized for cooperatively folding systems to evaluate disordered systems [5–8]. In the course of our efforts to systematically study IDP structure–function relationships by NMR, we have established a purely carbon-detected NMR protocol for generating IDP chemical shift assignments, which we demonstrate here through application to the C-terminal disordered region of the pancreatic duodenal homeobox protein 1 (Pdx1-C).

Solution NMR spectroscopy, which is well suited to studying biopolymers possessing heterogeneous conformational ensembles, stands as the most frequently successful method for IDP structure analysis. Two-dimensional (2D) heteronuclear correlation experiments have become a mainstay of biomolecular NMR spectroscopy because they generally provide good peak dispersion, serve readily as a component for building multidimensional experiments that yield chemical shifts and structure constraints, and provide a valuable detection platform for the measurement of spin relaxation, scalar couplings, or dipolar couplings. In particular, the ^{1}H,^{15}N heteronuclear single quantum coherence (HSQC) has become a workhorse experiment for protein NMR because it provides nominally one-to-one mapping between resonances and residues (with the absence of proline being the major exception) and can

* Corresponding author. Fax: +1 814 865 3314.
E-mail address: sas76@psu.edu (S.A. Showalter).

[1] *Abbreviations used:* IDP, intrinsically disordered protein; Pdx1-C, C-terminal disordered region of the pancreatic duodenal homeobox protein 1; 2D, two-dimensional; HSQC, heteronuclear single quantum coherence; 3D, three-dimensional; BMRB, Biological Magnetic Resonance Bank; DTT, dithiothreitol; SSP, secondary structure propensity; CAS, carbon-detected amino-acid-selective; IPAP, in phase anti-phase.

0003-2697/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ab.2013.12.005

be rapidly acquired with inexpensive samples. For most small and cooperatively folding proteins, the ^{1}H,^{15}N HSQC features an extremely well-dispersed set of resonances. In contrast, IDPs almost always display problematically finite chemical shift dispersion in the proton dimension of this experiment. This problem is seen quite clearly in the ^{1}H,^{15}N HSQC of Pdx1-C (Fig. 1A), for which the extreme spectral crowding in the random coil region (near 8.5 ppm on the ^{1}H axis) makes it impossible to unambiguously identify most individual resonances in the spectrum.

The principal advantage to choosing the ^{1}H,^{15}N HSQC as a detection platform for protein NMR is that it provides a simple spectrum with one 2D resonance per amino acid residue with the exception of proline. Similar mapping is achievable using, for example, the $^{13}C'$–^{15}N or $^{13}C'$–$^{13}C_{\alpha}$ as a reporting pair, but with the consequence that ^{13}C is a far less sensitive detection nucleus than ^{1}H. Fortunately, recent improvements in instrumentation, most notably the development of stable cryogenically cooled probes, has largely mitigated this concern, rendering carbon-detected biomolecular NMR practical for routine applications. This, coupled with intense interest in the inner workings of IDPs, has propelled a sudden burst of development in the field of carbon-detected biomolecular NMR, which can be used to report on the structural aspects of IDPs in atomic detail [5,8–13]. These carbon-detected methods make the study of IDPs feasible because the resulting 2D heteronuclear correlation spectra generally feature excellent chemical shift dispersion in both dimensions. In particular, we and others have found the ^{15}N,^{13}C CON spectrum, which records the correlation between the carbonyl carbon and amide nitrogen in the peptide plane of the protein backbone, to be a generally successful detection platform for IDPs. For Pdx1-C, the degree of improvement over conventional ^{1}H,^{15}N HSCQ detection is clearly evident in Fig. 1B.

Successful investigation of biomolecular structure and dynamics by NMR always rests on the comprehensive and accurate generation of chemical shift assignments for the system of interest. The excellent quality of ^{15}N,^{13}C CON spectra, like the one shown for Pdx1-C in Fig. 1B, suggests that achieving this goal should be straightforward using carbon-detected NMR (which is demonstrated by the full-page and fully annotated ^{15}N,^{13}C CON of Pdx1-C in Fig. S1 of the online Supplementary material). Even so, few efficient and generally practical protocols for generating chemical shift assignments, based purely on the collection of carbon-detected spectra, have been reported to date. Here we report a protocol, based on a suite of four three-dimensional (3D) experiments and additional amino acid-specific 2D spectra [5,9,11], that efficiently yields chemical shift assignments for IDPs that are soluble and stable under mildly acidic conditions. Our recommended strategy is based completely on carbon-detected spectra and yet required only 2 weeks of NMR time to complete for Pdx-C (BMRB [Biological Magnetic Resonance Bank] accession no. 19596). Furthermore, acquisition does not rely on high-dimensional (i.e., four- or five-dimensional) experiments or on nonuniform sampling, making it accessible to users without access to these techniques. Finally, our analysis strategy builds on traditional nearest neighbor daisy chaining through the protein backbone that will be familiar to anyone who has previously assigned a protein using conventional triple resonance methods [14,15]. To facilitate broader adoption of this highly efficient strategy, we have made our modified versions of each pulse program used available through our website (http://research.chem.psu.edu/sasgroup/pulseprograms.html) along with representative parameter sets.

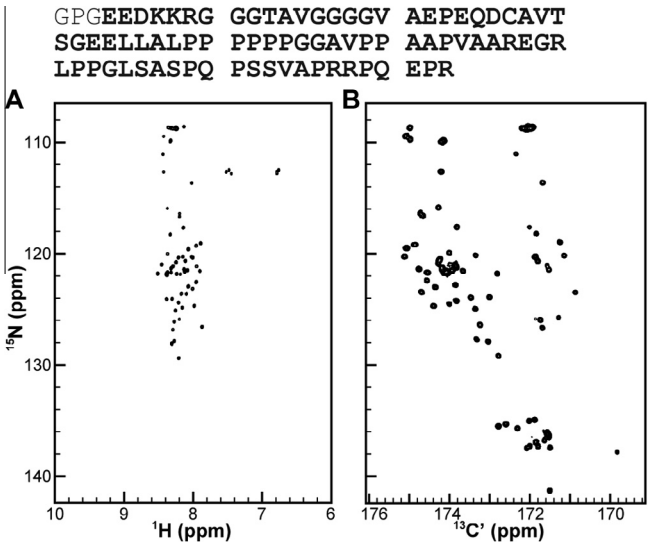

Fig.1. Two-dimensional heteronuclear correlation spectra provide the foundation for all biomolecular NMR studies, but traditional ^{1}H-detected methods do not work as well for IDPs as newer ^{13}C-detected methods. (A) The ^{1}H,^{15}N HSQC of Pdx1-C. (B) The ^{15}N,^{13}C CON of Pdx1-C. Of special value is the presence of peaks corresponding to proline residues, seen downfield shifted in the nitrogen dimension of the ^{15}N,^{13}C CON. Note that the amino acid sequence of Pdx1-C is provided at the top of the figure, with the N-terminal GPG cloning artifact shown in the thinner font. A larger version of panel B is provided in Fig. S1 of the Supplementary material, including complete annotation of chemical shift assignments.

Materials and methods

Purification of PDX1-C

A synthetic Pdx1 gene with codons optimized for *Escherichia coli* was purchased from Geneart, and the Pdx1 C terminus (amino acids 204–283 of the human sequence, subsequently referred to as Pdx1-C) was subcloned by polymerase chain reaction (PCR) into pET49b (Novagen) encoding a glutathione-*S*-transferase tag, a 6× His tag, and a 3C protease recognition site upstream of the cloning site. The recombinant plasmid was transformed into BL21(DE3) competent cells for protein overexpression. Cell growth conditions and the protein purification protocol for Pdx1-C were identical to our previously reported procedures for the Pdx1-homeodomain [16] except as noted below. As a final step to ensure full purification, Pdx1-C was subjected to size exclusion chromatography using a HiPrep 26/60 Sephacryl S-200 HR column (GE Life Sciences) in 50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM β-mercaptoethanol, and 1 mM ethylenediaminetetraacetic acid (EDTA). Following concentration using an Amicon Ultra centrifugal filter device (Millipore) that contained a PES 3000-MWCO (molecular weight cutoff) membrane, Pdx1-C was buffer exchanged into 50 mM cacodylate (pH 6.5), 50 mM KCl, and 1 mM dithiothreitol (DTT). Protein concentration was determined by direct detect Fourier transform infrared (FT-IR) (Millipore) using a molecular weight of 8089 g/mol.

NMR spectroscopy

All of the NMR experiments were recorded at 11.6 and 14.0 T on Bruker AVANCE-3 spectrometers operating at ^{1}H frequencies of 500.13 and 600.07 MHz, respectively, and equipped with TCI cryoprobes. All spectra were acquired on uniformly ^{15}N and ^{13}C isotope-enriched samples of Pdx1-C at concentrations ranging from 0.6 to 0.8 mM in 50 mM cacodylate buffer (pH 6.5), 50 mM KCl, 1 mM DTT, and 10% D_2O. All spectra were collected at 298 K.

Typical pulse times were 9.79 and 31 μs for hard ^{1}H and ^{15}N 90° pulses, respectively, with some sample-based variation in the ^{1}H pulse time. All pulsed field gradients (PFGs) used in the

experiments were applied for 1 ms with a sine shape. In all pulse sequences, unless otherwise noted, the 90° band-selective ^{13}C pulses have the Q5 shape (or time reversed, Q5tr) and the band-selective ^{13}C 180° pulses use the Q3 shape [17] with durations of 384 and 307 μs, respectively, at 11.6 T and durations of 320 and 256 μs, respectively, at 14.0 T. The timing and phase parameters specific to pulse sequences associated with the amino acid-filtered NMR experiments are reported in the relevant figure captions in the supplementary material. The ^{1}H and ^{15}N carriers were placed at 4.7 and 124 ppm, respectively. The pulses for excitation of carbonyl carbon and alpha carbon were centered at 172 and 54 ppm, respectively, whereas the ^{13}C carrier was changed at the positions indicated by vertical arrows to $^{13}C^{\alpha/ali(\beta)}$ = 39 ppm, $^{13}C^{\alpha}$ = 54 ppm, and $^{13}C'$ = 172 ppm. In all figures, the carbon pulses represented by solid shapes were applied on-resonance and those marked with slanted stripes were off-resonance pulses centered on the aliphatic region. The adiabatic inversion pulse (gray pulse) that inverts both C′ and C^{α} magnetizations, used during the nitrogen chemical shift labeling period, was a 500-ms CHIRP pulse with 60 Hz sweep and 25% smoothing [18]. 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. In all experiments, the recycle delay was set to 1.3 s except for the H_N-flip versions, where a 1.0-s delay was used.

Direct detection on $^{13}C'$ requires that steps be taken to refocus the active $^{13}C'$–$^{13}C^{\alpha}$ coupling during acquisition. In all pulse sequences used here, the carbonyl dimension is collected in the in-phase anti-phase (IPAP) manner, enabling $^{13}C'$–$^{13}C^{\alpha}$ couplings to be "virtually decoupled" as a post-acquisition processing step [9]. The 3D (H_N-flip)N(CA)CON experiment was performed with spectral widths of 20 × 40 × 40 ppm, a matrix size of 1024(C′) × 64(N) × 128(N) points, and 16 scans, amounting to a total acquisition time of 50 h (2 days 2 h). Owing to the lower intensities from 3D (H_N-flip)N(CA)NCO, the number of scans was increased to 32, yielding a total acquisition time of 100 h (4 days 4 h). The standard CCCON spectrum from the TopSpin pulse program library was collected over 2 days, with a matrix size of 1024(C′) × 64(N) × 128(C) points, spectral widths of 20 × 40 × 70 ppm, and 16 scans of signal averaging. The aliphatic protons in the side chains were assigned using H(CC)CON [5] collected over 2 days, with a matrix size of 1024(C′) × 64(N) × 128(H) points, spectral widths of 20 × 40 × 40 ppm, and 16 scans of signal averaging. All amino acid-filtered CON-based experiments were carried out with spectral widths of 20 × 40 ppm and matrix sizes of 1024(C′) × 256(N). The details of the pulses and delays used in these experiments are described in the figure captions below each amino acid-filtered experiment in the Supplementary material (Figs. S4–S18). All NMR data were processed in TopSpin 2.1 and converted to Sparky format [19] for data analysis.

Pdx1-C secondary structure analysis

Secondary structure propensity (SSP) was calculated from secondary chemical shifts to determine the per-residue probability of finding α-helix and β-sheet conformations in Pdx1-C [20]. The data are represented here by calculating the difference in $^{13}C^{\alpha}$ ($\Delta\delta C^{\alpha}$) and $^{13}C^{\beta}$ ($\Delta\delta C^{\beta}$) chemical shifts measured from those of random coils and plotting $\Delta\delta C^{\alpha}$ to $\Delta\delta C^{\beta}$. In addition, calculations were performed using the recently published δ2D methodology for predicting the secondary structures of disordered proteins using NMR chemical shifts [21]. This analysis was performed using the δ2D webserver (http://www-vendruscolo.ch.cam.ac.uk/d2D) that takes assigned $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N chemical shifts of each residue as inputs, predicting as output the relative populations of α-helix, extended β-sheets, polyproline II, and random coil present in the ensemble.

Results and discussion

Model system

Despite intense interest in evaluating structure–function relationships involving IDPs, there remains a significant composition difference [22] between databases of known protein sequences (e.g., the UniProt knowledge base [23]) and the sequences found in known 3D structures deposited in the Protein Data Bank [24], thereby reinforcing the assertion that disordered proteins are systematically under-represented in structural studies. To understand why IDPs pose such a challenge to structural biology, and to see the utility of the chemical shift assignment protocol we report here, it is helpful to understand how IDPs differ from cooperatively folding proteins on the molecular level. Bioinformatic studies have established a series of amino acid sequence hallmarks common to IDPs that render them distinct from known globular proteins, including the possession of long "medium-complexity" sequence segments and a tendency toward repeats [22]. Typically, IDPs are also depleted in nonpolar residues and enriched in polar and charged residues, whereas many display a pronounced enrichment in proline [25].

Each of these sequence hallmarks of disorder is found in the Pdx1 C-terminal domain (Fig. 1, top), which makes this protein an excellent model system for NMR method development. Pdx1 is a transcription factor known to play an important role in pancreatic development and β-cell maturation as well as the regulation of insulin production in mature organisms [26–31]. The well-folded homeodomain component of Pdx1 is well studied [16,29–31] but represents only a small fraction of the total polypeptide. Full-length Pdx1 includes both N- and C-terminal intrinsically disordered regions that are responsible for mediating known protein–protein interactions [32,33] and that include sites of polymorphism that correlate with type 2 diabetes in humans [34–37]. Consistent with the patterns reviewed above, the disordered segments of Pdx1 possess a low abundance of nonpolar residues, have low-complexity amino acid sequences, and feature repetitive glycine and proline-rich segments. Our laboratory has previously succeeded in using a combination of proton- and carbon-detected NMR spectra to fully assign and structurally characterize the intrinsically disordered C terminus of the human protein FCP1 [6–8,38]. In contrast, the prevalence of proline residues and otherwise extreme amino acid bias in Pdx1-C proved to be too much for effective analysis using our original methods [6], leading us to establish the protocol presented here for generating chemical shift assignments from purely carbon-detected NMR experiments.

Carbon-detected NMR approach for chemical shift assignment

The acquisition of essentially complete NMR chemical shift assignments is a prerequisite for any NMR-based approach characterizing the structures of biomolecules and the mechanisms driving their interactions with their cognate partners. For cooperatively folded proteins, the task of chemical shift assignment is usually accomplished by correlating the backbone carbon resonances (e.g., $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$) associated with a residue of interest to those of the nearest neighbor residue on the N-terminal side in the context of a suite of triple-resonance NMR experiments detected through the ^{1}H,^{15}N HSQC platform [14,15]. When coupled with prior knowledge of the amino acid sequence, this procedure is generally sufficient on three grounds: (i) the nearest neighbor connections bridge the peptide bond; (ii) there is sufficient redundancy built into such a large chemical shift set; and (iii) identifying the amino acid type of most residues based on statistical similarity to tabulated chemical shift values is straightforward and amenable

to automation. In most cases, this strategy fails for IDPs because the heterogeneous nature of their structure causes a collapse in the $^{1}H_N$ chemical shift dimension of the spectra and, in many cases, the prevalence of proline residues blocks the walk-along-the-backbone procedure that depends on all residues possessing an amide proton.

As discussed above, we work around the limitations inherent to proton-detected NMR spectroscopy of IDPs by recording the correlation between the carbonyl carbon and amide nitrogen in the peptide plane of the protein backbone, which gives rise to the ^{15}N,^{13}C CON spectrum. For Pdx1-C, the degree of improvement over conventional ^{1}H,^{15}N HSQC detection (Fig. 1A) is clearly evident in Fig. 1B, which displays the ^{15}N,^{13}C CON for this 80-residue IDP. It is worth noting that the ^{15}N,^{13}C CON also yields sharper line widths (at a given magnetic field strength) than the ^{1}H,^{15}N HSQC, further enhancing the practical resolution of the experiment. Finally, the ^{15}N,^{13}C CON has the convenient property of containing a resonance for each peptide bond involving a proline ^{15}N (easily identified by the downfield shift in the ^{15}N dimension).

One limitation of carbon-detected NMR spectroscopy is its inherently lower sensitivity, on a per-scan basis, compared with proton-detected NMR, which originates in the reduced gyromagnetic ratio of the ^{13}C nucleus. Although much of the early carbon-detected NMR literature focused on so-called "protonless" experiments, a wide range of "H-start" experiments have been developed more recently, yielding enhanced sensitivity. For proteins like Pdx1-C, which can be studied under mildly acidic conditions, we have found H_N-flip variants of the ^{15}N,^{13}C CON experiment to be especially effective because the increased per-scan sensitivity is coupled to a decrease in the necessary interscan delay created by a more rapid recovery of the amide protons to equilibrium [39,40]. A representative H_N-flip ^{15}N,^{13}C CON pulse sequence is shown in Fig. 2A and forms the basis for constructing the 3D pulse programs we discuss below. It is important to note that exclusive reliance on H_N-start experiments is not feasible for most IDPs due to their tendency to be enriched in proline, which is lost from the CON when the H_N-start format is used. In addition, fast solvent exchange caused by enhanced solvent exposure in the extended state renders IDPs highly susceptible to resonance intensity loss when exchangeable protons are included in the magnetization transfer pathway. For this reason, application of H_N-start CON techniques works best for IDPs that are soluble under mildly basic solution conditions. For proteins where these factors become limiting, and for the carbon-detected amino-acid-selective (CAS) NMR experiments we also use here, initial excitation on aliphatic proton resonances is the preferred means to enhancing sensitivity. Nearly all of the CAS–NMR experiments we discuss are formally similar to the H_α-start ^{15}N,^{13}C (HACA)CON, shown in Fig. 2B, which can also be used as a starting point for modification of other carbon start or H_N-start CON-based experiments.

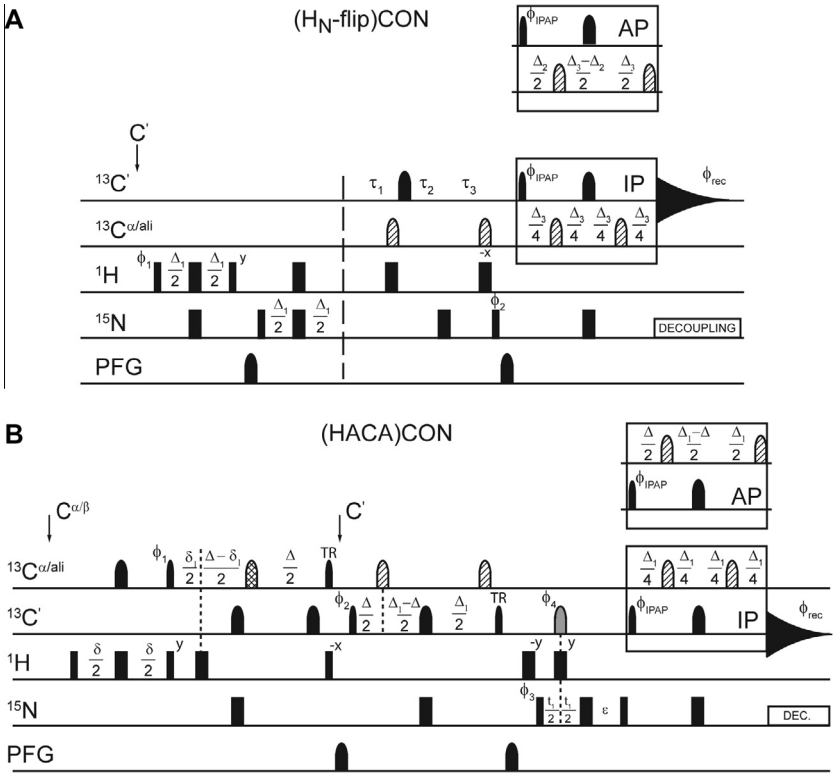

Fig.2. H-start versions of the ^{15}N,^{13}C CON form the basis of nearly all experiments used in this study. In both panels, narrow and wide rectangular pulses correspond to 90° and 180° hard pulses, respectively, whereas band-selective pulses on the ^{13}C channel are similarly represented with narrow (90°) and wide (180°) shapes. Pulsed field gradients (PFGs) are also indicated by shapes. (A) Pulse sequence for the (H_N-flip) CON–IPAP experiment. The delays are $\Delta = 5$ ms, $\Delta_1 = 4.6$ ms, $\Delta_2 = 9.0$ ms, and $\Delta_3 = 25$ ms. ^{15}N chemical shift evolution is measured during t_1 using a semi-constant time period with delays $\tau_1 = (\Delta_3 + t_1)/2$, $\tau_2 = (1 - \Delta_3/t_{1max})t_1/2$, and $\tau_3 = (1 - t_1/t_{1max})\Delta_3/2$, where t_{1max} is the maximum duration of the t_1 labeling period [47,48]. The phase cycle is $\varphi_1 = x, -x$; $\varphi_2 = y, y, -y, -y$; $\varphi_{IPAP}(IP) = x, x, x, x, -x, -x, -x, -x$; $\varphi_{IPAP}(AP) = -y, -y, -y, -y, y, y, y, y$; and $\varphi_{rec} = x, -x, -x, x, x, -x, -x, x$. Quadrature detection in the indirect dimension was obtained by States–TPPI (time proportional phase incrementation) of φ_2. (B) 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 $\varepsilon = t_1(0) + pC180$. The phase cycle is $\varphi_1 = 4(x), 4(-x)$; $\varphi_2 = 2(x), 2(-x)$; $\varphi_3 = x, -x$; $\varphi_4 = 8(x), 8(-x)$; $\varphi_{IPAP}(IP) = x$, $\varphi_{IPAP}(AP) = -y$; $\varphi_{REC} = x, -x, -x, x, -x, x, x, -x$. Quadrature detection in the indirect dimension is obtained by States–TPPI of φ_3. In both experiments, pulses are applied at the frequency indicated on the left of each line, with narrow and wide rectangles or shapes representing 90° and 180° pulses, respectively. All pulses are applied with x phase unless otherwise indicated. The gray pulse on ^{13}C indicates a band-selective $^{13}C'$ and $^{13}C^{\alpha}$ inversion pulse. Pulses filled with diagonal lines are off-resonance 180° Q3-shaped pulses centered at 54 ppm. The pulse filled with hashed lines is a higher selectivity 180° Q3-shaped pulse applied on-resonance with a duration of 1200 μs on a system operating at 11.7 T static field strength. Pulses labeled "TR" are time-reversed versions of the standard 90° Q5-shaped pulse.

Summarizing the procedure we outline here, our strategy is to combine 3D NMR experiments that help to establish nearest neighbor connections to adjacent residues with 3D NMR experiments fully characterizing the side chain chemical shifts. Acquisition of 2D amino acid-filtered experiments, which identify the amino acid types of as many resonances as possible through their spin topologies, further reduces the spectral degeneracy commonly associated with low-complexity protein sequences and yields numerous unambiguous entry points from which the process of assigning the backbone can be initiated.

3D NMR spectroscopy: connecting the backbone

The H_N-flip principle has been implemented in a pair of 3D carbon-detected NMR experiments described as (H_N-flip)N(CA)CON and (H_N-flip)N(CA)NCO [11] that collectively are sufficient to provide nearest neighbor linkages throughout the main chain. In our experience, optimal signal intensity in these two experiments was realized using the updated pulse sequences shown in Figs. S2 and S3 of the Supplementary material. The 3D (H_N-flip)N(CA)CON experiment is used to build nitrogen–nitrogen correlations between each signal observed in the CON and the amide nitrogen of the preceding residue in the chain. As displayed by the strips in Fig. 3A, the F1 dimension of the spectrum provides two correlations: the autocorrelation peak records the "self" amide nitrogen chemical shift, whereas the cross-peak records the nitrogen chemical shift of the preceding residue in the chain. For many systems, this one experiment should be sufficient to assign the observed magnetic resonances by daisy chain connecting each residue with its nearest neighbor to the N-terminal side. Unfortunately, owing to the H_N-start format, this experiment suffers from the same complication of proline breaking the chain as conventional proton-detected triple resonance experiments. Note, however, that the strips corresponding to non-proline residues that precede a proline do display their amide nitrogen correlations to the CON correlation featuring the proline, thereby providing some guidance for their placement. If the obstacle presented by proline-induced breaks in the walk along the backbone proves to be too great, the 3D (H_N-flip)N(CA)CON is amenable to conversion into a CANCO format.

For systems that are challenging to assign based on the 3D (H_N-flip)N(CA)CON alone, the 3D (H_N-flip)N(CA)NCO offers powerful complementary information by correlating each of the resonances in a CON with the amide chemical shift of the preceding, self, and following residues in the F1 dimension (Fig. 3B). Because the self and preceding amide nitrogen resonances are typically both known from (H_N-flip)N(CA)CON, it is usually quite easy to deduce which cross-peak corresponds to the amide resonance of the residue on the C-terminal side of the residue providing the CON peak. Note that the intensity of signals obtained from the (H_N-flip)N(CA)NCO is weaker as compared with the (H_N-flip)N(CA)CON; however, because this experiment allows one the liberty of moving either forward or back in the polypeptide daisy chain, the (H_N-flip)N(CA)NCO offers an especially powerful tool for difficult-to-assign proteins. When spectrometer time and sample stability are not limiting factors, the lower signal-to-noise ratio can generally be overcome with an increase in the number of scans acquired. As a starting point, we generally collect the (H_N-flip)N(CA)NCO with double the number of scans used for the (H_N-flip)N(CA)CON, which does not provide equal intensity but is generally sufficient to provide a usable signal-to-noise ratio.

The necessary task of correlating each peak in the $^{15}N,^{13}C$ CON with its position in the amino acid sequence of the protein chain is not sufficient to complete the assignment process for most purposes. In addition, it is desirable to record the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for the purpose of evaluating whether the IDP being studied possesses any secondary structure biases. Furthermore, full side chain chemical shift assignment can facilitate amino acid type

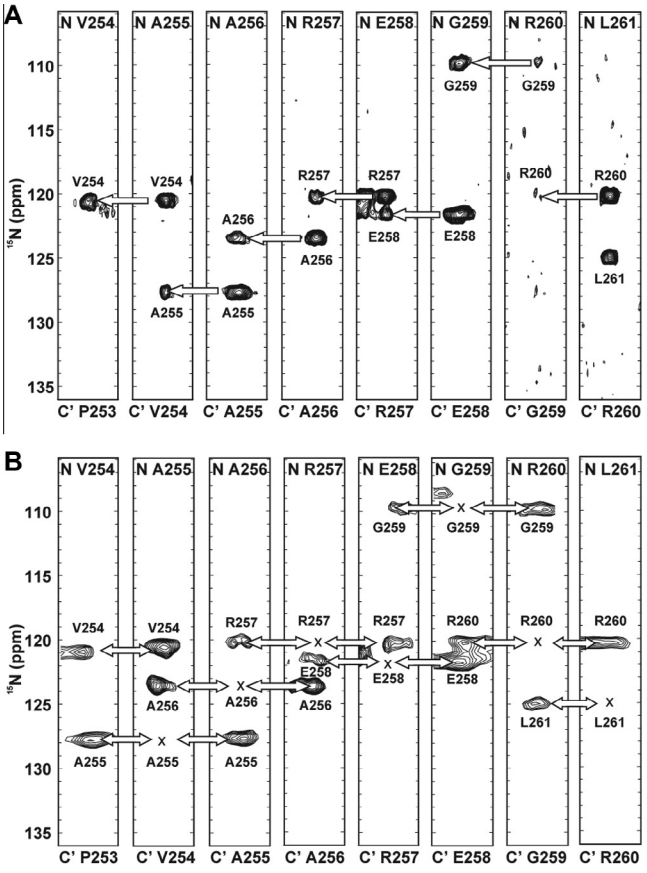

Fig.3. Strips corresponding to the labeled Pdx1-C resonances obtained from 3D (H_N-flip)N(CA)CON (A) and 3D (H_N-flip)N(CA)NCO (B) are shown. The bottom labels of the strips correspond to the carbonyl chemical shifts, whereas the top labels correspond to those of amide nitrogen in the F2 dimension. Nitrogen shifts in the F1 dimension are provided adjacent to each peak. The autocorrelation peak is rarely present in the 3D (H_N-flip)N(CA)NCO, so its position is marked with an *x* in the figure. The pulse sequences used to generate this data are presented in Figs. S2 and S3 of the Supplementary material.

assignment, among other objectives. The most efficient means we have found for completing the goal of comprehensive side chain chemical shift assignment is to collect an additional pair of 3D experiments: the 3D CCCON, used to obtain the aliphatic carbon side chain assignments [9], and the 3D H(CC)CON experiment, used to obtain the aliphatic proton side chain assignments [5]. In general, IDPs are sufficiently depleted in aromatic amino acids to render additional effort aimed at assigning the aromatic chemical shifts unnecessary, although the addition of extra experiments aiming to fill this gap remains an option.

2D NMR spectroscopy: CAS experiments

In many cases, collection of a high-resolution $^{15}N,^{13}C$ CON and the four 3D experiments discussed above is sufficient to assign the backbone of a target IDP. For each CON signal, 3D CCCON yields the complete aliphatic carbon resonances of the preceding residue, narrowing down the amino acid type, and the 3D (H_N-flip)N(CA)CON provides the nitrogen chemical shift of the preceding residue as well. Sometimes, however, this process leads to regions in the CON spectrum where, for the same nitrogen chemical shift, there is more than one resonance present, causing ambiguity. In these instances, the use of amino acid-filtered 2D experiments can provide the additional information needed to break the impasse at very little additional cost in terms of spectrometer time [12,41–44]. These experiments capitalize on amino acid type-specific spin physics, band-selective pulses exciting very unique

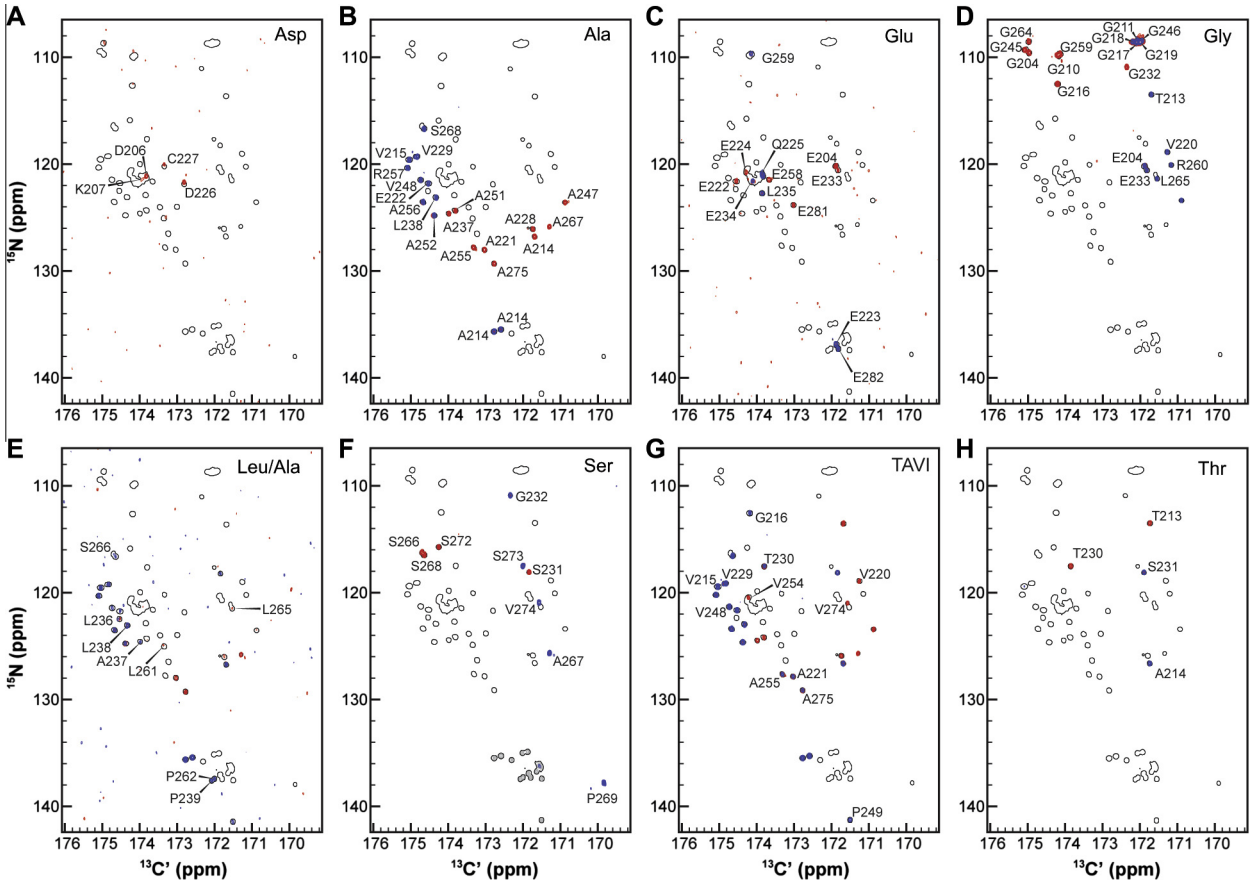

Fig.4. A suite of CAS–NMR experiments collected as CACON/CANCO for Asp (A), Ala (B), Glu (C), Gly (D), Leu/Ala (E), Ser (F), Thr-Ala-Val-Ile (TAVI) (G), and Thr (H) is shown. In each panel, the ^{15}N,^{13}C CON is shown as outlined peaks, with the CACON resonances overlaid in blue and the CANCO resonances in red. Note that the Ser-filtered CAS–CANCO contains artifact peaks arising from poor suppression of proline resonances (colored gray in panel F). The pulse sequences used to generate these data are presented in Figs. S4 to S18 of the Supplementary material. (For interpretation of the references to color in this figure legend and the text, the reader is referred to the Web version of this article.)

spectral regions, and specific coupling transfer steps to yield even greater amino acid filtering than is achievable by inference from the set of aliphatic carbon chemical shifts recorded in the CCCON alone. Originally designed by the Oschkinat group in an amide proton-detected format [41–44], these experiments were recently modified for ^{15}N,^{13}C CON readout as carbon-detected amino acid-specific spectra (CAS–NMR) [12]. As was the case with the 3D pulse sequences discussed above, we have found that in many cases minor modifications to the originally reported pulse sequences were required for optimal performance. Pulse programs for the entire suite of carbon-detected amino acid-specific 2D experiments that we employ in our laboratory are provided in the Supplementary material (Figs. S4–S18).

In general, the CAS–NMR data sets are composed of two experiments collected as a ^{15}N,^{13}C CACON and the closely related ^{15}N,^{13}C CANCO. The CACON-type experiment makes use of amino acid filters to select the desired amino acid type, followed by intra-residue transfer from the $^{13}C^{\alpha}$ to the $^{13}C'$, via the $^{1}J_{C\alpha}C'$ scalar coupling. The experiment concludes with standard readout of the $^{13}C'$–^{15}N correlation in the 2D CON format. The CANCO-type experiment also makes use of the same amino acid filters but employs a different set of echo times such that both the intra-residue $^{1}J_{C\alpha}C'$ and the inter-residue $^{2}J_{C\alpha}C'$ are active during transfer to the $^{13}C'$ prior to detection of $^{13}C'$–^{15}N pairs in a format visually indistinguishable from the 2D CON. Note, however, that the longer delay times required for refocusing both one- and two-bond coupling at the end of the CANCO experiment, along with splitting the initial magnetization between two correlations, lead to lower signal to noise, which is usually circumvented by an increase in the number of scans. The key to the CAS–NMR strategy is as follows. For a given selected amino acid type, the CAS-filtered ^{15}N,^{13}C CACON contains only peaks corresponding to the CON correlations of residues adjacent to the selected amino acid type on the C-terminal side. The CAS-filtered ^{15}N,^{13}C CANCO recaptures this same set of peaks and also the set of CON correlations directly corresponding to the amino acid type selected for. Although it is generally not possible to know a priori which self peak corresponds to which neighbor peak in the ^{15}N,^{13}C CANCO, the spectrum is greatly simplified and any ambiguity present in the 3D spectra discussed above is generally resolved.

To provide an example of how this procedure works, we present Pdx1-C CAS–NMR experiments acquired with eight different amino acid filters in Fig. 4 (the pulse sequences used are presented in Figs. S4–S18 of the Supplementary material). In each panel, the shadow of the full ^{15}N,^{13}C CON is displayed, with the CAS–CACON overlaid in blue and the CAS–CANCO in red. All CAS–NMR peaks have been annotated with an assignment corresponding to the residue contributing the ^{15}N resonance to the 2D correlation. Note that in all panels, half of the CAS–CANCO resonances are obscured by the CAS–CACON resonances. These spectra are satisfyingly free of bleed-through artifacts in all cases but one; all of the peaks arising from a proline amide nitrogen are present as artifacts in the Ser-filtered CAS–CANCO (and are colored gray in Fig. 4 for clarity). Note that this artifact poses little practical impediment to assignment.

The simplest illustration of how these experiments are used is provided by considering the Thr-filtered spectra in Fig. 4H (bottom

right). Pdx1-C contains two threonine residues (T213 and T230), resulting in the two annotated correlations from the Thr-filtered CAS–CANCO (red peaks in Fig. 4H) and the two inter-residue correlations annotated as A214 and S231, respectively (blue peaks in Fig. 4H). The identity of the blue peak appearing downfield in the ^{15}N dimension, annotated as A214 in Fig. 4H, was established by noting that it is also recorded in the Ala-filtered CAS–CANCO (see the red peak in the same position of the Ala panel in Fig. 4B). The dipeptide Thr-Ala appears only once in the sequence of Pdx1-C, providing unambiguous assignment of this resonance. Similarly, the blue resonance in the Thr-filtered CAS–CACON labeled as S231 is also present in the Ser-filtered CAS–CANCO, where it appears in red at the same location in Fig. 4F. Because the Thr-Ser dipeptide sequence is also unique, this resonance is unambiguously assigned. Next, the two red autocorrelations identified as threonine residues must be unambiguously assigned. This is similarly straightforward because T213 also is found in the Gly-filtered CAS–CACON (blue resonance in Fig. 4D), yielding positive identification; T230 is identified because all other possibilities are eliminated (although its assignment can also be verified from the Thr-Ala-Val-Ile (TAVI)-filtered CAS–CACON in Fig. 4G on additional inspection).

It is similarly straightforward to unambiguously identify the residue-specific assignments for all five resonances in the Ser-filtered CAS–CON because each serine in Pdx1-C is followed by an amino acid of a different type (G231, A267, P269, S273, or V274), all of which are used here to filter a CAS–CANCO spectrum. Following this procedure to its completion yields unambiguous assignment for a large fraction of peaks without any reference to 3D spectra and at least provides amino acid typing for many more. In other words, this procedure yields multiple robust starting points for analysis of the (H_N-flip)N(CA)CON experiment, generally resulting in in efficient resonance assignments of IDPs.

To summarize, a total of four 3D carbon-detected spectra and 15 2D carbon-detected spectra were used to generate thorough chemical shift assignment for the C-terminal disordered region of Pdx1 (the assignments are deposited in the BMRB under accession no. 19596) as a result of investing 2 weeks of NMR time. Although all of the pulse programs we used for this project have been published previously [5,9,11,12], we needed to modify a large number of them for use on our spectrometers; few others were included in the standard, manufacturer-provided pulse program library. To facilitate adoption of the protocol recommended in this article, we have made Bruker formatted versions of the pulse sequences used for this project (and others written by our laboratory) available through our website.

Outcomes for Pdx1-C

The experiments described here led efficiently to complete assignment of Pdx1-C despite its biased amino acid composition and the presence of long glycine and proline repeats. This project was motivated by a desire to study structure–function relationships for Pdx1-C, which is known to mediate interactions with other transcription factors and to harbor amino acid polymorphisms that correlate with the early onset of type 2 diabetes in humans [33,36]. In contrast to natively folding proteins, IDPs like Pdx1-C do not adopt a single conformation under native conditions but instead remain in a potentially heterogeneous ensemble of disordered conformations [45]. As a first step toward evaluating the structure of Pdx1-C, we calculated its SSP [20], based on the chemical shift assignments we report here, and compared the results with those we previously reported for the cooperatively folding homeodomain of Pdx1 [16]. Fig. 5 shows the difference between the offset from random coil for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei ($\Delta\delta C\alpha - \Delta\delta C\beta$) for each residue in the Pdx1 homeodomain (panel A) and Pdx1-C (panel B). When chemical shift data are represented in this

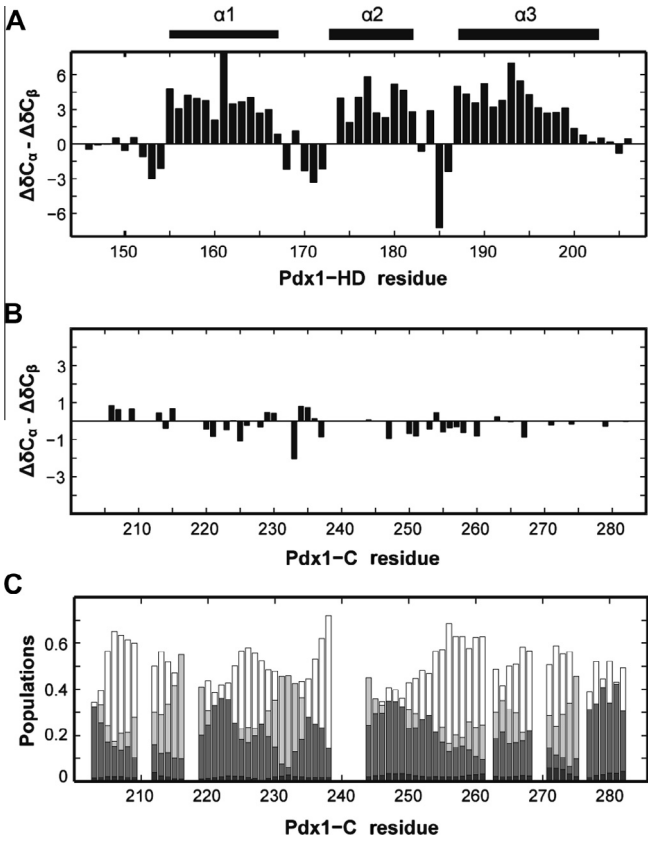

Fig.5. Secondary chemical shifts of Pdx1-C indicate that this protein adopts an extended coil-like structure. (A) For comparison, the secondary structure of Pdx1-HD (BMRB accession no. 19227) is shown as the difference between $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary chemical shifts, indicating the presence of an α-helix when long stretches of positive values are encountered. The secondary structure from the cocrystal of Pdx1 with a consensus DNA duplex is represented by bars above the figure for comparison. (B) Pdx1-C displays virtually no tendency toward secondary structure based on the difference between $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary chemical shifts. (C) δ2D calculations for Pdx1-C (see Materials and methods) provide additional insight into the nature of the extended structure found. For each residue, α-helix is shown in black, extended-β in dark gray, polyproline II in light gray, and coil structure as white (non-filled) bars.

way, stretches of values greater than or less than approximately 2 ppm correspond to α-helix or β-strand character, respectively, making it easy to see the three α-helical segments comprising the homeodomain (Fig. 5A). In contrast, the graph for Pdx1-C is almost completely devoid of features, suggesting that its conformation in solution is highly random. This assignment is confirmed through calculation of secondary structure tendencies in Pdx1-C using the recently published δ2D method [21], which predicts the ensemble to be a mixture of coil, polyproline II, and extended β conformation (Fig. 5C). The nearly complete absence of secondary structure bias predicted for Pdx1-C is somewhat unexpected because disordered fragments that mediate protein–protein interactions are often identified by secondary structural bias in the unbound state [46]. Therefore, it is intriguing to speculate that Pdx1-C may bind its partners in a highly extended conformation.

Conclusions

Using a suite of exclusively carbon-detected NMR experiments, we were able to successfully complete chemical shift assignment for the intrinsically disordered polypeptide chain of Pdx1-C. Our strategy combines nearest neighbor assignment, facilitated by 3D experiments that correlate amide nitrogen pairs and triples, with

CAS–NMR experiments composed of pairs of 2D CACON and CANCO spectra. The amalgamation of these two experiment types yielded a chemical shift assignment strategy that required only 2 weeks of instrument time and that was sufficient for the complete assignment of an 80-residue IDP, shown by secondary chemical shift information to be almost completely devoid of secondary structure. This chemical shift assignment of Pdx1-C now serves as the stepping stone for future structural and mechanistic studies. Of broad importance for the community, all of the pulse programs used to generate these results, as well as many others our laboratory has used and/or published, are now freely available through our website, making adoption of the protocol used here straightforward.

Acknowledgments

This work was supported by a National Science Foundation (NSF) CAREER award (MCB-0953918) and startup funds from The Pennsylvania State University to S.S. M.B. is the recipient of a National Institutes of Health (NIH) predoctoral fellowship (F31GM101936). We thank the Penn State ScholarSphere team (http://scholarsphere.psu.edu) for its help in establishing permanent and stable web hosting of our pulse programs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ab.2013.12.005.

References

[1] A. Tantos, K.H. Han, P. Tompa, Intrinsic disorder in cell signaling and gene transcription, Mol. Cell. Endocrinol. 348 (2012) 457–465.
[2] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[3] M. Wells, H. Tidow, T.J. Rutherford, P. Markwick, M.R. Jensen, E. Mylonas, D.I. Svergun, M. Blackledge, A.R. Fersht, Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain, Proc. Natl. Acad. Sci. USA 105 (2008) 5762–5767.
[4] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J.D. Forman-Kay, Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase, Structure 18 (2010) 494–506.
[5] B. O'Hare, A.J. Benesi, S.A. Showalter, Incorporating ^{1}H-chemical shift determination into ^{13}C-direct detected spectroscopy of intrinsically disordered proteins in solution, J. Magn. Reson. 200 (2009) 354–358.
[6] S.A. Showalter, NMR assignment of the intrinsically disordered C-terminal region of *Homo sapiens* FCP1 in the unbound state, Biomol. NMR Assign. 3 (2009) 179–181.
[7] C.W. Lawrence, A. Bonny, S.A. Showalter, The disordered C-terminus of the RNA polymerase II phosphatase FCP1 is partially helical in the unbound state, Biochem. Biophys. Res. Commun. 410 (2011) 461–465.
[8] C.W. Lawrence, S.A. Showalter, 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 (2012) 1409–1413.
[9] W. Bermel, I. Bertini, I.C. Felli, M. Piccioli, R. Pierattelli, ^{13}C-detected protonless NMR spectroscopy of proteins in solution, Prog. Nuclear Magn. Reson. Spectrosc. 48 (2006) 25–45.
[10] J. Novacek, N.Y. Haba, J.H. Chill, L. Zidek, V. Sklenar, 4D non-uniformly sampled HCBCACON and (1)J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins, J. Biomol. NMR 53 (2012) 139–148.
[11] W. Bermel, I. Bertini, I.C. Felli, L. Gonnelli, W. Kozminski, A. Piai, R. Pierattelli, J. Stanek, Speeding up sequence specific assignment of IDPs, J. Biomol. NMR 53 (2012) 293–301.
[12] W. Bermel, I. Bertini, J. Chill, I.C. Felli, N. Haba, M.V.V. Kumar, R. Pierattelli, Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs), ChemBioChem 13 (2012) 2425–2432.
[13] D. Pantoja-Uceda, J. Santoro, Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins, J. Biomol. NMR 57 (2013) 57–63.
[14] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, M. Rance, Protein NMR Spectroscopy: Principles and Practice, second ed., Academic Press, San Diego, 2005.
[15] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog. Nuclear Magn. Reson. Spectrosc. 34 (1999) 93–158.
[16] M. Bastidas, S.A. Showalter, Thermodynamic and structural determinants of differential Pdx1 binding to elements from the insulin and IAPP promoters, J. Mol. Biol. 425 (2013) 3360–3377.
[17] L. Emsley, G. Bodenhausen, Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators, J. Magn. Reson. 97 (1992) 135–148.
[18] J.M. Bohlen, G. Bodenhausen, Experimental aspects of CHIRP NMR–spectroscopy, J. Magn. Reson. A 102 (1993) 293–301.
[19] T. Goddard, D. Kneller, SPARKY 3 [computer software], University of California, San Francisco, 2006.
[20] J.A. Marsh, V.K. Singh, Z. Jia, J.D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: implications for fibrillation, Protein Sci. 15 (2006) 2795–2804.
[21] C. Camilloni, A. De Simone, W.F. Vranken, M. Vendruscolo, Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts, Biochemistry 51 (2012) 2224–2231.
[22] J.C. Wootton, Non-globular domains in protein sequences: automated segmentation using complexity measures, Comput. Chem. 18 (1994) 269–285.
[23] UniProt Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[25] P. Romero, Z. Obradovic, X.H. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Protein Struct. Funct. Dyn. 42 (2001) 38–48.
[26] A.S. Bernardo, C.W. Hay, K. Docherty, Pancreatic transcription factors and their role in the birth, life, and survival of the pancreatic beta cell, Mol. Cell. Endocrinol. 294 (2008) 1–9.
[27] E. Glick, D. Leshkowitz, M.D. Walker, Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter, J. Biol. Chem. 275 (2000) 2199–2204.
[28] M.A. Cissell, L. Zhao, L. Sussel, E. Henderson, R. Stein, Transcription factor occupancy of the insulin gene in vivo: evidence for direct regulation by Nkx2.2, J. Biol. Chem. 278 (2003) 751–756.
[29] K. Ohneda, R.G. Mirmira, J. Wang, J.D. Johnson, M.S. German, The homeodomain of PDX-1 mediates multiple protein–protein interactions in the formation of a transcriptional activation complex on the insulin promoter, Mol. Cell. Biol. 20 (2000) 900–911.
[30] P. Serup, J. Jensen, F.G. Andersen, M.C. Jorgensen, N. Blume, J.J. Holst, O.D. Madsen, Induction of insulin and islet amyloid polypeptide production in pancreatic islet glucagonoma cells by insulin promoter factor 1, Proc. Natl. Acad. Sci. USA 93 (1996) 9015–9020.
[31] A. Longo, G.P. Guanga, R.B. Rose, Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain, Biochemistry 46 (2007) 2948–2957.
[32] A. Liu, J. Oliver-Krasinski, D.A. Stoffers, Two conserved domains in PCIF1 mediate interaction with pancreatic transcription factor PDX-1, FEBS Lett. 580 (2006) 6701–6706.
[33] Y. Qiu, M. Guo, S. Huang, R. Stein, Insulin gene transcription is mediated by interactions between the p300 coactivator and PDX-1, BETA2, and E47, Mol. Cell. Biol. 22 (2002) 412–420.
[34] E.H. Hani, D.A. Stoffers, J.C. Chevre, E. Durand, V. Stanojevic, C. Dina, J.F. Habener, P. Froguel, Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus, J. Clin. Invest. 104 (1999) R41–R48.
[35] J. Weng, W.M. Macfarlane, M. Lehto, H.F. Gu, L.M. Shepherd, S.A. Ivarsson, L. Wibell, T. Smith, L.C. Groop, Functional consequences of mutations in the MODY4 gene (IPF1) and coexistence with MODY3 mutations, Diabetologia 44 (2001) 249–258.
[36] B.N. Cockburn, G. Bermano, L.L. Boodram, S. Teelucksingh, T. Tsuchiya, D. Mahabir, A.B. Allan, R. Stein, K. Docherty, G.I. Bell, Insulin promoter factor-1 mutations and diabetes in Trinidad: identification of a novel diabetes-associated mutation (E224K) in an Indo-Trinidadian family, J. Clin. Endocrinol. Metab. 89 (2004) 971–978.
[37] W.M. Macfarlane, T.M. Frayling, S. Ellard, J.C. Evans, L.I. Allen, M.P. Bulman, S. Ayres, M. Shepherd, P. Clark, A. Millward, A. Demaine, T. Wilkin, K. Docherty, A.T. Hattersley, Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes, J. Clin. Invest. 104 (1999) R33–R39.
[38] S. Kumar, S.A. Showalter, W.G. Noid, Native-based simulations of the binding interaction between RAP74 and the disordered FCP1 peptide, J. Phys. Chem. B 117 (2013) 3074–3085.
[39] W. Bermel, I. Bertini, I.C. Felli, R. Pierattelli, Speeding up ^{13}C direct detection biomolecular NMR spectroscopy, J. Am. Chem. Soc. 131 (2009) 15339–15345.
[40] I. Bertini, I.C. Felli, L. Gonnelli, M.V.V. Kumar, R. Pierattelli, High-resolution characterization of intrinsic disorder in proteins: expanding the suite of ^{13}C-detected NMR spectroscopy experiments to determine key observables, ChemBioChem 12 (2011) 2347–2352.
[41] P. Schmieder, M. Leidert, M. Kelly, H. Oschkinat, Multiplicity-selective coherence transfer steps for the design of amino acid-selective experiments: a triple-resonance experiment selective for Asn and Gln, J. Magn. Reson. 131 (1998) 199–202.

[42] M. Schubert, M. Smalla, P. Schmieder, H. Oschkinat, MUSIC in triple-resonance experiments: amino acid type-selective ^{1}H–^{15}N correlations, J. Magn. Reson. 141 (1999) 34–43.
[43] M. Schubert, L.J. Ball, H. Oschkinat, P. Schmieder, Bridging the gap: a set of selective ^{1}H–^{15}N correlations to link sequential neighbors of prolines, J. Biomol. NMR 17 (2000) 331–335.
[44] M. Schubert, H. Oschkinat, P. Schmieder, MUSIC and aromatic residues: amino acid type-selective ^{1}H–^{15}N correlations (part III), J. Magn. Reson. 153 (2001) 186–192.
[45] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci. 11 (2002) 739–756.
[46] A. Mohan, C.J. Oldfield, P. Radivojac, V. Vacic, M.S. Cortese, A.K. Dunker, V.N. Uversky, Analysis of molecular recognition features (MoRFs), J. Mol. Biol. 362 (2006) 1043–1059.
[47] S. Grzesiek, A. Bax, Amino-acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins, J. Biomol. NMR 3 (1993) 185–204.
[48] T.M. Logan, E.T. Olejniczak, R.X. Xu, S.W. Fesik, A general method for assigning NMR-spectra of denatured proteins using 3D HC(CO)NH–TOCSY triple resonance experiments, J. Biomol. NMR 3 (1993) 225–231.