Chapter 32

Analysis of structure and dynamics of intrinsically disordered regions in proteins using solution NMR methods

Nikita V. Saibo, Snigdha Maiti, Bidisha Acharya, and Soumya De
School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, India

1 Introduction

Under physiological conditions, an intrinsically disordered protein (IDP) does not fold into a stable, three-dimensional structure; instead, it exists as a flexible polypeptide chain.[1] The term IDP is broadly used to refer to proteins that are entirely disordered as well as proteins that carry long disordered regions, typically longer than 30 residues, appended to properly folded domains.[2,3] In this chapter, we will use the term intrinsically disordered regions and proteins (IDRPs) to refer to both these types of proteins. IDRPs are found in the proteome of all three kingdoms of life with 2%, 4%, and 33% disordered residues in archaea, eubacteria, and eukarya, respectively.[4] In eukaryotes, more than 70% of signaling proteins have long disordered regions.[5,6] The higher proportion of disordered regions in eukaryotes relative to archaea or eubacteria stems from the greater need for protein-mediated signaling and regulation in eukaryotes.[6–8] A majority of transcription factors in eukaryotes are predicted to carry long disordered regions.[9] Long disordered regions are rarely found in active enzymes, as they need specific three-dimensional structures to carry out their catalytic function. However, few active enzymes with disordered regions have recently been identified.[10–12] Despite such a significant abundance of IDRPs in various organisms and their involvement in important functions, the structural studies of IDRPs were largely hampered due to the lack of appropriate experimental methods.

Historically, structural biology methods have been developed to study folded proteins. Folded proteins have a funnel-shaped energy landscape where the well-folded native state is at the bottom of the funnel.[13,14] In contrast, disordered sequences have a flat energy landscape and lack an isolated global energy minimum.[15] Due to small energy differences between the various near-identical local minima, the disordered sequences adopt an ensemble of conformations.[16] This also confers multiple degrees of flexibility to the disordered regions in IDRPs. IDRPs are involved in a multitude of functions such as transcription regulation, cell cycle regulation, cell division, apoptosis, cell signaling, and signal transduction.[1,17–20] IDRPs often have multiple interaction motifs, which allow them to bind to various targets through these sites and thereby act as hub proteins with a large number of connections in protein interaction networks (PINs).[21–23] Mutations in IDRPs have also been implicated in several diseases.[20,24,25] Several oncogenes and other cancer-associated genes are found to encode IDRPs.[18,20] Due to their abundance and involvement in important biological functions, it has been clear that new structural biology methods need to be developed to study IDRPs. In this chapter, we will discuss various NMR methods, which have proved to be the most versatile for studying IDRPs.

2 NMR chemical shift assignments of intrinsically disordered sequences

The first step of any NMR study is to assign the chemical shifts of each NMR active nuclei in the molecule. Due to a large number of protons in IDRPs, one-dimensional spectra have severe signal overlap. This problem is mitigated by incorporating additional spin ½ isotopes ^{15}N and ^{13}C and collecting multi-dimensional heteronuclear spectra. Here we will discuss strategies to assign each ^{1}H, ^{15}N, and ^{13}C nuclei in the protein.

Advances in Protein Molecular and Structural Biology Methods. https://doi.org/10.1016/B978-0-323-90264-9.00032-5

2.1 Standard backbone assignment strategy

The most common NMR spectrum collected first for any protein is ^{1}H-^{15}N HSQC, which provides a fingerprint spectrum of the protein where all the N-H pairs present in the protein backbone and side chains appear as cross-peaks. For a folded protein, a well-resolved spectrum is obtained (Fig. 1A and B). The spectral resolution of disordered sequences is poor as they lack unique structure and sample very similar chemical environments. The lack of chemical shift dispersion is especially evident in the proton dimension. Cross-peaks occur in a narrow range from 7.5 to 8.5 ppm resulting in a highly crowded spectrum (Fig. 1C). Such narrow dispersion typically provides the first clue about the presence of a disordered sequence in a protein.

The cross-peaks in the ^{1}H-^{15}N HSQC spectra are assigned to their corresponding amino acid using multiple 3D experiments (Table 1) using a ^{13}C and ^{15}N dual-labeled protein sample.[26,27] For short disordered regions (up to 50 residues) appended to folded domains, this standard sequential backbone assignment method may be adequate. The presence of repeat

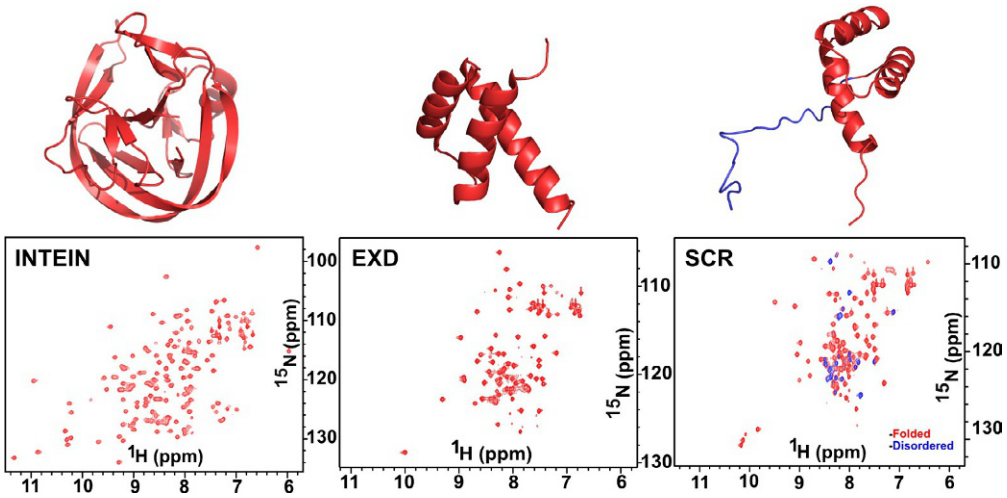

FIG. 1 ^{1}H-^{15}N HSQC spectra reflect protein structure. Intein *(left)* is an all beta-stranded protein and has large chemical shift dispersion. Extradenticle (EXD) is an all alpha-helical protein and has narrower ^{1}H dispersion *(center)*. SCR *(right)* has a folded domain *(red)* and intrinsically disordered residues *(blue)*. Peaks corresponding to the folded domain and the disordered region are shown in red and blue, respectively. The disordered residues appear in a narrow range from 7.5- to 8.5-ppm in the proton dimension giving rise to a crowded spectrum.

TABLE 1 NMR experiments for backbone assignments.

^{1}H detected experiments		
HSQC	**N_i-H^N_i**	
CBCA(CO)NH	$C\beta_{i-1}$-$C\alpha_{i-1}$-N_i-H^N_i	Grzesiek and Bax[32]
HNCACB	H^N_i-N_i-$C\alpha_i$-$C\beta_i$ and H^N_i-N_i-$C\alpha_{i-1}$-$C\beta_{i-1}$	Grzesiek and Bax[32]
HNCO	H^N_i-N_i-C'_{i-1}	Kay et al.[33]
HN(CA)CO	H^N_i-N_i-C'_i and H^N_i-N_i-C'_{i-1}	Clubb et al.[34]
HNN	N_{i-1}-N_i-N_{i+1}	Panchal et al.[28]
^{13}C detected experiments		
CON, (HACA)CON	N_i (F1), C'_{i-1} (F2)	Bermel et al.[35]
(HACA)N(CA)CON	N_i (F1), N_{i-1} (F2), C'_{i-1} (F3)	Bastidas et al.[36]
CCCON	C^{ali}_{i-1} (F1), N_i, C'_{i-1} (F3)	Bermel et al.[35]
H(CC)CON	H^{ali}_{i-1} (F1), N_i, C'_{i-1} (F3)	O'Hare et al.[37]

sequences such as poly glutamines, glycine-serine repeats, or polyprolines can complicate the assignment of even short disordered sequences. Since ^{15}N chemical shift dispersion is better than $^{1}H^{N}$ chemical shifts even for IDRPs, the HNN experiment is found to be very useful in sequential backbone assignment. It provides through bond connection to the previous and the following amide ^{15}N chemical shift of a residue[28] (Fig. 2A). The phase of the peaks, i.e., positive or negative, depends on the adjacent residue type and can aid direct identification of ^{15}N shifts for glycines and prolines.[29] In a dual receiver NMR spectrometer, the 3D HNN(N_{i-1}-N_i-N_{i+1}) and 2D $^{13}C^{\alpha}$-^{15}N ($C^{\alpha}_{i/i-1}$ -N_i) can be simultaneously acquired.[30] The 2D $^{13}C^{\alpha}$-^{15}N spectrum utilizes the wide dispersion range of ^{13}C and ^{15}N nuclei chemical shifts and also provides directionality for a sequential walk. The 2D version of the HNN experiment, named 2D-(HN)NH, also presents a way to identify amino acid type.[31] Its three variants are 2D-(HN)NH-G, 2D-(HN)NH-A, and 2D-(HN)NH-ST, which identify glycine, alanine, and serine/threonine, respectively, which aid the assignment by providing multiple starting points for a sequential walk.

2.2 Backbone assignment using ^{13}C detection

Although the ^{1}H-^{15}N HSQC experiment is the workhorse of NMR studies, it can be of limited use for IDRPs with long disordered regions. Apart from signal overlap in the ^{1}H dimension, ^{1}H-^{15}N HSQC spectra also have reduced signal intensity due to faster exchange of the amide protons in the solvent-exposed disordered sequences. Also, IDRPs are often proline-rich sequences.[38] Prolines lack amide protons and hence are not detected in ^{1}H-^{15}N HSQC spectra. Most of these issues can be circumvented using ^{13}C-^{15}N CON and ^{13}C-^{15}N CAN as alternatives.[39,40] The 2D ^{13}C-^{15}N CON and ^{13}C-^{15}N CAN experiments correlate the amide nitrogen with the carbonyl carbon and alpha carbon, respectively (Fig. 2B). These experiments have increased spectral resolution, are unaffected by the solvent exchange of amide protons, and can efficiently detect proline residues. In these experiments, ^{13}C is directly detected. Due to the lower gyromagnetic ratio, the sensitivity of ^{13}C is much less than ^{1}H, which in turn increases the experiment time. However, the use of 'proton start' experiments,[41] cryogenically cooled probes, and sparse data sampling can mitigate this problem to some extent.

The 2D ^{13}C-^{15}N CON experiment also provides a fingerprint spectrum and gives better resolution compared to ^{1}H-^{15}N HSQC (Fig. 3). For sequential backbone assignment, 3D (HACA)N(*CA*)CON is used, which connects the amide nitrogen and carbonyl carbon of each residue with the amide nitrogen of the following residue (N_{i-1}-C'_{i-1}-N_i).[36] For highly repetitive sequences, 3D (HACA)N(*CA*)NCO is particularly useful as it provides bidirectional correlation between each C'_{i-1}-N_i spin-pair and the amide nitrogen of i-1th and i + 1th residue (N_{i-1}-C'_{i-1}-N_i-N_{i+1}).

2.3 Fast data acquisition to reduce experiment time

IDRPs are often susceptible to proteolytic cleavage and have a smaller half-life compared to folded proteins. Many IDRPs are also prone to aggregation. Multi-dimensional NMR experiments can take up to two weeks of NMR time. Direct detection of low sensitivity ^{13}C nuclei can further increase the experimental time. This requires a significant reduction in data acquisition time. Several techniques have been developed to get good resolution in a relatively short period of time. Among them, Non-Uniform Sampling (NUS) and Band selective Excitation Short Transient (BEST) methods are widely used.[42]

In NUS, sparse data are collected in the indirect dimension, i.e., some of the points in the FID are skipped in the indirect dimension or randomly sampled to save time, and later on, during processing, the whole data set is reconstructed. In conventional NMR, signals are collected in discrete steps for each dimension. For 2D experiments, the indirectly detected

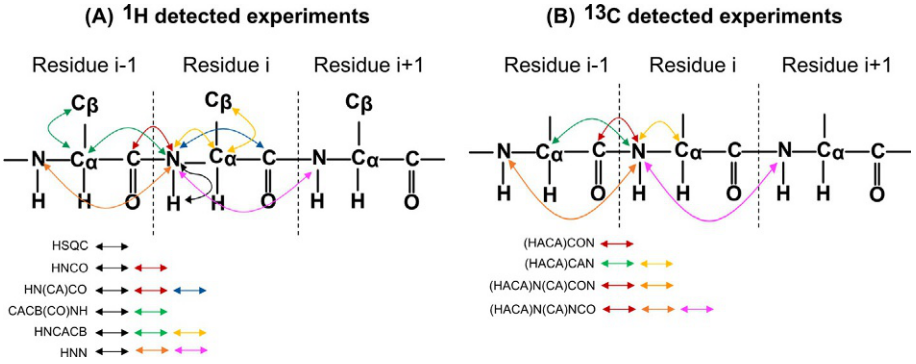

FIG. 2 Experiments for sequential assignment of IDRPs. The set of ^{1}H-detected (A) and ^{13}C-detected (B) experiments are shown. For each experiment, the set of *arrows* indicate the observed nuclei and the logical connections between them.

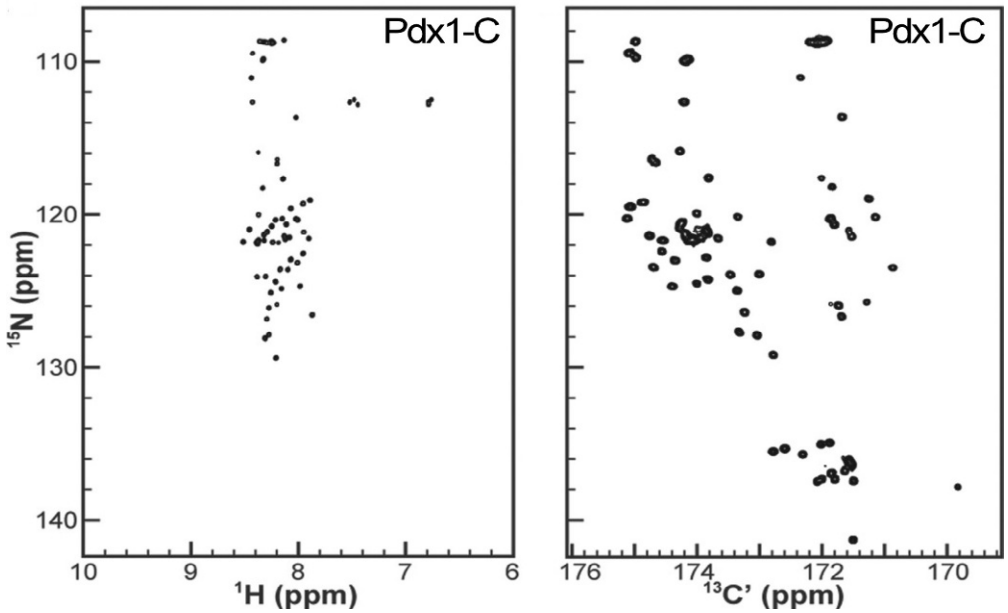

FIG. 3 Comparison of ^{1}H-^{15}N HSQC and ^{15}N-^{13}C CON spectra of Pdx1-C. ^{15}N-^{13}C CON *(right)* provides better chemical shift dispersion as compared to the ^{1}H-^{15}N HSQC spectra *(left)* of Pdx1-C. (Published with permission from Sahu et al. (2014) *Anal Biochem* 2014;449:17–25.)

dimension is sampled as a series of 1D experiments with delays between pulses (evolution time), which is incremented at a fixed amount. The experiment time increases with the number of increments that are being measured. In NUS, a subset of 1D data points between the first and last points are skipped in a semi-random manner. The missing data points are reconstructed using mathematical algorithms. Unlike traditional FID, which is processed by Fourier transformation (FT), non-uniform sampling requires special processing systems. Some of the frequently used such systems are multi-dimensional decomposition (MDD),[43] Iterative Soft Thresholding method (IST),[44] Sparse Multi-dimensional Iterative Lineshape-Enhanced (SMILE),[45] and deep neural networks.[46] These processing schemes are commonly offered in modern NMR spectrometers.

Band selective excitation short-transient (BEST) reduces the NMR experiment time by shortening the delay between scans. The delay between consecutive scans is provided to restore the equilibrium spin polarization via longitudinal relaxation. In BEST experiments, the amide protons (in 6 to 10 ppm range) are selectively excited without perturbing the aliphatic protons (in 0 to 5.5 ppm range). Dipolar interactions between a large number of aliphatic protons with the observed amide protons significantly reduce the longitudinal relaxation times, thereby quickly restoring the spin polarization of the amide protons.[47] This enables very short recycle delays and saves time. A combination of NUS-BEST pulses has been successfully used to assign the backbone of α-synuclein.[42]

2.4 Segmental isotope labeling of IDRPs

Segmental isotope labeling is a unique approach to selectively study a specific region within a protein by NMR spectroscopy. One-third of all eukaryotic proteins contain long intrinsically disordered regions (IDRs) along with one or more folded domains.[48] Characterization of such disordered regions should be done in the context of the full protein to get a complete structure–function relationship between the disordered region and the appended folded domain.[49] Eliminating the signals from the folded part of the protein can significantly reduce spectral crowding. This can be achieved by expressed protein ligation (EPL), protein trans-splicing (PTS), or sortase-mediated ligation (SML) methods (Fig. 4).

EPL and PTS methods[50] utilize intein enzymes, which catalyze the formation of a peptide bond between the polypeptide chains appended at their two terminals and are themselves excised out.[27] In the EPL method, the target disordered region is expressed with intein fused to its C-terminus followed by an affinity tag. This recombinant protein is expressed in bacteria and isotopically labeled with ^{13}C and ^{15}N. The folded domain is expressed as an unlabelled protein with a cysteine at its N-terminus. The IDR-intein is cleaved with a thiol to form an alpha-thioester, which reacts with the N-terminal cysteine of the folded domain, resulting in a thioester bond between the IDR and the folded domain. The thioester bond spontaneously rearranges to a peptide bond, producing the desired product (Fig. 4A). In the PTS method, naturally occurring or engineered split-inteins are used.[51] The disordered part of the target protein along with one half of the split intein is isotopically labeled

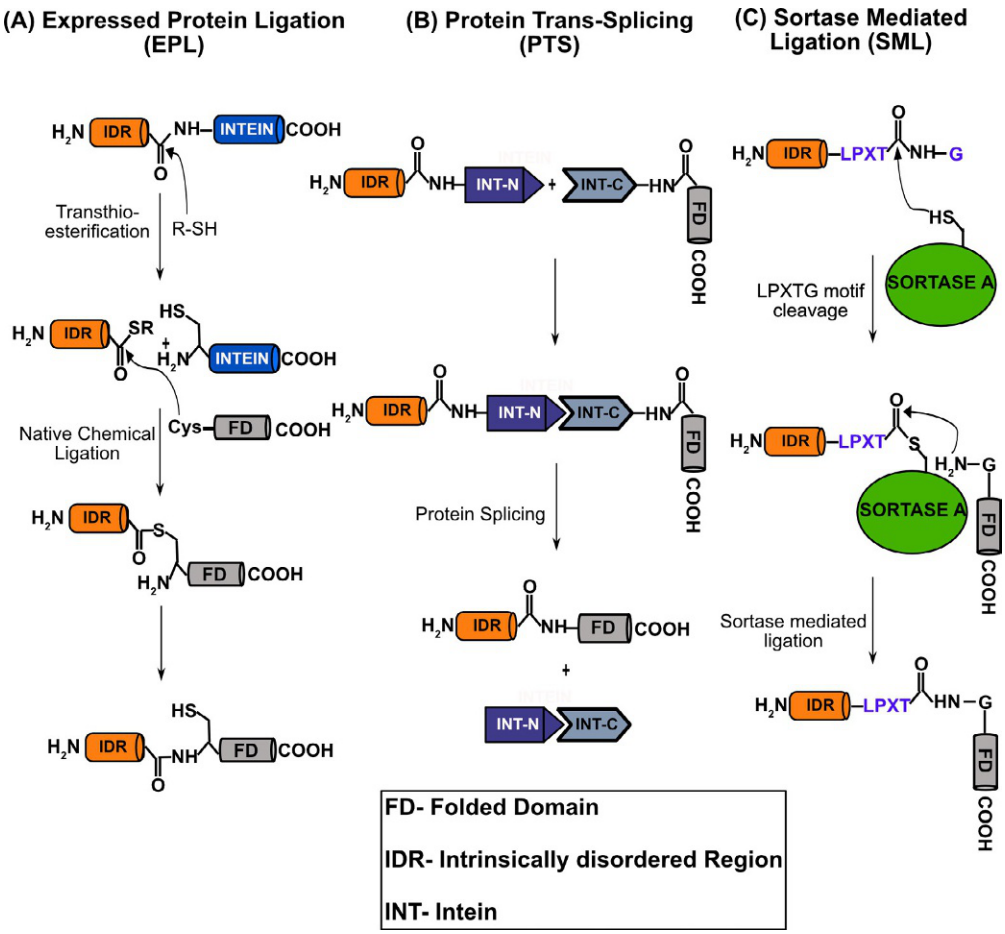

FIG. 4 Segmental isotope labeling of IDRPs. (A) In expressed protein ligation (EPL) approach, the intrinsically disordered region (IDR) and the folded domain (FD) are expressed separately and ligated via native chemical ligation. (B) Protein trans-splicing (PTS) of IDR and DF is accomplished using split intein. (C) For sortase-mediated ligation (SML), Sortase A recognizes the C-terminal LPXTG motif of IDR and cleaves the Gly residue. The acyl-enzyme intermediate is attacked by the Gly of FD, resulting in the formation of the ligated protein and the removal of Sortase A.

with ^{13}C and ^{15}N, and the folded part along with the remaining half of the split intein is expressed without isotope enrichment. These purified recombinant proteins are mixed to produce the functional intein enzyme, which catalyzes a splicing reaction forming a peptide bond between the disordered and folded parts of the target protein, and the intein itself is excised out (Fig. 4B).

In the SML method, the sortase enzyme is used, which catalyzes the peptide bond formation between an LPXTG motif in the C-terminus of proteins and the poly-glycine bridge in the cell wall of gram-positive bacteria.[52] The disordered region, containing the LPXTG motif in its C-terminus, is expressed as ^{15}N and ^{13}C labeled, and the folded domain, with triglycine in its N-terminus, is expressed as unlabelled protein. The purified proteins are linked by a peptide bond by the sortase enzyme (Fig. 4C). In all three methods, i.e., EPL, PTS, and SML, the disordered sequence can be appended to the C-terminus of a folded domain by swapping the constructs as described here (Fig. 4).

All three methods have advantages and disadvantages and should be chosen accordingly. Since in the EPL method, a thiol is added to form the alpha-thioester, this method should be avoided if the folded domain contains disulfide linkages. The PTS and SML methods can be used in such cases. However, the PTS method may have poor expression and solubility issues depending on the target sequences appended to the split inteins. In the SML method, a non-native sequence of LTXTGGG is inserted between the disordered and folded sequences.

2.5 Cell-free protein synthesis for IDRPs

IDRPs, which are toxic to the cell or are prone to proteolytic cleavage, can be produced by cell-free protein synthesis. Cell lysate containing all the cellular components for protein expression are typically derived from *E. coli* and used for cell-free

protein synthesis.[53] Due to the open nature of this system, protease inhibitors, stabilizers, scrambling inhibitors, or solubilizing agents can be added directly during the protein synthesis, thus enhancing IDRP production. Severe spectral crowding in IDRPs with long disordered regions can be reduced using selectively labeled amino acids instead of uniform ^{15}N and ^{13}C labeling. Unfortunately, labeled amino acids or precursors can be converted to other amino acids via various metabolic pathways in the cell.[54] In the cell-free system, amino acid scrambling is significantly reduced due to the absence of most metabolic enzymes. Scrambling can be further reduced by adding inhibitors.[55]

A selectively labeled sample cannot be used for sequential assignment. However, selective labeling can be used to fill in the gaps after sequential assignment using a uniformly labeled sample. Several combinatorial selective labeling strategies have been developed to complete the assignment using a minimal number of samples.[56]

3 Structural characterization of IDRPs

Intrinsically disordered sequences lack well-defined structural elements such as secondary and supersecondary structures as found in folded proteins. However, disordered regions do possess functional elements spread throughout the length of the sequence, which facilitate their interaction with multiple partner molecules (small molecules, DNA, RNA, or proteins). Depending on their structural features and length, these functional elements are classified into two distinct categories, namely linear motifs and molecular recognition features (MoRFs). Linear motifs are 3–15 residues long, which form interaction motifs.[57] They are also referred to as short linear motifs (SLiMs) or eukaryotic linear motifs (ELMs). MoRFs are partially structured segments of 10–70 residues that undergo disorder-to-order transitions upon partner recognition.[58] Four different categories of MoRFs have been identified depending upon the dominant secondary structure formed in the bound state: α-helix forming α-MoRFs; β-strands forming β-MoRFs; i-MoRFs, which form irregular structures; and a mixture of the secondary structure containing complex-MoRFs.[58–60] Thus, disordered sequences have distinct structural features, and a thorough understanding of IDRP function requires the structural characterization of these functional elements. Here, we will describe the NMR methods by which an ensemble structure of IDRPs can be obtained.

3.1 Chemical shift-based methods

Chemical shifts of a nucleus are influenced by its local environment. The backbone chemical shifts (^1H_N, ^{15}N, $^{13}C'$, $1H_\alpha$, $^{13}C_\alpha$, $^{13}C_\beta$) are sensitive to the local secondary structures and can be used to determine the φ and ψ torsion angles.[61,62] The ^1H_N and ^{15}N chemical shifts are also sensitive to hydrogen bonding and electrostatic interactions.[63–65] Deviations of experimentally determined chemical shifts from random coil values, called the secondary shifts (Δδ), have been found to correlate with the protein backbone structure. Several datasets with random coil chemical shifts have been developed based on experimentally measured chemical shifts of peptides[66–68] or chemical shifts deposited in the Biological Magnetic Resonance Data Bank.[69–73] Random coil chemical shifts are influenced by neighboring residues,[70,71] pH, and temperature.[68,74,75] Hence, the random coil chemical shift library should be carefully selected for analysis.

In disordered sequences, different regions interchange between various conformations and, hence, their backbone chemical shifts have population-weighted average values of these conformers. Many approaches have been developed to determine the populations of these conformations from their chemical shifts. The neighbor-corrected Structural Propensity Calculator (ncSPC) algorithm uses a random coil database with corrections for neighboring residues and experimentally determined chemical shifts to obtain the populations of α-helix and β-extended conformations for each residue.[71,76] In another approach, RefDB[72] can be used to obtain amino acid-specific chemical shifts for the random coil, α-helical and β-extended conformations. The experimental chemical shifts are then used to determine the populations of these conformations. The δ2D method[77] extends the ncSPC algorithm also to predict the polyproline II conformations, which are frequently found in IDRPs.[38] Thus, the backbone chemical shifts, measured during the standard process of sequential assignment of IDRPs, can be used to determine the population distribution of local structures in the disordered sequences.

3.2 Residual dipolar coupling (RDC)

Residual Dipolar Coupling (RDC) can be used to determine local and long-range conformational sampling in IDRPs. A nucleus in a magnetic field behaves as a magnetic dipole and is influenced by nearby nuclei. This dipolar coupling between two nuclei depends on the angle that the internuclear vector forms with the external magnetic field. Due to the isotropic tumbling of molecules in the solution, dipolar coupling averages to zero. Weakly aligning media, such as lipid bicelles, filamentous phages, and stretched polyacrylamide gels, can be used to introduce bias in the tumbling of the molecules resulting in the appearance of residual dipolar coupling.[78–80] In IDRPs, localized secondary structural elements preferentially

align along the magnetic field, resulting in parallel and orthogonal orientation of the N-H bond vector in α-helical and β-extended conformations, respectively. This, in turn, results in measurable positive and negative RDC values for these conformations. Apart from local structural propensities, RDC is also sensitive to transient long-range order in IDRPs.[81,82]

3.3 Paramagnetic relaxation enhancement (PRE)

Paramagnetic Relaxation Enhancement (PRE) has been widely used to detect long-range contacts in IDRPs[83,84] and also for probing sparsely populated short-lived preformation of binding sites.[15] In PRE, a paramagnetic group, termed as the spin-label, is covalently attached to a cysteine sidechain of a protein. Different types of paramagnetic moieties can be introduced into the peptide chain.[85] PRE of a nucleus occurs due to its dipolar interaction with the unpaired electron on the spin-label. The signals of the amino acids that transiently approach the paramagnetic spin-label relax more efficiently, resulting in lowered signal intensities.[83] This method can detect transient interactions at distances of 20–30Å.[86]

3.4 Determination of ensemble structure of IDRPs

Due to their flexible nature, a single three-dimensional structure is not adequate to describe IDRPs. Instead, an ensemble description, which captures the various conformations sampled by different regions of IDRPs, is more appropriate.[87,88] The overall approach is to computationally generate a large set of IDRP conformations and then select conformations that agree with experimental results. This is referred to as the sample and select (SAS) approach.[89]

The initial pool of IDRP conformations can be generated using statistical coil generators,[90–93] where the polypeptide chain is created by adding amino acids with torsion angles randomly drawn from a statistical coil model. Alternatively, molecular dynamics (MD) simulations, with energy functions designed for disordered sequences,[94] can be used to generate the initial IDRP conformations.[15] Sampling enough conformations by MD simulations is a challenge, and techniques such as replica exchange molecular dynamics,[95] metadynamics,[96] and accelerated molecular dynamics,[97] have been used to address this problem. Restrained MD simulations have also been used for conformational sampling, where experimental data are used to bias the simulation.

The generated ensemble of conformations is pruned using experimental data to get the final representative ensemble of structures (Fig. 5). NMR parameters are calculated for each structure in the ensemble. Average values are calculated for a subset of structures and compared to the experimentally determined values, thereby leading to the final selection of structures in the IDRP ensemble. Several algorithms have been devised for the selection of conformations. The algorithm ENSEMBLE incorporates several NMR parameters such as chemical shifts, RDC, and PRE, for ensemble selection. It assigns weights to different conformations to maximize agreement with experimental data.[98,99] Another

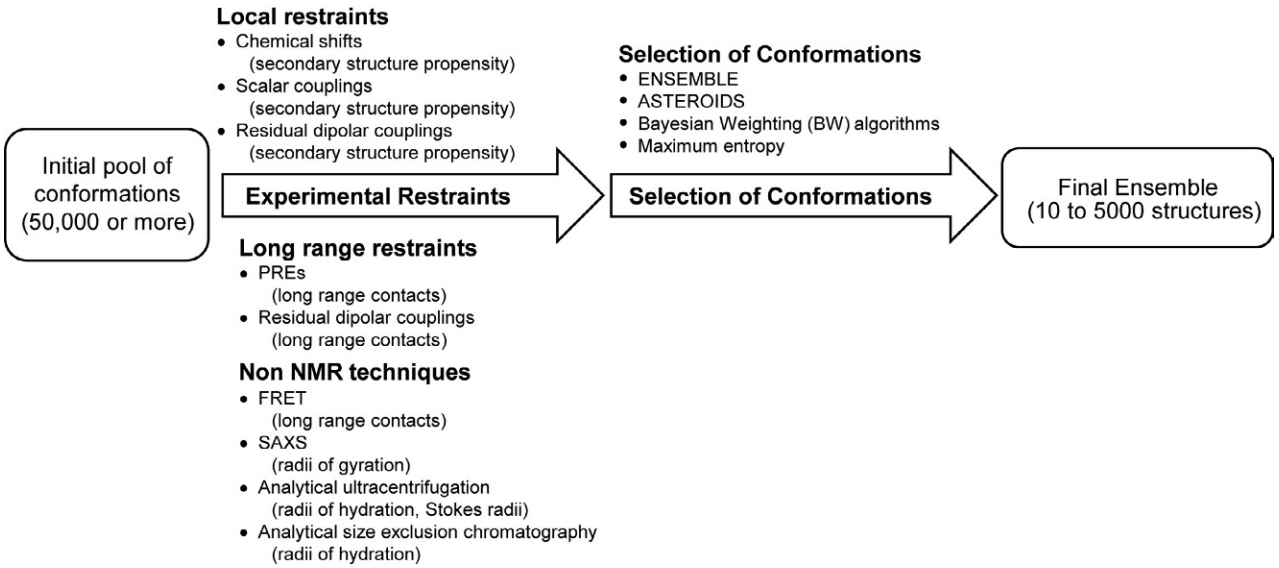

FIG. 5 Generation of ensemble structure of IDRPs. Schematic representation of the determination of ensemble structure of IDPs using the SAS approach. A computationally generated pool of IDRP conformations is pruned using experimental data to generate the final ensemble description.

program, ASTEROIDS, selects conformations from the initial pool using a genetic algorithm to generate the representative ensemble.[100] Statistical methods such as Bayesian weighting[101] and maximum entropy theory[102] have also been used for the selection of conformations.

It is important to note that the generation of an ensemble of conformers for an IDRP is an underdetermined problem, i.e., the number of conformations is much more than the number of restraints obtained from experiments. Hence, it is important to collect multiple experimental data. Apart from NMR spectroscopy, other biophysical techniques such as small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), analytical ultracentrifugation (AUC), and analytical size exclusion chromatography (SEC) can also be used to obtain complementary data (Fig. 5). Several IDRP ensemble structures are deposited in the Protein Ensemble Database.[103]

4 Characterization of IDRP dynamics

Apart from structural information, NMR spectroscopy also provides detailed information on protein dynamics at atomic resolution in a wide range of timescales (Fig. 6). IDRPs undergo librational motions in picoseconds (ps); segmental, chain-like, and torsion angle dynamics in nanoseconds (ns); binding interactions in microseconds (μs) to milliseconds (ms); and folding upon binding in ms or slower timescale.[104–109] In this section, we will describe the NMR experiments to study these dynamic motions in IDRPs.

4.1 Measuring fast timescale (ps-ns) dynamics

Backbone ^{15}N spin relaxation is sensitive to the fast ps - ns timescale motions of a protein. Relaxation is a process by which the bulk magnetization is restored to its thermal equilibrium after a perturbation is done to perform an NMR experiment. A spin ½ nucleus relaxes via dipole–dipole interactions between two nuclei and chemical shift anisotropy. These processes depend on the random tumbling motions of a molecule and result in small fluctuations in the magnetic field experienced by a nucleus. The ensemble average of these fluctuations is expressed as a correlation function $g(\tau)$

$$g(\tau) = \frac{1}{5} e^{-|\tau|/\tau_c}$$

where τC is the correlation time. The correlation function exponentially decays to insignificant values within a few nanoseconds. It decays fast for disordered regions and smaller proteins and slowly for larger proteins. The relative distribution of the frequencies of all motions in the ensemble can be obtained by the Fourier transformation of the correlation function $g(\tau)$, which results in the spectral density function $J(\omega)$.

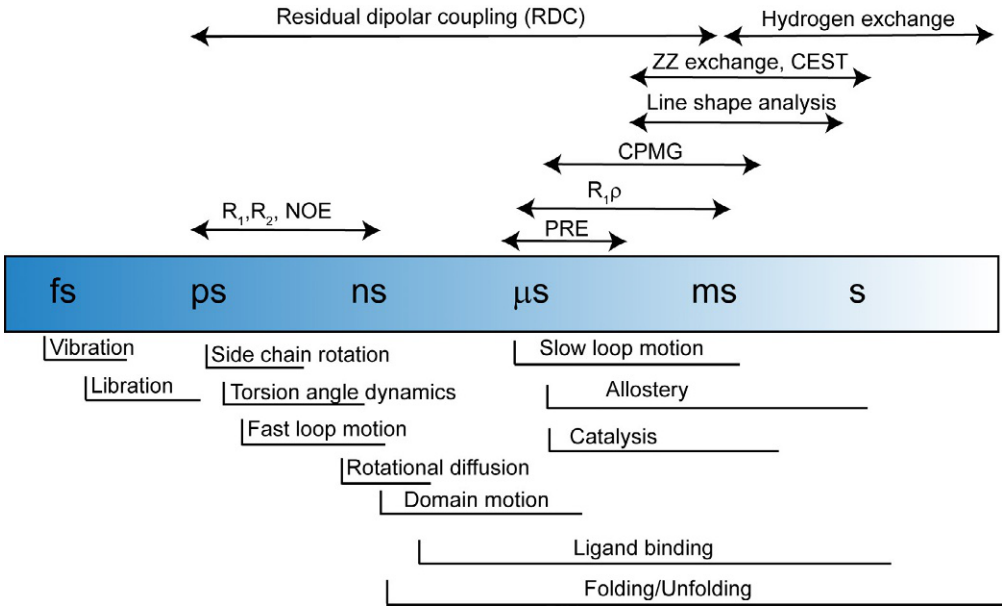

FIG. 6 Protein dynamics probed by NMR. Protein dynamics ranging from picoseconds to slower than seconds timescales can be probed at an atomic resolution using various NMR experiments.

$$J(\omega) = \frac{2}{5}\frac{\tau_C}{1+\omega^2\tau_C^2}$$

In practice, three relaxation parameters are measured by experiment: the rate of decay of ^{15}N bulk magnetization along the transverse (x-y) plane (R_2), the rate of its growth along the longitudinal z-axis (R_1), and ^{15}N polarization ($\{^1H\}$-^{15}N NOE) at the steady-state with the saturation of protons. These three parameters are measured for each backbone amide and are related to the spectral density functions evaluated at the frequencies 0, ω_N, ω_H, $\omega_H - \omega_N$ and $\omega_H + \omega_N$, where ω_N and ω_H are the Larmor frequencies of ^{15}N and ^1H nuclei, respectively. Since ω_H is tenfold larger than ω_N, the spectral density functions at the higher frequencies, i.e., $J(\omega_H - \omega_N)$, $J(\omega_H)$ and $J(\omega_H + \omega_N)$ can be combined to a single function of $J(0.87\omega_H)$.[110] The three spectral density functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ can be evaluated from the experimentally measured R_1, R_2, and $\{^1H\}$-^{15}N NOE.

$$J(0.87\omega_H) = \frac{4}{5}\frac{1}{d^2}\frac{\gamma_N}{\gamma_H}R_1(\mathrm{NOE}-1)\,\mathrm{srad}^{-1}$$

$$J(\omega_N) = \frac{1}{3d^2+4c^2}\left[4R_1 - 7d^2J(0.87\omega_H)\right]\mathrm{srad}^{-1}$$

$$J(0) = \frac{R_2 - \left(\frac{3}{8}d^2 + \frac{1}{2}c^2\right)J(\omega_N) - \frac{13}{8}d^2J(0.87\omega_H)}{\frac{d^2}{2}+\frac{2}{3}c^2}$$

where,

$$d = \frac{\mu_0 h\gamma_H\gamma_N}{8\pi^2}\frac{1}{r_{NH}^3}$$

$$c = \frac{\omega_N}{\sqrt{3}}(\sigma_\parallel - \sigma_\perp)$$

$$r_{NH} = 1.02^-$$

$$\sigma_\parallel - \sigma_\perp = -160\,\mathrm{ppm\ for\ backbone\ NH}; -89.6\,\mathrm{ppm\ for\ Trp\ residue\ NH}$$

These three frequencies depend on the magnetic field strength of the spectrometer. For a spectrometer with 600 MHz proton Larmor frequency, $\omega_N = 3.8\times10^8$ rad/s and $0.87\omega_H = 3.3\times10^9$ rad/s. Thus, $J(0)$ and $J(0.87\omega_H)$ represent the relative contributions of the slower and faster motions in a molecule, respectively. For the transcription factor sex combs reduced (SCR), which has a folded domain and an intrinsically disordered region, the slower motions are dominant in the folded domain, whereas the faster motions are present more in the disordered region (Fig. 7A).

4.2 Identifying rigid segments in IDRPs

IDRPs have small linear motifs, which facilitate their interactions with partner molecules. The eukaryotic linear motif resource lists ~3615 short linear motifs (SLiMs), and it is estimated that the eukaryotic proteome may contain ~10^5 such motifs.[111] Identification of such motifs in IDRPs by bioinformatics methods fails due to the high rate of mutation and very low sequence conservation in IDRPs. Moreover, a lack of knowledge of the binding partner and weak interactions between IDRPs and partner molecules make it difficult to experimentally identify these linear motifs. For the transcription factors SCR and deformed (DFD), it has been shown that these functional motifs are also dynamically rigid and can be identified from the backbone ^{15}N relaxation data.[109]

Residue-wise flexibility was determined from R_1, R_2, and heteronuclear ^{15}N-NOE for two *Drosophila* HOX transcription factors SCR and DFD. Reduced spectral density analysis revealed varying degrees of flexibility in the disordered region of these proteins (Fig. 7A). Rigid and flexible segments in the disordered region were identified from the relative contribution of the lower frequency ($J(0)$ and $J(\omega_N)$) and higher frequency ($J(0.87\omega_H)$) components. For rigid segments, the spectral density function was dominated by the lower frequency components, and for flexible segments, the higher frequency component also made a significant contribution. Based on this observation, the product $J(0)^*J(\omega_N)/J(0.87\omega_H)$ was suggested as a sensitive parameter to identify rigid and flexible segments within a disordered sequence.

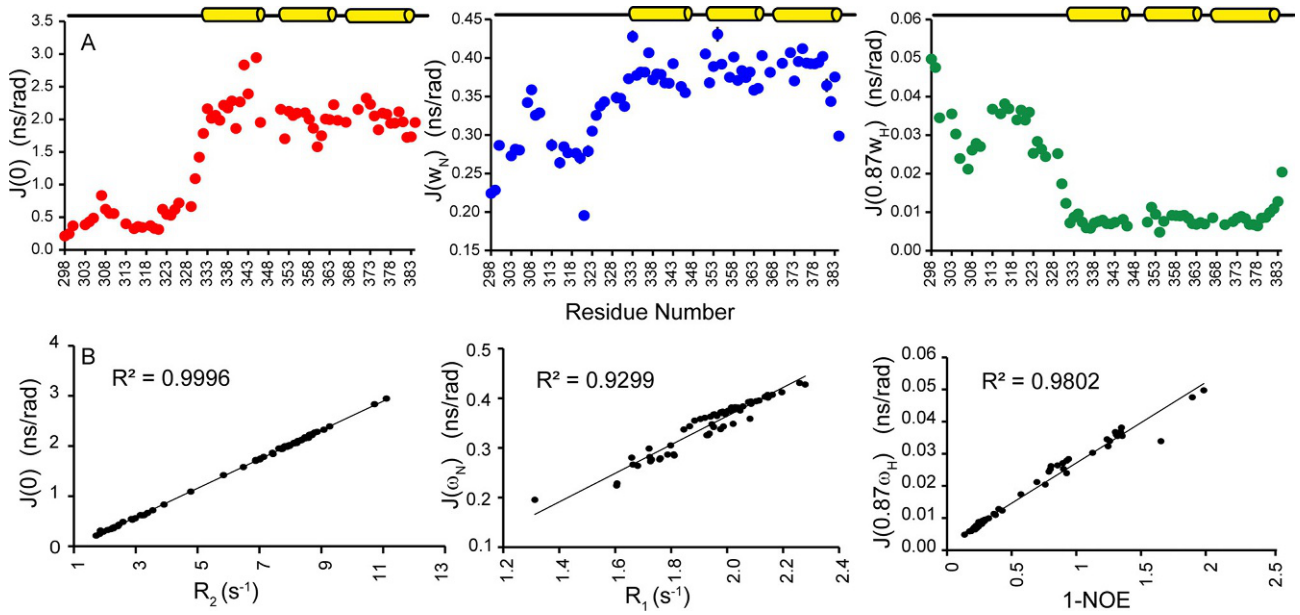

FIG. 7 Spectral density mapping. (A) Reduced spectral density values for SCR. Cylinders represent helices in the folded domain, and the N-terminus residues are disordered. (B) Correlation between spectral density values and their corresponding NMR relaxation parameters for all the residues for SCR. (Adapted with permission from Maiti et al. (2019) J Mol Biol 2019;431:1353–1369.)

It turns out that the spectral density functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ are dominated by R_2, R_1, and (1-NOE) values, respectively (Fig. 7B). Hence, the product $J(0)*J(\omega_N)/J(0.87\omega_H)$ was replaced by the equivalent product $R_1R_2/(1\text{-NOE})$ to obtain residue-wise rigidity. Consecutive residues with $R_1R_2/(1\text{-NOE})$ values significantly more than the average of the disordered region were identified as a rigid segment (Fig. 8). This analysis identified a short stretch of residues in the disordered region of both SCR and DFD, which was also shown to interact specifically with a partner transcription factor extradenticle.[109] Thus, a simple method, based on well-established NMR relaxation experiments, was presented to identify functionally important rigid segments in intrinsically disordered regions of proteins.

4.3 Determination of global flexibility of disordered sequences

The model-free formalism,[112] which is extensively used to interpret the backbone dynamics of folded proteins, cannot be used for disordered sequences as the global tumbling motion of the protein and residue-wise internal motions do not have significantly different timescales and, hence, cannot be separated. Instead, a segmental motion model[113] can be used. For disordered sequences, the torsion angles (φ and ψ) sampled by an amino acid are influenced by its neighbors.[114] Thus, a polypeptide chain is not as flexible as a freely jointed chain, and its flexibility depends on the amino acid composition. One way to measure the overall flexibility of a disordered sequence is to determine its persistence length; the longer the persistence length, the stiffer is the polypeptide chain. Schwalbe et al. have shown that the transverse relaxation rate (R_2) can be fitted to the following equation to determine the persistence length of a polypeptide chain.[113]

$$R_2(\mathrm{i}) = R_{\mathrm{int}} \sum_{J=\mathrm{i}}^{N} \mathrm{e}^{\frac{|i-j|}{\lambda_0}}$$

where R_{int} is the intrinsic relaxation rate, λ_0 is the persistence length in a number of residues, and N is the total number of residues in the polypeptide chain. Ri_{nt} depends on the temperature and viscosity of the solution.

4.4 Slow dynamics (μs-ms) in IDRPs

IDRPs also contain molecular recognition features (MoRFs), which are 10 to 70 residues long and undergo folding upon binding to partner molecules. Structural changes in these disordered sequences upon binding can be tracked by NMR titration experiments. In a typical experiment, the IDRP is ^{15}N-labeled, and the unlabelled partner molecule (protein, DNA or RNA) is titrated in. At each titration point, ^{15}N HSQC or HMQC spectra are collected. Interacting residues can be identified by the change in their chemical shifts. Representative examples include folding of disordered p21 residues upon binding CDK[115],

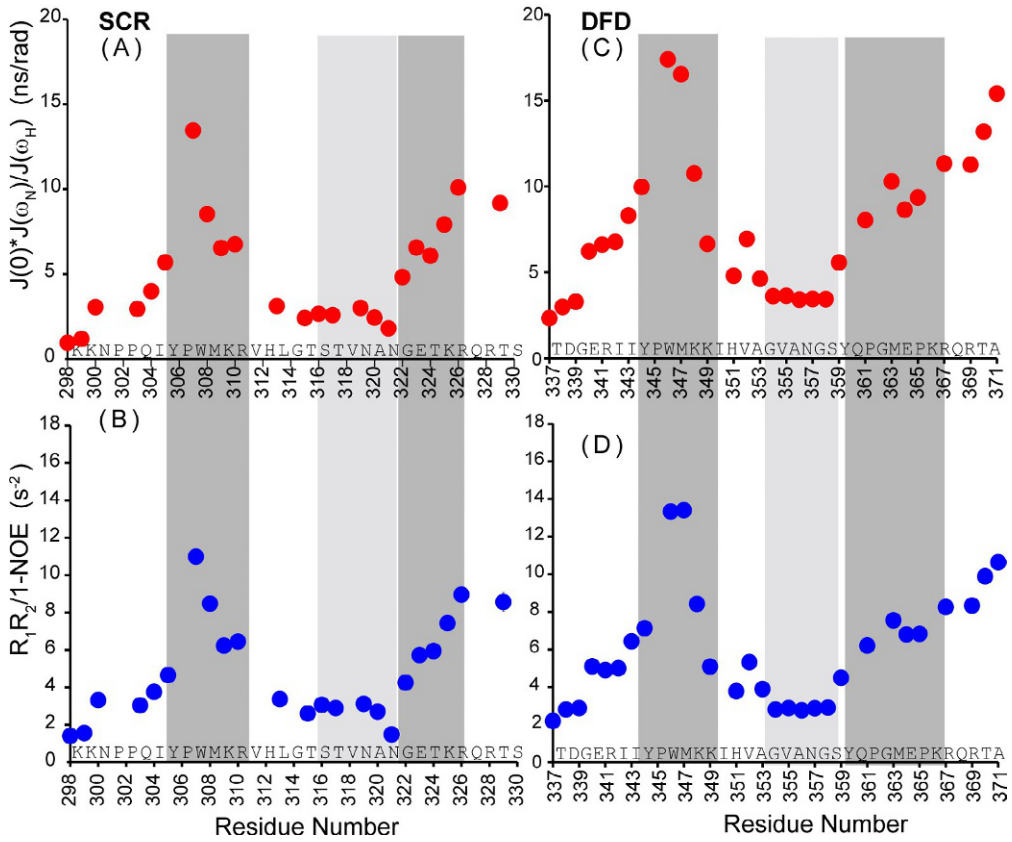

FIG. 8 Identification of rigid segments. The rigid and flexible segments in the IDRs are readily identified from the residue-wise rigidity plot of $J(0)*J(\omega_N)/J(0.87\omega_H)$ for the disordered region of (A) SCR and (C) DFD. The same segments can also be identified directly from the plot of $R_1*R_2/(1\text{-NOE})$ for (B) SCR and (D) DFD. The rigid and flexible segments are highlighted in *dark gray* and *light gray*, respectively. (Adapted with permission from Maiti et al. (2019) J Mol Biol 2019;431:1353–1369.)

formation of stable helices upon DNA binding by LEF-1,[116] folding upon binding of the disordered pKID to the KIX domain of CREB binding protein,[117] binding of the disordered region of p53 to TAZ2 domain of CBP[118], binding of disordered C-terminal region of Artemis to the DNA binding domain of Ligase-IV[119], and folding of 4E-BP2 upon eIF4E-binding.[120]

Binding and folding of IDRPs occur at the slower μs-ms timescale and can be studied by relaxation dispersion,[121] CEST[122], and lineshape analysis experiments.[123] Often IDRPs form weak complexes so that the NMR peaks corresponding to the bound state are broadened due to chemical exchange and become undetectable. NMR peaks corresponding to the bound state can also become undetectable if the IDRP binds to a large protein. In such cases, the invisible bound state of the IDRP can be characterized by relaxation dispersion and CEST experiments.[117,119,124,125] A ^{15}N and ^{13}C labeled IDRP is titrated with the sub-stoichiometric amount of unlabelled partner and relaxation dispersion and CEST experiments can be used to determine the amide ^{15}N, $^{13}C'$, $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts of the bound state, which can further be used to predict the secondary structure of the bound state of IDRP.

IDRP binding to partner molecules may also occur via multiple steps. Using a three-state model, the disordered pKID was shown to bind the KIX domain via an intermediate.[117] Recently, a general approach for the analysis of multiple site exchange of relaxation dispersion data has been proposed.[126] The titration data can also be analyzed by lineshape analysis, which can accommodate more complex binding schemes.[127] It provides both thermodynamic parameters, i.e., equilibrium (K_{EQ}) and dissociation constants (K_D), and kinetic parameters, i.e., binding (k_{ON}) and dissociation (k_{OFF}) rates for each step of the binding scheme. Lineshape analysis was used to analyze the six-state binding and catalysis model of the disordered cytoplasmic tail of the amyloid precursor protein (APP) and the enzyme Pin1.[128,129]

5 *In-cell* NMR experiments

The recent development of *in-cell* NMR techniques allowed the study of proteins under physiological conditions compared to in-vitro methods. The intracellular environment can influence proteins' structural conformations, dynamics, and

function. Folded proteins suffer from slower tumbling motions due to the crowding effect in a cell, which results in peak broadening and poor detection of NMR signals. IDRPs have faster motions resulting in relatively sharper peaks and hence are better suited for study by *in-cell* NMR. In-cell NMR helps to study the effects of increased viscosity, molecular crowding, and other intracellular environments on IDRPs and hence, aid in characterizing the IDRPs structure,[130] dynamics,[131] and interactions.[132] To observe the *in-cell* NMR spectra, the protein of interest is labeled with ^{13}C and ^{15}N isotopes and over-expressed in bacterial cells. The bacterial cells are packed in NMR tubes for experiments. Overexpression of the target protein is sufficient to suppress signals from other bacterial proteins.[133]

For NMR studies in a eukaryotic cell, the ^{13}C and ^{15}N labeled proteins are overexpressed and purified from bacteria and delivered into the eukaryotic cell. The labeled proteins are incorporated into cells by microinjection,[134,135] cell-penetrating peptide-mediated endocytic transportation,[130] diffusion through pore-forming toxins,[136] or electroporation.[131] HSQC, HMQC, and CON are used as fingerprint spectra. These are assigned using ^{15}N and ^{13}C labeled in-vitro samples, and the assignments are transferred to the in-vivo spectra. However, different chemical environment in cells can cause peak shifts, and these should be assigned using *in-cell* triple resonance experiments.[137] Due to the limited lifetime of cells, rapid measurement of multi-dimensional experiments is achieved using non-uniform sampling and BEST experiments.

The disordered microtubule-associated protein Tau is involved in the regulation and stabilization of microtubules and actin networks, and its aggregation is implicated in several neurodegenerative diseases.[138] *In-cell* NMR studies of Tau in *Xenopus laevis* oocytes[139] and HEK293 cells[140] showed that the protein remains disordered in a cell. Chemical shift changes and line broadening helped to identify phosphorylation and protein interaction sites on Tau in the cellular environment. Human α-synuclein (αSyn) forms amyloid aggregations in the brains of patients with Parkinson's disease[141] and is shown to be intrinsically disordered by in-vitro studies. Using *in-cell* NMR experiments of αSyn in non-neuronal A2780 and HeLa cells, and neuronal B65, SK-N-SH, and RCSN-3 cells, this protein was found to be indeed disordered.[131] Interestingly, these studies also revealed that *in-cell* αSyn adopts a more compact conformation than in-vitro conditions and exists in a predominantly monomeric state.

6 Conclusions and future perspectives

In this chapter, we have discussed important developments in NMR methods to study the structure and dynamics of IDRPs. Although impressive progress has been made in this field in the past 15 years, there is scope for further improvements. One particular improvement that will help tremendously is the development of force fields that can simulate realistic structures for both folded and disordered sequences. Currently, available force fields work better for one or the other and fail for IDRPs containing long disordered regions appended to folded domains. NMR is a technique that provides an ensemble average of the measured structural and dynamic properties. Since folded proteins have a unique structure with limited flexible regions, these properties can be interpreted with a small ensemble of 10 to 20 structures. For IDRPs, it is not clear how many ensemble structures are needed for such an interpretation. However, it is understood that the number of structures in the ensemble should increase with the number of residues and flexibility of the disordered region and should be in the range of several thousand structures. In addition, further improvement in spectral resolution, sensitivity, and fast data acquisition will help studies of IDRPs by NMR spectroscopy.

References

1. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* 2005;6:197–208.
2. Dunker AK, et al. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disord Proteins.* 2013;1, e24157.
3. Van Der Lee R, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014;114:6589–6631.
4. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol.* 2004;337:635–645.
5. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. *Biochemistry.* 2002;41:6573–6582.
6. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol.* 2002;323:573–584.
7. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform.* 2000;11:161–171.
8. Dunker AK, et al. Intrinsically disordered protein. *J Mol Graph Model.* 2001;19:26–59.
9. Liu J, et al. Intrinsic disorder in transcription factors. *Biochemistry.* 2006;45:6873–6888.
10. Palombo M, et al. The relationship between folding and activity in UreG, an intrinsically disordered enzyme. *Sci Rep.* 2017;7:1–10.
11. Rumi-Masante J, Rusinga FI, Lester TE, et al. Structural basis for activation of calcineurin by calmodulin. *J Mol Biol.* 2012;415:307–317.

12. Ding Z, et al. A 22-mer segment in the structurally pliable regulatory domain of metazoan CTP: phosphocholine cytidylyltransferase facilitates both silencing and activating functions. *J Biol Chem.* 2012;287:38980–38991.
13. Radford SE. Protein folding: progress made and promises ahead. *Trends Biochem Sci.* 2000;25:611–618.
14. Dill KA, Chan HS. From levinthal to pathways to funnels. *Nat Struct Biol.* 1997;4:10–19.
15. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol.* 2011;21:426–431.
16. Papoian GA. Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proc Natl Acad Sci U S A.* 2008;105:14237–14238.
17. Galea CA, Wang Y, Sivakolundu SG, Kriwacki RW. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry.* 2008;47:7598–7609.
18. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol.* 2002;323:573–584.
19. Xie H, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res.* 2007;6:1882–1898.
20. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys.* 2008;37:215–246.
21. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets: the roles of intrinsic disorder in protein interaction networks. *FEBS J.* 2005;272:5129–5148.
22. Haynes C, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol.* 2006;2:0890–0901.
23. Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol.* 2008;4:179.
24. Uversky VN. Introduction to intrinsically disordered proteins (IDPs). *Chem Rev.* 2014;114:6557–6560.
25. Vacic V, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol.* 2012;8, e1002709.
26. Basak AJ, et al. Structural insights into N-terminal IgV domain of BTNL2, a T cell inhibitory molecule, suggests a non-canonical binding interface for its putative receptors. *J Mol Biol.* 2020;432:5938–5950.
27. Boral S, Maiti S, Basak AJ, Lee W, De S. Structural, dynamic, and functional characterization of a DnaX Mini-intein derived from Spirulina platensis provides important insights into Intein-mediated catalysis of protein splicing. *Biochemistry.* 2020;59:4711–4724.
28. Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. *J Biomol NMR.* 2001;20:135–147.
29. Bhavesh NS, Panchal SC, Hosur RV. An efficient high-throughput resonance assignment procedure for structural genomics and protein folding research by NMR. *Biochemistry.* 2001;40:14727–14735.
30. Chakraborty S, Paul S, Hosur RV. Simultaneous acquisition of 13C α- 15N and 1H- 15N-15N sequential correlations in proteins: application of dual receivers in 3D HNN. *J Biomol NMR.* 2012;52:5–10.
31. Kumar D, Paul S, Hosur RV. BEST-HNN and 2D-(HN)NH experiments for rapid backbone assignment in proteins. *J Magn Reson.* 2010;204:111–117.
32. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc.* 1992;114(16):6291–6293. https://doi.org/10.1021/ja00042a003.
33. Kay LE, et al. Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson (1969).* 1990;89(3):496–514. https://doi.org/10.1016/j.jmr.2011.09.004.
34. Clubb RT, et al. A constant-time three-dimensional triple resonance pulse scheme to correlate Intra residue 'H N, 15N, and 13C' chemical shifts in "N-13C-labeled proteins". *J Magn Reson.* 1992;97(1):213–217. https://doi.org/10.1016/0022-2364(92)90252-3.
35. Bermel W, et al. 13C Direct-detection biomolecular NMR. *Concepts in Magnetic Resonance Part A.* vol. 32A. Wiley InterScience; 2008:183–200.
36. Bastidas M, Gibbs EB, Sahu D, Showalter SA. A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts Magn Reson Part A.* 2015;44:54–66.
37. O'Hare B, et al. Incorporating ^1H chemical shift determination into ^{13}C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J Magn Reson.* 2009;200(2):354–358. https://doi.org/10.1016/j.jmr.2009.07.014.
38. Theillet F-X, et al. The alphabet of intrinsic disorder: I. Act like a Pro: on the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins.* 2013;1:24360.
39. Goradia N, et al. An approach to NMR assignment of intrinsically disordered proteins. *ChemPhysChem.* 2015;16:739–746.
40. Felli IC, et al. Exclusively heteronuclear NMR experiments for the investigation of intrinsically disordered proteins: focusing on proline residues. *Magn Reson.* 2021;2:511–522.
41. Pierattelli R, Bermel W, Felli IC, Ku R. ^{13}C Direct-detection biomolecular NMR. *Concepts Magn Reson Part A.* 2008;183–200. https://doi.org/10.1002/cmr.a.20109.
42. Rao Kakita VM, et al. An efficient combination of BEST and NUS methods in multidimensional NMR spectroscopy for high throughput analysis of proteins. *RSC Adv.* 2018;8:17616–17621.
43. Orekhov VY, Jaravine VA. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog Nucl Magn Reson Spectrosc.* 2011;59:271–292.
44. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. *J Biomol NMR.* 2012;52:315–327.
45. Ying J, Delaglio F, Torchia DA, Bax A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR.* 2017;68:101–118.
46. Hansen DF. Using deep neural networks to reconstruct non-uniformly sampled NMR spectra. *J Biomol NMR.* 2019;73:577–585.

47. Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc.* 2006;128:9042–9043.
48. Bogatyreva NS, Finkelstein AV, Galzitskaya OV. Trend of amino acid composition of proteins of different taxa. *J Bioinforma Comput Biol.* 2006;4:597–608.
49. Tompa P. On the supertertiary structure of proteins. *Nat Chem Biol.* 2012;8:597–600.
50. Muir TW. Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem.* 2003;72:249–289.
51. Aranko AS, et al. Structure-based engineering and comparison of novel split inteins for protein ligation. *Mol BioSyst.* 2014;10:1023–1034.
52. Mao H, Hart SA, Schink A, Pollok BA. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc.* 2004;126:2670–2671.
53. Kigawa T, Muto Y, Yokoyama S. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J Biomol NMR.* 1995;6:129–134.
54. Staunton D, Schlinkert R, Zanetti G, Colebrook SA, Campbell ID. Cell-free expression and selective isotope labelling in protein NMR. *Magn Reson Chem.* 2006;44:2–9.
55. Yokoyama J, Matsuda T, Koshiba S, Tochio N, Kigawa T. A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Anal Biochem.* 2011;411:223–229.
56. Hoffmann B, Löhr F, Laguerre A, Bernhard F, Dötsch V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Prog Nucl Magn Reson Spectrosc.* 2018;105:1–22.
57. Davey NE, et al. Attributes of short linear motifs. *Mol BioSyst.* 2012;8:268–281.
58. Oldfield CJ, et al. Coupled folding and binding with α-helix-forming molecular recognition elements. *Biochemistry.* 2005;44:12454–12470.
59. Disfani FM, et al. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics.* 2012;28:75–83.
60. Vacic V, et al. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res.* 2007;6:2351–2366.
61. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR.* 1994;4:171–180.
62. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR.* 1999;13:289–302.
63. De Dios AC, et al. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab Initio approach. *Adv Sci.* 1993;260:1491–1496.
64. Asakura T, Taoka K, Demura M, Williamson MP. The relationship between amide proton chemical shifts and secondary structure in proteins. *J Biomol NMR.* 1995;6:227–236.
65. Moon S, Case DA. A new model for chemical shifts of amide hydrogens in proteins. *J Biomol NMR.* 2007;38:139–150.
66. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR.* 1995;5:67–81.
67. Schwarzinger S, Kroon GJA, Foss TR, Wright PE, Dyson HJ. Erratum: Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView (Journal of Biomolecular NMR (2000) vol. 18 (43–48)). *J Biomol NMR.* 2001;19:95.
68. Kjaergaard M, Poulsen FM. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR.* 2011;50:157–165.
69. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol.* 1991;222:311–333.
70. De Simone A, Cavalli A, Hsu STD, Vranken W, Vendruscolo M. Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J Am Chem Soc.* 2009;131:16332–16333.
71. Tamiola K, Acar B, Mulder FAA. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc.* 2010;132:18000–18003.
72. Zhang H, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR.* 2003;25:173–195.
73. Peti W, Smith LJ, Redfield C, Schwalbe H. Chemical shifts in denatured proteins: resonance assignments for denatured ubiquitin and comparisons with other denatured proteins. *J Biomol NMR.* 2001;19:153–165.
74. Jensen MR, Ruigrok RWH, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol.* 2013;23:426–435.
75. Kjaergaard M, et al. Temperature-dependent structural changes in intrinsically disordered proteins: formation of α-helices or loss of polyproline II? *Protein Sci.* 2010;19:1555–1564.
76. Tamiola K, Mulder FAA. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem Soc Trans.* 2012;40:1014–1020.
77. Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry.* 2012;51:2224–2231.
78. Tjandra N, Bax A. Crystalline medium. *Science.* 1997;278:1111–1114.
79. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol.* 1998;5:1065–1074.
80. Clore GM, Starich MR, Gronenborn AM. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc.* 1998;120:10571–10572.
81. Jensen MR, et al. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure.* 2009;17:1169–1185.

82. Salmon L, et al. NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc.* 2010;132:8407–8418.
83. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol.* 1997;268:158–169.
84. Clore GM, Tang C, Iwahara J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol.* 2007;17:603–616.
85. Su XC, Otting G. Paramagnetic labelling of proteins and oligonucleotides for NMR. *J Biomol NMR.* 2010;46:101–112.
86. Marius Clore G, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev.* 2009;109:4108–4139.
87. Kosol S, Contreras-Martos S, Cedeño C, Tompa P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules.* 2013;18:10802–10828.
88. Salvi N. *Ensemble Descriptions of IDPs and IDRs: Integrating Simulation and Experiment. Intrinsically Disordered Proteins.* Elsevier Inc.; 2019. https://doi.org/10.1016/b978-0-12-816348-1.00002-8.
89. Jensen MR, Zweckstetter M, Huang JR, Blackledge M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev.* 2014;114:6632–6660.
90. Feldman HJ, Hogue CWV. A fast method to sample real protein conformational space. *Proteins Struct Funct Genet.* 2000;39:112–131.
91. Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci U S A.* 2005;102:13099–13104.
92. Bernadó P, et al. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A.* 2005;102:17002–17007.
93. Ozenne V, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 2012;28:1463–1470.
94. Huang J, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2016;14:71–73.
95. Sugita Y, Okamoto Y. Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol Biol.* 2007;350:205–223.
96. Leone V, Marinelli F, Carloni P, Parrinello M. Targeting biomolecular flexibility with metadynamics. *Curr Opin Struct Biol.* 2010;20:148–154.
97. Pierce LCT, Salomon-Ferrer R, De Oliveira CAF, McCammon JA, Walker RC. Routine access to millisecond time scale events with accelerated molecular dynamics. *J Chem Theory Comput.* 2012;8:2997–3002.
98. Marsh JA, Forman-Kay JD. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J Mol Biol.* 2009;391:359–374.
99. Krzeminski M, Marsh JA, Neale C, Choy WY, Forman-Kay JD. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics.* 2013;29:398–399.
100. Nodet G, et al. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc.* 2009;131:17908–17918.
101. Fisher CK, Huang A, Stultz CM. Modeling intrinsically disordered proteins with Bayesian statistics. *J Am Chem Soc.* 2010;132:14919–14927.
102. Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol.* 2014;10:1–9.
103. Lazar T, et al. PED in 2021: a major update of the protein ensemble database for intrinsically disordered proteins. *Nucleic Acids Res.* 2021;49:D404–D411.
104. Khan SN, et al. Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation. *Biophys J.* 2015;109:988–999.
105. Abyzov A, et al. Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J Am Chem Soc.* 2016;138:6240–6251.
106. Mollica L, et al. Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment. *Front Mol Biosci.* 2016;3:1–18.
107. Narayanan C, Bafna K, Roux LD, Agarwal PK, Doucet N. Applications of NMR and computational methodologies to study protein dynamics. *Arch Biochem Biophys.* 2017;628:71–80.
108. Salvi N, Abyzov A, Blackledge M. Analytical description of NMR relaxation highlights correlated dynamics in intrinsically disordered proteins. *Angew Chem Int Ed.* 2017;56:14020–14024.
109. Maiti S, et al. Dynamic studies on intrinsically disordered regions of two paralogous transcription factors reveal rigid segments with important biological functions. *J Mol Biol.* 2019;431:1353–1369.
110. Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR.* 1995;6:153–162.
111. Van Roey K, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014;114:6733–6778.
112. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc.* 1982;104:4559–4570.
113. Schwalbe H, et al. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry.* 1997;36:8977–8991.
114. Pappu RV, Srinivasan R, Rose GD. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc Natl Acad Sci U S A.* 2000;97:12565–12570.
115. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A.* 1996;93:11504–11509.

116. Love JJ, Li X, Chung J, Dyson HJ, Wright PE. The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry.* 2004;43:8725–8734.
117. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 2007;447:1021–1025.
118. Arai M, Ferreon JC, Wright PE. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J Am Chem Soc.* 2011;134:3792–3803.
119. Charlier C, et al. Structure and dynamics of an intrinsically disordered protein region that partially folds upon binding by chemical-exchange NMR. *J Am Chem Soc.* 2017;139:12219–12227.
120. Dawson R, et al. The N-terminal domain of p53 is natively unfolded. *J Mol Biol.* 2003;332:1131–1141.
121. Palmer III AG, Kroenke CD, Loria JP. Nuclear magnetic resonance emethods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 2001;339:204–211.
122. Vallurupalli P, Bouvignies G, Kay LE. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc.* 2012;134:8148–8161.
123. Waudby CA, Christodoulou J. Chapter 24. *Breast Cancer.* 2020;2141:507–538.
124. Kragelj J, et al. Structure and dynamics of the MKK7-JNK signaling complex. *Proc Natl Acad Sci U S A.* 2015;112:3409–3414.
125. Schneider R, et al. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *J Am Chem Soc.* 2015;137:1220–1229.
126. Koss H, Rance M, Palmer AG. Algebraic expressions for Carr-Purcell-Meiboom-Gill relaxation dispersion for N-site chemical exchange. *J Magn Reson.* 2020;321:106846.
127. Kovrigin EL. NMR line shapes and multi-state binding equilibria. *J Biomol NMR.* 2012;53:257–270.
128. Greenwood AI, et al. Complete determination of the Pin1 catalytic domain thermodynamic cycle by NMR lineshape analysis. *J Biomol NMR.* 2011;51:21–34.
129. De S, et al. Complete thermodynamic and kinetic characterization of the isomer-specific interaction between Pin1-WW domain and the amyloid precursor protein cytoplasmic tail phosphorylated at Thr668. *Biochemistry.* 2012;51:8583–8596.
130. Inomata K, et al. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature.* 2009;458:106–109.
131. Theillet FX, et al. Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature.* 2016;530:45–50.
132. Burz DS, Dutta K, Cowburn D, Shektman A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR) David. *Nat Methods.* 2006;3:91–93.
133. Serber Z, Ledwidge R, Miller SM, Dötsch V. Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J Am Chem Soc.* 2001;123:8895–8901.
134. Bodart JF, et al. NMR observation of Tau in Xenopus oocytes. *J Magn Reson.* 2008;192:252–257.
135. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc Natl Acad Sci U S A.* 2006;103:11904–11909.
136. Ogino S, et al. Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J Am Chem Soc.* 2009;131:10834–10835.
137. Hamatsu J, et al. High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc.* 2013;135:1688–1691.
138. Lee V-Y, Goedert M, Trojanowski JQ. Neurodegenerative auopathies. *Annu Rev Neurosci.* 2001;24:1121–1161.
139. Selenko P, et al. In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat Struct Mol Biol.* 2008;15:321–329.
140. Zhang S, et al. In-cell NMR study of Tau and MARK2 phosphorylated Tau. *Int J Mol Sci.* 2019;20:1–14.
141. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. *Nat Rev Neurol.* 2013;9:13–24.