



**HAL**  
open science

# Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance

Elise Delaforge, Tiago N Cordeiro, Pau Bernadó, Nathalie Sibille

► **To cite this version:**

Elise Delaforge, Tiago N Cordeiro, Pau Bernadó, Nathalie Sibille. Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance. Graham A. Webb. Modern Magnetic Resonance, Springer International Publishing, pp.381-399, 2018, 978-3-319-28387-6. 10.1007/978-3-319-28388-3\_52 . hal-02348118

**HAL Id: hal-02348118**

**<https://hal.science/hal-02348118>**

Submitted on 21 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Conformational characterization of intrinsically disordered proteins and its biological significance

Elise Delaforge<sup>1</sup>, Tiago N. Cordeiro<sup>2</sup>, Pau Bernadó<sup>1</sup>, and Nathalie Sibille<sup>1,\*</sup>

<sup>1</sup>*Centre de Biochimie Structurale, CNRS UMR 5048 - INSERM U1054 - Université de Montpellier, 29, rue de Navacelles, 34090 Montpellier, France*

<sup>2</sup>*X-ray and Neutron Science, Niels Bohr Institute, Copenhagen, 2100, Denmark*

*Corresponding author \*:* Nathalie Sibille, [nathalie.sibille@cbs.cnrs.fr](mailto:nathalie.sibille@cbs.cnrs.fr)

## Abstract

Intrinsically Disordered Proteins (IDPs) perform a large variety of functions that are crucial for signalling, cell regulation and homeostasis. Functions performed by IDPs are complementary to those executed by their globular counterparts demonstrating that the biophysical properties of disordered proteins dictate their functional mechanisms. Conformational plasticity, large solvent accessibility, and transient structuration are inherent characteristics of IDPs that are ideal to modulate partner recognition in signalling processes. As a consequence, the characterization of the structural features of IDPs in their free state and in complex with the relevant biological partners are crucial to reveal the molecular basis of signalling and cell regulation. Nuclear Magnetic Resonance (NMR) is the only structural biology technique to derive structural and dynamic information of IDPs at the residue level. NMR can probe the conformational landscape of IDPs and monitor the changes exerted by environmental conditions, post-translational modifications, or recognition events. Through multiple parameters, NMR is a rich source of structural and dynamic information covering residue-specific conformational, and long-range intramolecular interactions. Additionally, NMR data can be complemented by information obtained from other biophysical techniques such as Small-Angle X-ray Scattering (SAXS) that probe the overall properties of the protein in solution. This chapter aims at describing the main technical and conceptual developments that have enabled NMR to decipher the structural basis of the biological role of IDPs in multitude of crucial biological processes.

## Introduction

During the last decades, intrinsically disordered proteins (IDPs) have been revealed as pivotal actors in diverse biological processes. Indeed, it has been demonstrated that the characterization of intrinsically disordered proteins (IDPs) or domains is needed as more than 30 % of eukaryotic proteins have disordered regions of length equal or superior to 50 consecutive residues. These proteins, that lack stable tertiary and/or secondary structures, are more common in Eukaryota than in Bacteria and Archaea [1], probably due to their biological complexity. IDPs have a very specific amino acid composition that is rich in charged, structure-breaking residues (proline and glycine), and in alanine, while they are depleted in order-promoting amino acids such as hydrophobic and aromatic residues [2].

IDPs are associated to distinct and complementary functions and localizations to those of their globular counterparts. Order is found in enzymes performing catalytic functions, whereas disordered is linked to signalling, regulation, and transport. A key aspect to explain this preponderance is linked to how disordered chains recognize their partners. Partner recognition of IDPs is normally performed through short linear motifs (SLiMs) that are embedded in disordered

poorly conserved regions. A crucial feature of IDPs lies on the entropic cost associated to the disorder-to-order transitions observed in SLiMs upon partner recognition. This entropic cost very often implies low to moderate affinities. This particularity in combination with the high specificity of the interactions, make IDPs very well suited agents for signalling processes [3]. It also explains the capacity of many IDPs to present one-to-many and many-to-one interactions, or binding promiscuity, explaining the enrichment of disorder in interactome hubs [4][5]. Moreover, many IDPs are submitted to additional regulation processes governed by alternative splicing and post-translational modifications (PTMs) (i.e. phosphorylation, acetylation, N- or O-linked glycosylation, methylation, ubiquitination...), which are facilitated by the accessibility of modifying enzymes to disordered chains. Given the important number of above-mentioned sensitive biological processes linked with disorder, it is not surprising that the vast majority of proteins involved in cancer, cardiovascular diseases or neurodegeneration are predominantly disordered.

The relevance of disorder in biology and disease has fostered and intensified research aiming at deciphering the structural basis of their function. In that context, Nuclear Magnetic Resonance (NMR) is the most adapted structural biology technique to address these questions. In this chapter, we describe how NMR can probe the structural and dynamic properties of IDPs at residue level. Its sensitivity to changes in the chemical environment enables NMR to monitor partner-recognition processes or the incorporation of PTMs [6–8]. We also address the technical and conceptual challenges associated to ensemble representations, and how the combination of multiple NMR parameters and/or information from other techniques can help to obtain more accurate conformational ensembles of IDPs and their complexes.

## Frequency Assignment

Several features of IDPs complicate their NMR frequency assignment. First, they display low amide proton dispersion (7.5 to 8.5 ppm), resulting in crowded spectrum, which has been a bottleneck to frequency assignment in the past [9]. However, the availability of spectrometers of increasingly high magnetic field (up to 1 GHz), and the possibility to record high dimensional NMR spectra in reasonable time thanks to non-uniform sampling have led to a significant improvement of the capacity to resolve this crowding [10]. Alternatively, in the case of very large proteins, a “divide-and-conquer” approach is often used to reduce the complexity of the spectra [11].

Second, the amino acid composition found in IDPs, which can be highly repetitive and enriched in prolines, often cause difficulties for frequency assignment [12],13]. Prolines can be either in *cis* or *trans* conformation, and they interconvert on a slow timescale, leading to the existence of a second population of resonances in the spectrum [14]. To overcome this issue,  $H\alpha$  and carbon detection are increasingly used, allowing to find chain connectivities independently to the presence of prolines [15].

An especially difficult case for the assignment is the presence of Low-Complexity Regions (LCRs), which correspond to protein segments with a very high amino acid composition bias. Homorepeats, long stretches of the same amino acid, are specially eye-catching examples of LCRs and often linked to human diseases [16]. LCRs in general and homorepeats in particular represent especially difficult case for NMR assignment due to the similar chemical environment of these residues; the result is a more crowded spectrum that prevents conventional assignment. A recent study on a N-terminal region of the androgen receptor containing 25 consecutive glutamines demonstrated that in some circumstances the assignment of homorepeats can be achieved [17].

Finally, the exchange rate of amide protons with water molecules in IDPs is fast due to large solvent exposition of their backbone. Fortunately, IDPs are usually stable at a wide range of pH, which allows one to work at pH below 7, typically pH 6, to reduce the exchange rate [18]. Moreover, due to their high flexibility, IDPs give rise to narrow high intensity resonances. Lowering temperature can thus also be used to reduce the exchange whilst retaining a good quality spectrum.

A significant research effort has also been devoted in recent years to automate the assignment of IDPs, and several software packages have been developed to speed-up this tedious but unavoidable process [19]. One of the most striking recent achievements was the automated assignment of Tau protein (441 amino acids) using 5D and 7D spectra, GAPRO for cross-peaks picking, and MARS for automated assignment, resulting in 92 % of assigned residues within 5 days – including the time to record spectra [20].

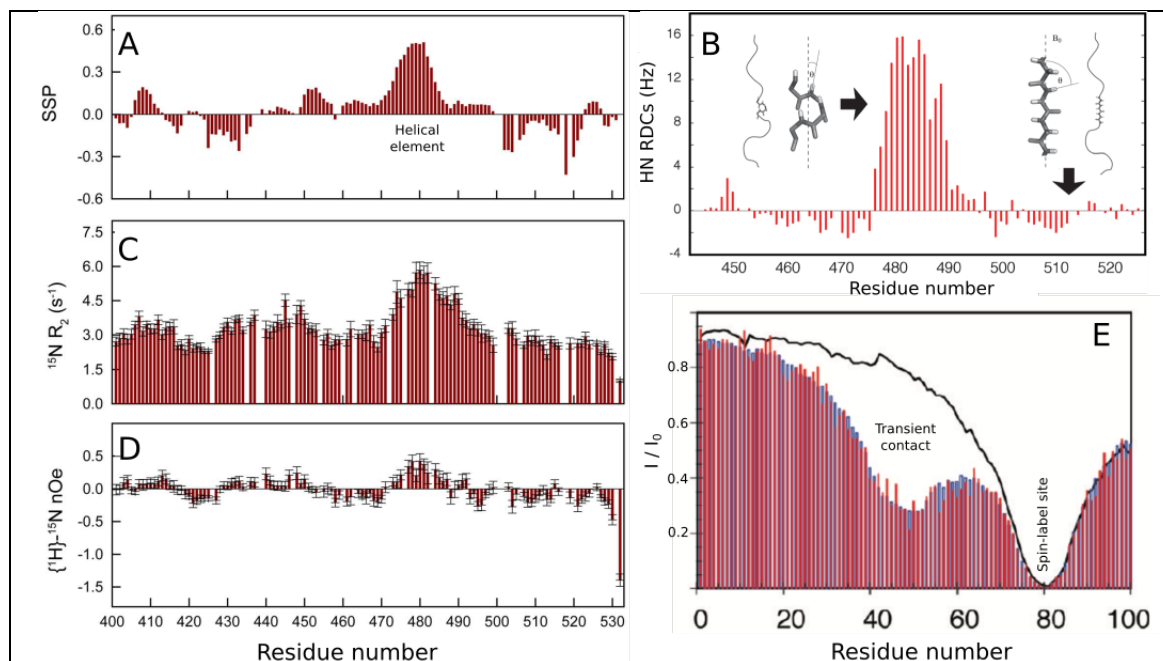
## NMR Observables

NMR is a rich source of structural and dynamic information for IDPs. Multiple NMR observables can be measured and they can probe local and global characteristics of disordered proteins. Importantly, NMR parameters correspond to time- and ensemble-averaged values due to the fact that the interconversion rate between different co-existing conformations is faster than the chemical shifts time-scale (ms). This averaging process must be considered when using these parameters to build structural models (see below). In this section, we succinctly describe the most relevant NMR parameters that are used for the structural and dynamic characterization of IDPs by NMR.

### Chemical shifts

The chemical shift (CS) is the most directly accessible parameter of a NMR spectrum, and given its sensitivity to the chemical environment of the nucleus, it contains information relative to the presence of secondary structure. Using the secondary chemical shift (SCS), *i.e.* the difference between the observed chemical shift and the chemical shift of the same amino acid in a *random coil* conformation, it is possible to determine the structural propensities within an IDP [21]. Indeed, the SCS of a carbon alpha (CA) or carbonyl (CO) will be positive in the case of helical propensity and negative in the case of  $\beta$ -strand or extended propensity, whereas a carbon beta (CB), nitrogen (N), aliphatic (HA), and amide (HN) proton chemical shifts will display the opposite behaviour. However, since IDPs are not – fully – structured, the SCS values obtained are small, and require the use of reliable databases for the *random coil* values. Several databases have been created to obtain these reference chemical shifts following different rationale. The earliest database was created using the chemical shifts of individual amino acids inserted in the middle of a glycine peptide (GGXGG) in 8 M Urea at pH 2, and using a correction for residues following prolines [22]. More recently, other databases were created, including correction factor for all amino-acids [23] and using QQQQQ peptides, with the rationale that glycines were not representative amino acids, using a lower denaturant concentration, and at neutral pH to better represent physiological conditions [24]. Following a different approach, Tamiola and Mulder created a *random coil* database (ncIDP) [25]. In their study, the CSs of 14 native IDPs were used to parameterize a database that uses both left- and right-neighbour correction factors, improving considerably the accuracy of chemical shift prediction for natively unfolded proteins. As the deviation from *random coil* behaviour can be subtle in the case of IDPs, one further step of analysis of SCSs was the addition of the contributions coming from CA and CB, which are insensitive to errors from chemical shift referencing [26]. For example, the SSP algorithm developed by Julie Forman-Kay's laboratory calculates the SCS using the RefDB *random coil* chemical shifts (which uses all CS from the BioMagResBank (BMRB)), and computes a secondary structure propensity (SSP) score using a smoothing function over five residues of the CA and CB SCS (Figure 1A) [27].

CSs are also a powerful tool to study PTMs, which often participate in the regulation of IDPs interactions. The most commonly studied PTM by NMR is phosphorylation, but it is also possible to follow glycosylation, acetylation, methylation, oxidation, propionylation... [28]. NMR allows not only to identify the presence of a PTM on a protein, to locate it at the sequence level, but also to follow its kinetics [6,29] and the structural changes it causes [8,30], which are key to understand the role of PTMs in the context of partner recognition [31].



**Fig. 1.** Illustration of the sensitivity of NMR parameters to the presence of residual secondary structure and long-range contacts in IDPs. **A.** The secondary structure propensity (SSP) calculated from experimental CA and CB chemical shifts using the software SSP is close to zero when no structure is present and increases in the helical element. **B.**  $^1D_{NH}$  residual dipolar couplings are negative in the disordered chain, where the bond vectors are on average perpendicular to the field; whereas in the helical region they are positive due to the tendency of the bond vector to be aligned with the field. **C.** The  $^{15}N$   $R_2$  relaxation rates also show an increase in the region of the helical element. **D.** The  $\{^1H\}$ - $^{15}N$  steady state heteronuclear Overhauser effects displays a value close to zero in the disordered chain and is more positive in the helical element. **E.** Paramagnetic relaxation enhancement intensity ratio: the presence of the probe affects the relaxation of residues close to it in the primary sequence, but also residues further apart, which are making transient contact with the region of the probe. The black line represents the data in the case where there are no transient contacts in the chain. Reproduced with permission from Communie G, et al. PLoS Pathog. 2013;9:e1003631 (A,C,D), reproduced with permission from Schneider R, et al. Mol. Biosyst. 2012;8:58–68 (B,E).

### Scalar Couplings ( $J$ )

Another parameter that contains backbone structural information is the scalar coupling. In particular, the relation between the  $^3J_{HNHA}$  and the dihedral torsion angle  $\Phi$  of the peptide bond is described by the Karplus equation [32]:  $^3J(\phi) = A \cos^2 \phi + B \cos \phi + C$ . Several parameterizations of this equation have been made over the years, the most recent one is  $A=7.97$  Hz,  $B=-1.26$  Hz, and  $C=0.63$  Hz [33]. Traditionally, the measurement of  $^3J_{HNHA}$  couplings in proteins can be obtained using a 3D  $^{15}N$ -separated quantitative J-correlation spectrum [34]. Recently, the accurate measurement of these couplings, directly from splittings in the  $^1H$  dimension of a WATERGATE-optimized  $^1H$ - $^{15}N$  TROSY- HSQC spectrum, have been reported for IDPs [35]. Secondary structure propensities are reflected in the measured  $^3J_{HNHA}$  coupling values. Residues in helices display couplings below 5 Hz, in *random coil* between 6 and 8 Hz, and in extended structures more than 8 Hz. An advantage of  $^3J_{HNHA}$  compared to CSs is that it is possible to distinguish type II polyproline helix (PPII) structures from  $\beta$ -strands. Indeed, residues in PPII have

a  $^3J_{\text{HNHA}}$  coupling value slightly higher than those of helices, between 5 and 6 Hz, due to the fact that PPII is a relatively open structure and has no intramolecular hydrogen bond [36]. As all the other NMR parameters,  $^3J_{\text{HNHA}}$  scalar couplings are time- and ensemble-averaged values, thus transiently populated secondary structures are difficult to identify. One solution could be, as for CSs, to subtract the value of a *random coil*  $^3J_{\text{HNHA}}$  coupling from the experimental data to locate deviations due to the presence of transient secondary structures. However, the only *random coil*  $^3J_{\text{HNHA}}$  databases that exist were either measured in the presence of a chemical denaturant (guanidinium chloride), or predicted from coil regions of crystallographic structures, and may thus not be representative of native IDPs [37,38].

### Residual Dipolar Couplings (RDCs)

The most sensitive and accurate parameters reporting on IDP conformation are residual dipolar couplings (RDCs) measured by partially aligning the studied protein in solution [39]. RDCs have been widely used to study the structure of multi-domain molecules but their use for IDPs has increased over the past few years and robust tools for predicting their values for both globular and disordered proteins have been developed [40–42]. In the case of IDPs, measured RDCs report on the orientation of the measured bond vector averaged over all existing conformers co-existing in solution. Although resulting RDCs are small due to the averaging process, they are measurable due to the inherent elongated structure of disordered chains [43]. However, secondary structure elements have a statistically averaged preferential orientation with respect to the magnetic field. For helices, the N-H bond vectors tend to be parallel to the magnetic field giving rise to more positive RDC values, whereas for extended structures, such as PPII or  $\beta$ -strands, the N-H bond vectors are on average orthogonal to the magnetic field yielding to more negative RDC values (Figure 1B) [44]. In addition to local conformational bias, RDCs also report on long-range intramolecular interactions by modulating the RDC profile along the chain [45]. However, decomposing both contributions is difficult. Many alignment media have been developed so far: PEG-alcohol, bicelles, stretched or compressed acrylamide gels, filamentous phages... [46] When measuring RDCs, one has to be extremely careful that the studied protein does not interact with or is not disturbed by the alignment medium: constant line width of the signal and no significant chemical shift differences with respect to a non-aligned sample must be observed. Indeed, strong electrostatic interaction can intervene decrease the tumbling rate of the protein concomitantly causing increased signal line-widths, as for example with filamentous phages, which are negatively charged and can thus interact with positively charged parts of a protein.

### Dynamics via Relaxation parameters

The combination of the longitudinal ( $R_1$ ), transverse ( $R_2$ ) relaxation rates and heteronuclear cross relaxation effect – or Nuclear Overhauser Effect (het-NOE) – of  $^{15}\text{N}$  nuclei has been extensively used to study protein dynamics in solution (Figure 1C,D) [47]. These rates report on motions faster than the correlation time of the protein (in the range of tenths of picoseconds to tenths of nanosecond). In the case of IDPs, the interpretation of relaxation rates is highly complex and is most often used in a qualitative manner applying the reduced spectral density mapping [48,49]. This approach translates the relaxation rates in the probability function of finding motions at specific angular frequencies,  $J(\omega)$ , thereby allowing the qualitative assessment of flexibility at residue level and thus permitting the localization of secondary structure formation segments. More recently, a specific mathematical framework has been proposed for the interpretation of relaxation rates in disordered proteins. It has been demonstrated in an extensive study of relaxation rates in Sendai virus' archetypal IDP,  $\text{N}_{\text{TAIL}}$ , showed that three dynamic modes are involved in IDPs dynamics: librational motions, local backbone sampling of the Ramachandran space, and segmental motions [50].

### Paramagnetic Relaxation Enhancement (PRE)

Long-range contacts influence the NMR parameters such as RDCs and need to be taken into account to get a more realistic structural description of IDPs [51]. By attaching a paramagnetic probe such as nitroxide spin label or a lanthanide to a specific residue on a protein, one can measure its influence on residues that are close – transiently or not – in space – up to 25-30 Å to the label [52]. Depending on the paramagnetic probe used, the effect on the surrounding residues is different. Some probes, such as lanthanides, induce a change of the chemical shift of nearby nuclei. These pseudo contact shifts (PCS) give information on the orientation between the observed nucleus and the paramagnetic center relative to the magnetic susceptibility tensor induced to the protein [53]. Other useful probes are stable free radicals, which increase the relaxation of nearby nuclei proportionally to the inverse sixth power of the distance to the probe (Figure 1E). This second type of probes is the most commonly used for IDPs as it only induces a line shape broadening of the signal corresponding to nuclei in the proximity of the paramagnetic center, due to a transverse paramagnetic relaxation enhancement (PRE). Common probes include MTSL spin label or 3-(2-iodoacetamido)-proxyl (PROXYL), which can be attached to a cysteine residue via a disulphide bond or a C-S bond, respectively [54,55]. The latter is often preferred due to its stability in presence of reducing agents that are needed, for example for interaction experiments with a cysteine containing protein. These probes can easily be reduced, commonly using ascorbic acid. Therefore, the same sample can be used for the measurement of both the paramagnetic and diamagnetic conditions, allowing simple analysis of the data using the intensity ratio  $I_{\text{para}}/I_{\text{dia}}$  extracted from  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, avoiding normalization issues and sample preparation imprecisions.

## From NMR Parameters to Conformational Ensembles

As described in previous sections, NMR represents an especially rich source of information to structurally and dynamically characterize IDPs. Importantly, through different observables, NMR offers information at the local (residue) level, but also on the overall size, and the presence of long-range interactions between remote positions of the chain [56,57]. However, NMR structural parameters are ensemble- and time-averaged observables, and therefore they have to be interpreted in a framework accounting for the coexistence of multiple conformations [58]. Although this feature is general for all kinds of biomolecules, IDPs represents an especially significant challenge due to their vast and flat conformational landscape. The number of degrees of freedom required to accurately describe disordered proteins is much larger than the amount of information that can be recorded experimentally. During the last decade multiple approaches have appeared to alleviate this intrinsic problem. The common feature of all these strategies is the application of sophisticated computational approaches that use previous knowledge on the chemical nature of peptide chains to reduce the number of degrees of freedom. The final aim is to have a description of the protein of interest that identifies its structural features and to link them with a specific biological function [59]. In the following paragraphs we will succinctly describe the different approaches used to derive structural ensembles of IDPs using NMR data.

Two general approaches have been developed in order to derive conformational ensembles from NMR data. The first one uses restrained Molecular Dynamics (MD) simulations by running in parallel multiple copies of the same molecule that collectively provide a good description of the experimental data used as restraints [60]. This approach combines an empirical energy force field with experimental information, and it is philosophically equivalent to the classical simulated annealing approaches for the structural determination of globular proteins by NMR. Restrained MD simulations have been applied to describe CSs [61], PREs [62,63], and RDCs [64] or their combination [65,66]. In these approaches, it remains unclear how the introduction of experimental restraints affects the exploration of the conformational space, which is key point for solving underdetermined problems. To ensure a more extensive sampling, some authors use Replica Exchange Molecular Dynamics (REMD), where conformations are exchanged between parallel

simulations according to a Monte Carlo criteria [67]. Another inherent limitation of the approach is the potential inaccuracy of present force fields to describe disordered protein conformational fluctuations and solvent behaviour. The improvement of force fields to adapt them to study IDPs is a very active field of research [68,69].

The second approach is based on the selection from a large ensemble of a subset of conformations that collectively describe the experimental data. This approach aims at enriching the ensemble with the conformations containing the structural features that are coded in the experimental data [57]. The generation of the initial pool is a key aspect of the approach, and it should explore as broadly as possible the conformational landscape of the protein. To achieve this aim, enhanced sampling approaches based on MD simulations have been used [70,71]. Alternatively, the conformational space can be sampled using simpler approaches based on statistical coil models derived from the backbone dihedral angle of loop regions of crystallographic structures [38,72–75]. In the absence of secondary structural elements or long-range contacts, Flexible-Meccano (FM), which is the most widely used statistical coil model, can describe conformational fluctuations at residue level probed by RDCs and the overall dimensions of the ensemble reported by SAXS [74]. Importantly, the presence of scarcely populated secondary structure or transient long-range contacts can be validated by comparing conformationally biased FM ensembles with experimental RDCs, CSs or PREs [14,45,76,77].

In IDPs, the need to simultaneously describe both local conformations and overall features has prompted strategies with the capacity to integrate multiple NMR observables in combination with other biophysical measurements. Among these approaches, the programs ENSEMBLE [78,79] and ASTEROIDS [80–82] are the two most popular ones. The combination of RDCs and PREs is especially useful as PREs probe long-range intramolecular contacts whereas RDCs probe both local and long-range features but both contributions are difficult to deconvolute [41]. When applying this combination to  $\alpha$ -synuclein, it was unambiguously proven that an interaction between the N- and C-termini of the protein [80]. Residue-specific NMR information can also be complemented with information reporting on the fluctuations of the overall properties of the ensemble such as SAXS [83,84] or smFRET [85].

It is widely recognized that a conformational ensemble that is in agreement with experimental data is not necessarily a proof of its accuracy [56,58]. The reliability of an ensemble is increased when multiple experimental parameters reporting with different sensitivity on similar features are integrated in the same minimization. Some strategies have been developed to assess the accuracy of data-driven ensembles. Cross-validation represents a very useful strategy to assess model accuracy, although it requires extensive experimental information. In cross-validation, a fraction of the experimental data is not used for the minimization and the capacity of the final model to describe these non-used data is evaluated. This strategy has been applied for the characterization of long-range interaction in Tau protein using 11 PRE datasets [86], and for the same protein and  $\alpha$ -synuclein using an extensive NMR dataset including CSs, RDCs, PREs and SAXS [87]. An alternative to cross validation is the use of Bayesian statistics [88]. In this approach, the relative population of all members of the initial ensemble and (importantly) their uncertainty are defined based on the experimental data [71,89]. The development of robust strategies to derive and validate ensemble models of IDPs is one of the future directions of the field. In that context, databases collecting experimental data and derived ensembles can be highly valuable [90].

## NMR Studies of Disordered Proteins Complexes

With their flexibility and inherent plasticity, IDPs play an important role in protein-protein interactions in many cellular processes, such as signalling and regulation of gene expression. Providing a comprehensive description of the biomolecular recognition processes for IDPs is thus of great importance for the understanding of key biological functions that are orchestrated by this family of proteins. Despite their relevance, detailed experimental evidence of binding mechanisms



employed by disordered proteins is still scarce. In this regard, NMR has been used to establish and characterize the vital role of protein disorder in the formation of biomolecular complexes. These highly flexible complexes are challenging or even impossible to study by crystallography. Although the study of these flexible complexes can be nicely complemented by other structural techniques, such as SAXS, the ability to build structural ensembles and to map dynamics of IDPs within complexes at the atomic scale makes NMR a unique tool.

Many IDPs present conformational bias at specific regions of the chain, which very often are associated to protein-protein interaction sites. An example of IDP that has been structurally characterized studied in its free state and in complex with its natural partner is the proliferating-cell-nuclear-antigen-associated factor (p15<sup>PAF</sup>), which is overexpressed in several types of human cancer. Although highly flexible, this protein populates a complex equilibrium of conformations in its free state, comprising partially structured states at specific segments of the protein. NMR relaxation and RDC analysis revealed that those segments coincide with functional binding sites, such as the PCNA interacting protein motif (PIP-box) and the KEN-box, which is recognized by the ubiquitin ligase to target p15<sup>PAF</sup> for degradation [14]. In complex with Proliferating Cell Nuclear Antigen (PCNA), p15<sup>PAF</sup> remains highly flexible with the exception of the PIP-box that binds with moderate affinity to PCNA adopting a conformation close to that found in its free state [91,92].

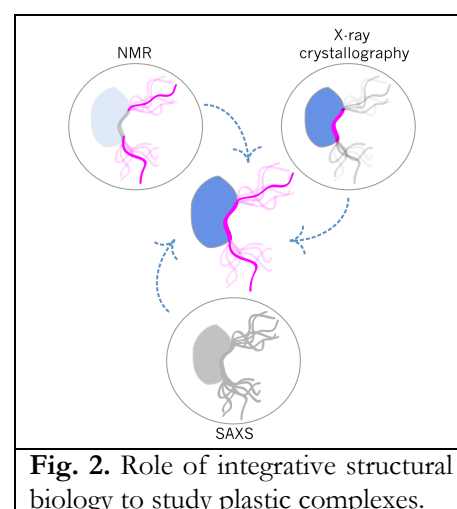
As in the case of p15<sup>PAF</sup>, many IDPs present pre-formed structures in the free state that resemble those of the bound state (“conformational selection”), which influence the binding affinity and the interaction kinetics. In other cases, the binding site undergoes a conformational transition only after association with the target (“induced fit”). These are two extreme cases, and generally both mechanisms co-exist and proteins present a disorder-to-order transition upon binding to their partner, referred to as the coupled folding and binding mechanism [93].

NMR relaxation dispersion probes protein dynamics in the microsecond to millisecond time range providing atomic resolution insights into coupled folding and binding processes of IDPs. This is exemplified in the of the folding upon binding process of the phosphorylated kinase-inducible activation domain (pKID) of the transcription factor CREB upon binding to its target, the KIX domain of the transcriptional co-activators CBP/p300 [7]. pKID is largely disordered in the free state and binds through a dynamic encounter complex, stabilized primarily by non-specific hydrophobic contacts, and ultimately folds into the fully bound state without dissociation from KIX. Unlike pKID, the disordered protein c-Myb binds KIX without formation of observable intermediates and its association strongly depends on the residual helical structure present in the N-terminal portion of c-Myb [94]. Akin to c-Myb, the unbound disordered domain of Sendai virus nucleoprotein (N<sub>TAIL</sub>) has a helix-forming propensity that affects its binding to the C-terminal domain of the phosphoprotein (PX). However, the interaction of N<sub>TAIL</sub> with its folded partner stabilizes a preformed helical state in an initial dynamic encounter complex docked to PX, previously to the formation of the final bound state [95]. In general, NMR and binding kinetics data show clearly that coupled folding and binding reactions of IDPs do not follow a single, general mechanism. Rather, binding occurs by a variety of mechanisms, depending on the inherent conformational propensities of the IDP, as well as the conformational dynamics of their partners.

The coupled folding-binding process might give the impression that IDPs are in fact proteins “waiting for a partner to fold” [96], and subsequently, to adopt its conformational unambiguous functional state. This might be true for some IDPs, but cannot be generalized for all of them. In this sense, NMR has allowed to put in evidence that the preservation of protein disorder by experiencing transient contacts in the bound state is an essential property of some functional IDP complexes. Similarly to fuzzy logics in mathematics, such flexible assemblies are termed as fuzzy complexes [97]. Disorder in fuzzy complexes can be local or global, and it is often tuneable by environmental or PTMs. Fuzziness in IDP complexes alleviates the entropic cost of binding and enables an ultra-sensitive response to signals, as well as the adaptability of disordered region to bind multiple partners.

The application of NMR to study fuzzy complexes of IDPs with local or global disorder is illustrated for G-actin bound to  $\beta$ -thymosin/WH2 domains and the signalling protein Sic1 interacting with Cdc4.  $\beta$ -thymosin/WH2 ( $\beta$ -t/WH2) domains are widespread short disordered regions (25-50 residues) able to recognize G-actin and regulate actin-assembly. When bound to G-actin, the N-terminal half of  $\beta$ -t/WH2 adopts a well-ordered amphipathic helix, whereas the C-terminal part remains highly dynamic, but tuneable by electrostatic interactions. NMR analyses showed that at low ionic strength  $\beta$ -t/WH2 sequesters G-actin in a polymerization incompetent state, where the dynamic interactions of the C-terminal part restrain it to a single main conformational and functional state [98]. NMR has also uncovered the example of a fuzzy complex while studying the interaction of cyclin-dependent kinase inhibitor Sic1 and the SCF subunit of Cdc4 [99]. Sic1 is a disordered protein comprising nine phosphorylation sites that can bind to Cdc4 in a phosphorylation dependent manner [100]. Interestingly, Cdc4 has only one high-affinity binding pocket for a phosphate group. NMR indicates that Sic1 remains largely disordered also when bound to Cdc4, forming a high affinity fuzzy complex only after six phosphorylations. The conformational plasticity of Sic1 allows the interchange between phosphorylation sites, which are in constant equilibrium with each other to occupy the single Cdc4 binding site. Along the order-disorder continuum, Sic1 remains on the disorder extreme of the spectrum even upon binding, exemplifying a clear case where structural disorder and multiplicity are used to perform a highly specialized function.

A successful structural and dynamic elucidation of highly plastic complexes requires the integration of information derived from complementary techniques. Indeed, since NMR is limited in the size of the particle under study, and by its sensitivity to heterogeneous chemical environments, structural characterisation of flexible complexes can become unmanageable by NMR. However, as illustrated in Figure 2, the disappearance of NMR signals for residues directly involved in the formation of the complex can be used to localize the interacting region, to then isolate and co-crystallize it in order to get the detailed crystallographic structure of the complex. On top of that, SAXS provides the overall envelope of the entire system, and can be used to integrate in a common model both the NMR and the crystallographic information.



**Fig. 2.** Role of integrative structural biology to study plastic complexes.

## Outlook

In this chapter, we have described how NMR has contributed to the conformational characterization of intrinsically disordered proteins and their complexes, and to decipher the structural bases of their biological function.

During the last two decades, tremendous efforts that have been undertaken to link structural features in disordered regions with the specific tasks performed *in vivo*. One of the most striking observations has been the evolutionary conservation of functional stretches in IDPs, which is in contrast to the non-functional parts that exhibit much lower sequence conservation due to the lack of structural or functional requirements [2]. In the vast majority of cases, these observations arise from bioinformatics studies, and therefore require experimental validation, which necessarily must come from NMR. It has been shown that NMR is the tool of choice to highlight the structural hot spots within disordered chains. Moreover NMR enables the localization of residues directly implicated in PTMs and recognition events. The identification of functional elements within an IDP is an essential information to guide the rational design of small molecules or peptide mimics able to modulate specific signalling cascades for pharmacological applications.

Along this chapter we have detailed the enormous technical and conceptual progress that NMR has experienced that nowadays enables the structural and dynamic characterization of IDPs and their complexes. Despite this progress there are multiple aspects that still require the attention of the NMR community. Probably the most urgent need is to develop robust and statistically meaningful methods to derive conformational ensembles from NMR data. Researchers still differently address the definition of the best combinations of experimental observables, and the integration of parameters with different information content or averaging properties. Even more fundamental is the assessment of the reliability of conformational ensembles, and how their accuracy depends on the experimental data used or the strategy applied. It will be also necessary to better define conformational transitions in IDPs. This dynamic phenomenon is poorly understood due to the vast and flat energy landscape explored by IDPs. Here, the development of computational approaches enabling to explore this landscape will be required. Reliable force fields and strategies to integrate relaxation rates into these simulations seem to be of fundamental relevance to understand how IDPs exchange from one conformation to another. The above-mentioned technical and conceptual improvements will be necessary to address the next biological challenges of the field. Low-complexity regions presenting high degeneracy of chemical shifts are an especially relevant one. To disentangle their spectroscopic complexity, either novel high-dimensional pulse sequences or tailored isotopic labelling schemes will have to be developed.

The formation of biomolecular complexes is on the origin of many biological functions of IDPs. Understanding the molecular bases of these interactions is therefore crucial but remains challenging in many instances, as they are fuzzy and/or only formed transiently. The combination of multiple relaxation rates probing a broad range of time-scales (from ps to ms) can deliver a kinetic perspective of these interactions. Very often IDPs can simultaneously interact with multiple partners, or similar regions of the same disordered chain compete for a single binding site. Subtle perturbations in these equilibria can drive to different signalling or metabolic pathways. By exploiting its atomic resolution, NMR is the most suited tool to study these complex and polydisperse scenarios.

IDPs were discovered two decades ago. Since then, our understanding of these proteins and their role in fundamental biological processes has enormously grown. However, we are convinced that there are still many structural and functional aspects of disordered proteins that remain unknown. NMR, as the sole structural biology technique able to probe these proteins at atomic level, has the unique opportunity to unveil new facets of this fascinating family of proteins.

## Acknowledgments

Financial support by the French Infrastructure for Integrated Structural Biology (FRISBI – ANR-10-INSB-05-01) and the Structural Biology RIO/IbiSA platform (Centre de Biochimie Structurale) is gratefully acknowledged.

Keywords (Intrinsically disordered proteins, conformational ensemble, disordered complexes, integrative methods)

## References

1. 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. 2004;635–45.

2. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem. Rev.* [Internet] 2014;114:6561–88. Available from: <http://dx.doi.org/10.1021/cr400514h>
3. Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.* 2015;35:49–59.
4. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 2005;272:5129–48.
5. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 2008;37:215–46.
6. Landrieu I, Lacosse L, Leroy A, Wieruszkeski J-M, Trivelli X, Sillen A, et al. NMR analysis of a Tau phosphorylation pattern. *J. Am. Chem. Soc.* [Internet] 2006;128:3575–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16536530>
7. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 2007;447:1021–5.
8. Sibille N, Huvent I, Fauquant C, Verdegem D, Amniai L, Leroy A, et al. Structural characterization by nuclear magnetic resonance of the impact of phosphorylation in the proline-rich region of the disordered Tau protein. *Proteins* 2012;80:454–62.
9. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 2004;104:3607–22.
10. Yao X, Becker S, Zweckstetter M. A six-dimensional alpha proton detection-based APSY experiment for backbone assignment of intrinsically disordered proteins. *J. Biomol. NMR* 2014;60:231–40.
11. Mark W-Y, Liao JCC, Lu Y, Ayed A, Laister R, Szymczyna B, et al. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.* 2005;345:275–87.
12. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 2000;41:415–27.
13. Mäntylähti S, Hellman M, Permi P. Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. *J. Biomol. NMR* [Internet] 2011 [cited 2012 Oct 12];49:99–109. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21259120>
14. De Biasio A, Ibáñez De Opakua A, Cordeiro TN, Villate M, Merino N, Sibille N, et al. P15PAF Is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophys. J.* 2014;106:865–74.
15. Bermel W, Bertini I, Felli IC, Lee Y-M, Luchinat C, Pierattelli R. Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. *J. Am. Chem. Soc.* 2006;128:3918–9.
16. Jorda J, Kajava A V. Protein homorepeats sequences, structures, evolution, and functions. *Adv. Protein Chem. Struct. Biol.* 2010;79:59–88.
17. Eftekhazadeh B, Piai A, Chiesa G, Mungianu D, Garcia J, Pierattelli R, et al. Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract. *Biophys. J.* 2016;110:2361–6.
18. Kim S, Wu K-P, Baum J. Fast hydrogen exchange affects <sup>15</sup>N relaxation measurements in intrinsically disordered proteins. *J. Biomol. NMR* 2013;55:249–56.
19. Baran MC, Huang YJ, Moseley HNB, Montelione GT. Automated Analysis of Protein NMR Assignments and Structures. *Chem. Rev.* 2004;104:3541–56.
20. Narayanan RL, Durr UHN, Bibow S, Biernat J, Mandelkow E, Zweckstetter M. Automatic assignment of the intrinsically disordered protein Tau with 441-residues. *J. Am. Chem. Soc.* 2010;132:11906–7.
21. Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy.

- Biochemistry 1992;31:1647–51.
22. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD.  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 1995;5:67–81.
  23. Schwarzing S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* [Internet] 2001;123:2970–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11457007>
  24. 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* [Internet] 2011 [cited 2012 Oct 12];50:157–65. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21604143>
  25. Tamiola K, Acar B, Mulder FAA. Sequence Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* 2010;132:18000–3.
  26. Zhang H, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* 2003;25:173–95.
  27. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci. A Publ. Protein Soc.* 2006;15:2795–804.
  28. Theillet F-X, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon M-K, et al. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR* 2012;54:217–36.
  29. Cordier F, Chaffotte A, Terrien E, Prehaud C, Theillet F-X, Delepierre M, et al. Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. *J. Am. Chem. Soc.* 2012;134:20533–43.
  30. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, Thomas PJ, et al. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* 2007;14:738–45.
  31. Okuda M, Nishimura Y. Real-time and simultaneous monitoring of the phosphorylation and enhanced interaction of p53 and XPC acidic domains with the TFIIF p62 subunit. *Oncogenesis* [Internet] 2015;4:e150. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4753521/>
  32. Karplus M. Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* 1963;85:2870–1.
  33. Vögeli B, Ying J, Grishaev A, Bax A. Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *J. Am. Chem. Soc.* 2007;129:9377–85.
  34. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in  $^{15}\text{N}$ -enriched proteins. *J. Am. Chem. Soc.* [Internet] 1993;115:7772–7. Available from: <http://dx.doi.org/10.1021/ja00070a024>
  35. Roche J, Ying J, Bax A. Accurate measurement of  $(^3\text{J})\text{J}(\text{HNH}\alpha)$  couplings in small or disordered proteins from WATERGATE-optimized TROSY spectra. *J. Biomol. NMR* 2016;64:1–7.
  36. Oh K-I, Lee K-K, Park E-K, Jung Y, Hwang G-S, Cho M. A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. *Proteins* 2012;80:977–90.
  37. Plaxco KW, Morton CJ, Grimshaw SB, Jones J a, Pitkeathly M, Campbell ID, et al. The effects of guanidine hydrochloride on the “random coil” conformations and NMR chemical shifts of the peptide series GGXGG. *J. Biomol. NMR* [Internet] 1997;10:221–30. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20700831>
  38. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native

- and random coil conformations. *J. Mol. Biol.* 1996;255:494–506.
39. Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 1997;278:1111–4.
  40. Prestegard JH, Bougault CM, Kishore AI. Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* 2004;104:3519–40.
  41. Huang J, Ozenne V, Jensen MR, Blackledge M. Direct prediction of NMR residual dipolar couplings from the primary sequence of unfolded proteins. *Angew. Chem. Int. Ed. Engl.* 2013;52:687–90.
  42. Kosol S, Contreras-Martos S, Cedeño C, Tompa P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* 2013;18:10802–28.
  43. Louhivuori M, Pääkkönen K, Fredriksson K, Permi P, Lounila J, Annala A. On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* [Internet] 2003;125:15647–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14664613>
  44. Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, Blackledge M. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* 2008;130:8055–61.
  45. Bernado P, Bertocini CW, Griesinger C, Zweckstetter M, Blackledge M. Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* 2005;127:17968–9.
  46. Prestegard JH, Kishore AI. Partial alignment of biomolecules: an aid to NMR characterization. *Curr. Opin. Chem. Biol.* [Internet] 2001;5:584–90. Available from: <http://www.sciencedirect.com/science/article/pii/S1367593100002477>
  47. Korzhnev DM, Billeter M, Arseniev a. S, Orekhov VY. NMR studies of Brownian tumbling and internal motions in proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* [Internet] 2001;38:197–266. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0079656500000285>
  48. Bracken C. NMR spin relaxation methods for characterization of disorder and folding in proteins. *J. Mol. Graph. Model.* 2001;19:3–12.
  49. Peng JW, Wagner G. Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry* [Internet] 1992;31:8571–86. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1390643>
  50. Abyzov A, Salvi N, Schneider R, Maurin D, Ruigrok RWH, Jensen MR, et al. Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J. Am. Chem. Soc.* 2016;138:6240–51.
  51. Uversky VN. Seven lessons from one IDP structural analysis. *Structure* [Internet] 2010 [cited 2016 Feb 26];18:1069–71. Available from: <http://www.sciencedirect.com/science/article/pii/S0969212610002753>
  52. 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.* [Internet] 1997;268:158–69. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9149149>
  53. Yagi H, Pilla KB, Maleckis A, Graham B, Huber T, Otting G. Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. *Struct. (London, Engl. 1993)* 2013;21:883–90.
  54. Berliner LJ, Grunwald J, Hankovszky HO, Hideg K. A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal. Biochem.* 1982;119:450–5.
  55. Ramos A, and Gabriele Varani\*. A New Method To Detect Long-Range Protein–RNA Contacts: NMR Detection of Electron–Proton Relaxation Induced by Nitroxide Spin-Labeled RNA. *J. Am. Chem. Soc.* [Internet] 1998;120:10992–3. Available from: <http://dx.doi.org/10.1021/ja982496e>

56. Jensen MR, Zweckstetter M, Huang J, Blackledge M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* 2014;114:6632–60.
57. Jensen MR, Ruigrok RWH, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* 2013;23:426–35.
58. Kruschel D, Zagrovic B. Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol. Biosyst.* 2009;5:1606–16.
59. Eliezer D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 2009;19:23–30.
60. Vendruscolo M. Determination of conformationally heterogeneous states of proteins. *Curr. Opin. Struct. Biol.* 2007;17:15–20.
61. Robustelli P, Stafford KA, Palmer AG 3rd. Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc.* 2012;134:6365–74.
62. Allison JR, Varnai P, Dobson CM, Vendruscolo M. Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* 2009;131:18314–26.
63. Ganguly D, Chen J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J. Mol. Biol.* 2009;390:467–77.
64. Esteban-Martin S, Fenwick RB, Salvatella X. Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings. *J. Am. Chem. Soc.* 2010;132:4626–32.
65. Huang J, Grzesiek S. Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. *J. Am. Chem. Soc.* 2010;132:694–705.
66. Wu K-P, Weinstock DS, Narayanan C, Levy RM, Baum J. Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J. Mol. Biol.* 2009;391:784–96.
67. Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* 2013;138:94112.
68. Henriques J, Cragnell C, Skepo M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* 2015;11:3420–31.
69. Mercadante D, Milles S, Fuertes G, Svergun DI, Lemke EA, Gräter F. Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *J. Phys. Chem. B* 2015;119:7975–84.
70. Mukrasch MD, Markwick P, Biernat J, Bergen M von, Bernado P, Griesinger C, et al. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* 2007;129:5235–43.
71. Fisher CK, Huang A, Stultz CM. Modeling intrinsically disordered proteins with bayesian statistics. *J. Am. Chem. Soc.* 2010;132:14919–27.
72. Fitzkee NC, Fleming PJ, Rose GD. The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins* 2005;58:852–4.
73. 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;102:13099–104.
74. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102:17002–7.
75. Ozenne V, Bauer F, Salmon L, Huang J-R, Jensen MR, Segard S, et al. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 2012;28:1463–70.

76. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105:5762–7.
77. Perez Y, Gairi M, Pons M, Bernado P. Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *J. Mol. Biol.* 2009;391:136–48.
78. Choy WY, Forman-Kay JD. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* 2001;308:1011–32.
79. Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 2013;29:398–9.
80. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, et al. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 2010;132:8407–18.
81. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* 2009;131:17908–18.
82. Ozenne V, Schneider R, Yao M, Huang J, Salmon L, Zweckstetter M, et al. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *J. Am. Chem. Soc.* 2012;134:15138–48.
83. Sibille N, Bernadó P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* [Internet] 2012;40:955–62. Available from: <http://www.biochemsoctrans.org/bst/040/0955/bst0400955.htm>
84. Cordeiro TN, Herranz-Trillo F, Urbanek A, Estana A, Cortes J, Sibille N, et al. Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr. Opin. Struct. Biol.* 2016;42:15–23.
85. Delaforge E, Milles S, Bouvignies G, Bouvier D, Boivin S, Salvi N, et al. Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin alpha. *J. Am. Chem. Soc.* 2015;137:15122–34.
86. Bibow S, Ozenne V, Biernat J, Blackledge M, Mandelkow E, Zweckstetter M. Structural impact of proline-directed pseudophosphorylation at AT8, AT100, and PHF1 epitopes on 441-residue tau. *J. Am. Chem. Soc.* 2011;133:15842–5.
87. Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, et al. Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and  $\alpha$ -Synuclein in Solution from {NMR} and Small Angle Scattering. *Structure* [Internet] 2014;22:238–49. Available from: <http://www.sciencedirect.com/science/article/pii/S096921261300436X>
88. Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 2011;21:426–31.
89. Ullman O, Fisher CK, Stultz CM. Explaining the structural plasticity of alpha-synuclein. *J. Am. Chem. Soc.* 2011;133:19536–46.
90. Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, et al. pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.* 2014;42:D326-35.
91. De Biasio A, de Opakua AI, Mortuza GB, Molina R, Cordeiro TN, Castillo F, et al. Structure of p15(PAF)-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nat. Commun.* 2015;6:6439.
92. Cordeiro TN, Chen P-C, De Biasio A, Sibille N, Blanco FJ, Hub JS, et al. Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. *Nucleic Acids Res.* 2016;
93. Wright PE, Dyson HJ. Linking folding and binding. *Curr. Opin. Struct. Biol.* 2009;19:31–8.
94. Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically



- disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U. S. A.* 2015;112:9614–9.
95. Schneider R, Maurin D, Communie G, Kragelj J, Hansen DF, Ruigrok RWH, et al. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *J. Am. Chem. Soc.* 2015;137:1220–9.
  96. Janin J, Sternberg MJE. Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol. Rep.* 2013;5:2.
  97. Miskei M, Antal C, Fuxreiter M. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.* 2016;
  98. Didry D, Cantrelle F-X, Husson C, Roblin P, Moorthy AME, Perez J, et al. How a single residue in individual beta-thymosin/WH2 domains controls their functions in actin assembly. *EMBO J.* 2012;31:1000–13.
  99. Mittag T, Orlicky S, Choy W-Y, Tang X, Lin H, Sicheri F, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105:17772–7.
  100. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, et al. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 2001;414:514–21.