Intrinsically Disordered Proteins: Methods for Structure and Dynamics Studies

Scott A. Showalter

The Pennsylvania State University, State College, PA, USA

Intrinsically disordered proteins (IDPs) partially or completely lack a cooperatively folded structure under native conditions, preventing their equilibrium state from being adequately described by a single structural model. Reaching the goal of quantitatively describing IDP structure–function relationships has required the development of novel experiments aiming to characterize their structure and dynamics. This article will cover the state of the art in defining IDP structures, with particular emphasis on the role NMR has to play in resolving this grand challenge. In addition, the article will close with a survey of contemporary computational approaches to utilizing NMR structural constraints for IDP model generation. As a case study aiming to motivate the points made, examples will be provided from our investigation of the intrinsically disordered C-terminal region of FCP1 and its interaction with the cooperatively folded winged-helix domain from Rap74.

Keywords: NMR resonance assignment, residual dipolar couplings, paramagnetic relaxation enhancement, spin relaxation, carbon detection, ensemble model, structure determination, molecular dynamics, IDP, FCP1

How to cite this article:
eMagRes, 2014, Vol 3: 181–190. DOI 10.1002/9780470034590.emrstm1360

Introduction

Intrinsically disordered proteins (IDPs) partially or completely lack a cooperatively folded structure under native conditions, preventing their equilibrium state from being adequately described by a single structural model. At least superficially, these proteins challenge traditionally narrow formulations of the structure–function paradigm: IDPs lack spatially and temporally stable structure and yet retain significant biological functions. Our view is that, like cooperatively folding domains, IDPs do possess native structure that is responsible for imparting their specific functions; understanding the structure–function relationships of IDPs simply requires broadening our traditionally narrow concept of the biologically functional native state. Reaching the goal of quantitatively describing IDP structure–function relationships has required the development of novel experiments aiming to characterize their structure and dynamics, because the nature of IDP structure is different from the nature of the classical targets for structural biology. This article will cover the state of the art in defining IDP structures, with particular emphasis on the role NMR has to play in resolving this grand challenge. Unsurprisingly, the nature of the conformational space adopted by IDPs is hotly debated. Resolving this debate requires the development of quantitative and comprehensive tools for generating IDP ensembles and for validating their biological relevance. Thus, the article will close with a survey of contemporary computational approaches to utilizing NMR structural constraints for IDP model generation. As a case study aiming to motivate the points made, examples will be provided from our investigation of the intrinsically disordered C-terminal region of FCP1 and its interaction with the cooperatively folded winged-helix domain from Rap74.

IDPs are Structurally and Functionally Distinct from Cooperatively Folding Proteins

It is now generally accepted that disordered 3-D structure is as well suited to biological function as orderly (i.e., cooperatively folded) structure, but this was not always so. Over the course of the 1990s, numerous examples of biologically functional IDPs began to emerge, although their study was far from systematic. In 1999, Wright and Dyson[1] recognized the prevalence of intrinsically unstructured proteins (as they were named at the time) and argued the perspective that protein function derives from native structure, regardless of whether or not folding occurs. Thus, it has been established that even in the absence of spatial and temporal order, native structure must always be considered significant.

Early computational studies of intrinsic disorder revealed the chemical properties inherent to disordered sequences that have rendered these proteins especially challenging to characterize biophysically and structurally. Long 'medium complexity' sequence segments frequently occur in proteins, possessing compositional properties that are distinct from those of known globular proteins.[2] Bioinformatic studies have clarified that the 'medium complexity' sequences we now recognize as IDPs are relatively enriched in polar residues and proline and depleted in hydrophobic residues.[3] As a result, IDPs lack the hydrophobic core that is necessary to drive protein folding. Although some

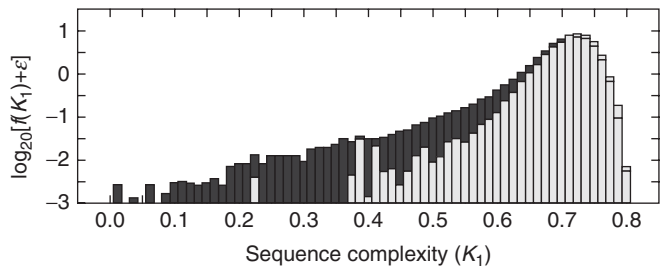

Figure 1. Coverage of protein sequences in the form of deposited structures in the protein data bank is poor for sequences with lower complexity. The parameter K1 measures the information content inherent to the given protein's amino acid sequence, as defined by Wooton. IDPs tend to have complexities of 0.5 or less, whereas the peak near 0.7 is composed largely of enzymes and other cooperatively folding proteins

recent progress has been made, regions of low-to-medium sequence complexity remain systematically underrepresented in structure databases [e.g., the protein data bank (PDB)], as demonstrated in Figure 1. While we now understand that regions of dynamic structure in proteins are common, we still have only rudimentary knowledge of the structures adopted in the natively disordered state.

One might fairly question why the structural features of the natively disordered state are worth attempting to quantify, but there are two compelling reasons for doing so. First, while our ability to infer structural features of ordered protein domains from their sequences is improving, significant progress is still needed before de novo structure prediction from sequence alone will become general and robust. Perhaps detailed investigation of intrinsically disordered structure, and an appreciation for the sequence-encoded characteristics giving rise to it, will also advance the field of protein folding. Second, intrinsic disorder is vital for fundamental cellular functions, including transcription, translation, and cellular signal transduction.[4] In performing these functions, IDPs nearly always form complexes with other macromolecules, necessitating the determination not only of their unbound-state structures, but also of the conformations they adopt in complexes.

Many recent studies have investigated the remarkable disorder-to-order transition that couples IDP folding to ligand binding.[5] Importantly, while IDPs are generally depleted in hydrophobic residues, those which are present disproportionately participate in forming protein–protein interfaces upon complex formation.[6] Thus, molecular recognition fragments (or MoRF sequences) in disordered proteins are readily identified and represent a fertile sequence space for productive enumeration of structure–function relationships.[7] Most IDPs are found to adopt one or more temporally stable and highly populated structures in complex with binding partners, suggesting that conformational adaptation or selection may be a critical aspect of IDP function, in analogy to the observed importance of dynamic structural rearrangements for enzymology. As such, there is significant interest in documenting the binding mechanism for IDPs and addressing whether they fold before, during, or after the formation of an initial complex with their binding partners.

Characterizing IDP Structure

How to best describe the conformational features of the unfolded state, whether intrinsic to the polypeptide's sequence or induced by denaturants, remains an open question. Current methods for protein structure determination and models for interaction mechanisms are primarily formulated under the assumption of a temporally stable and folded protein state. Motivated by the biological significance of the natively disordered state, rapid progress is being made toward developing methods for determining IDP structural ensembles. While it is self-evident that crystallographic methods are unlikely to yield biologically relevant structures of IDPs (at least in their unbound state), small angle X-ray scattering (SAXS) has been established as a powerful tool to describe the macroscopic features of IDP structure ensembles and, in some cases, even to define local structure.[8] Although recent advances in SAXS applications to IDPs are exciting in their own right, their discussion lies outside the scope of this article. Finally, although it will not be discussed here because of space constraints, solid-state NMR spectroscopy of vitrified IDP samples may play a vital role in our field, because it can readily yield constraints on the backbone φ and ψ torsion angles.[9]

Solution NMR spectroscopy is a uniquely powerful tool for studying the dynamic structural ensembles of IDPs.[10,11] Two remarkable structural models of the intrinsically disordered segments in p53 and Sic-1 have recently been solved, based primarily on solution NMR constraints.[12,13] Despite the enormous potential seen in these examples, general study of IDPs by NMR has been limited by the extremely poor ^{1}H-amide chemical shift dispersion typically observed in their spectra and the traditional reliance on this nucleus for detection in biomolecular NMR experiments. Where special cases with good spectral quality have been found, as illustrated by the p53 and Sic-1 examples cited in the preceding text, progress has been made. However, given the biological imperative to better understand IDP structure–function relationships, more systematically successful methods for NMR applications to IDPs are needed. Recently, ^{13}C direct-detected NMR spectroscopy has emerged as the tool of choice for achieving more systematic success when screening IDP systems for high spectral quality. Owing to a recent burst of effort across multiple laboratories, all of the NMR methods discussed in this article have been implemented in both ^{1}H-detected and ^{13}C-detected formats,

leaving the spectroscopist at liberty to choose whichever approach appears to be the most efficient for a given system of interest.

NMR Spectroscopy of IDPs

Solution NMR spectroscopy has great potential both to qualitatively define IDP conformational ensembles and to go even further: to determine their site-specific interactions with other proteins or nucleic acids. Owing to the lack of a tightly packed hydrophobic core, IDP structures appear to be determined mainly by torsional and other constraints on their polypeptide backbones. Among the sparse structure constraints available from disordered ensembles, the chemical shift and residual dipolar couplings (RDCs) have proven to be the most generally valuable, as has the measurement of paramagnetic relaxation enhancement from extrinsic paramagnetic probes.[14] In addition, the measurement of backbone dynamics by NMR spin relaxation is emerging as an integral component of IDP ensemble generation and refinement, owing to the more dynamic nature of the disordered state. Application of each of these techniques for the constraint of IDP ensemble models will be discussed in this section, but first methods for generating a well-resolved 2-D correlation spectrum, onto which the observations can be encoded, must be established.

The Advantages of ^{13}C-Detected NMR Methods

Two-dimensional heteronuclear correlation experiments have become a mainstay of biomolecular NMR spectroscopy because they generally provide good peak dispersion and are easily recorded in phase-sensitive, absorption-mode formats. In particular, the ^{1}H,^{15}N-heteronuclear single quantum coherence (HSQC) has become a workhorse experiment for protein NMR, because it provides nominally one-to-one mapping between resonances and residues (the absence of proline being the major exception) and can be rapidly acquired with inexpensive samples. Furthermore, the ^{1}H,^{15}N-HSQC serves readily as a component for building multidimensional experiments that yield chemical shifts and structure constraints, and as a detection platform for the measurement of spin relaxation, scalar couplings, or dipolar couplings. For most small and cooperatively folding proteins, the ^{1}H,^{15}N-HSQC features an extremely well-dispersed set of resonances, as shown for human ubiquitin in Figure 2(a). In contrast, nearly all IDPs yield poorly dispersed spectra, particularly in the $^{1}H^{N}$ dimension, characterized by extreme spectral crowding, as demonstrated for the

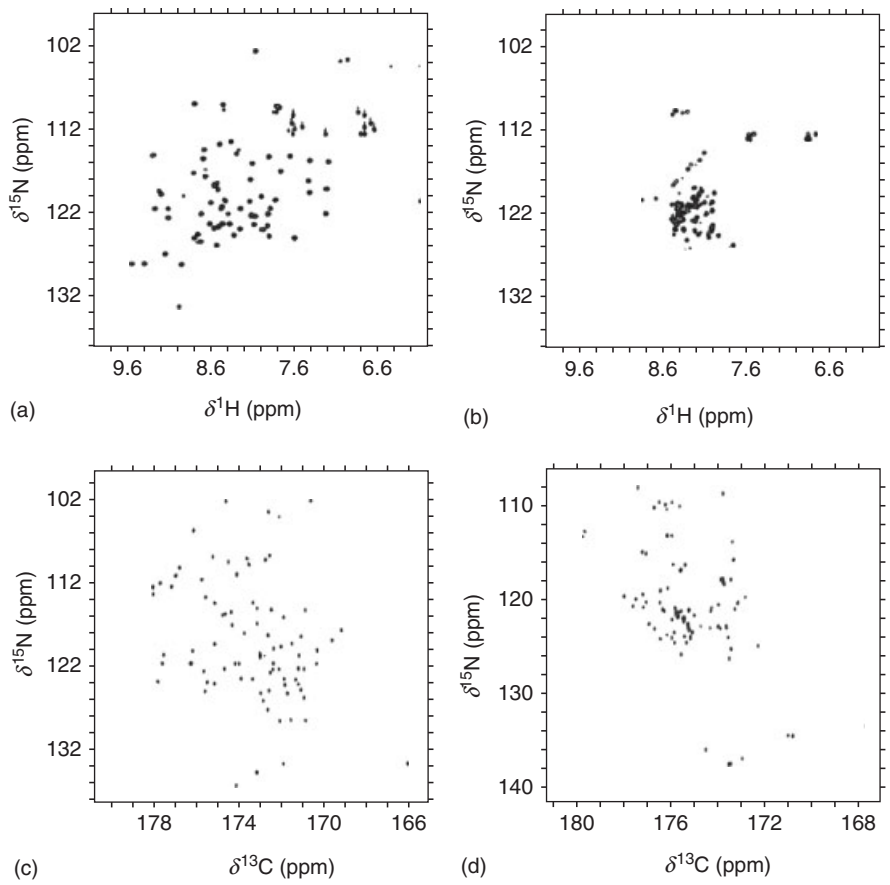

Figure 2. 2-D-heteronuclear correlation spectra provide the foundation for all biomolecular NMR studies, but traditional ^{1}H-detected methods do not work as well for IDPs as newer ^{13}C-detected methods. (a) The ^{1}H,^{15}N-HSQC of the cooperatively folded ubiquitin shows excellent peak dispersion, whereas (b) ^{1}H,^{15}N-HSQC of the intrinsically disordered FCP1 C-terminal tail does not. (c) The ^{15}N,^{13}C-CON of ubiquitin and (d) the ^{15}N,^{13}C-CON of FCP1 both show excellent peak dispersion

C-terminal domain of human FCP1 in Figure 2(b). While some IDPs yielding well-dispersed ^{1}H,^{15}N-HSQC spectra have been found, to truly advance our understanding of IDP function requires that a new platform which is generally likely to yield high-quality spectra be identified.

The principal advantage to choosing the ^{1}H,^{15}N-HSQC as a preferred detection platform for protein NMR is that it provides a simple spectrum with one 2-D-resonance per amino acid residue, except for proline. Similar mapping is achievable using, e.g., the $^{13}C'-^{15}N$ or $^{13}C'-^{13}C\alpha$ as a reporting pair, but choosing either of these pairs comes with consequences, most notably the fact that ^{1}H is a far more sensitive detection nucleus than ^{13}C. Recent advances in cryogenic probe technology, particularly the incorporation of cryogenically cooled carbon coils, have resulted in the reemergence of ^{13}C direct-detection spectroscopy as a tool for studying proteins in solution.[15] Carbon-13 detection has been incorporated into multiple 2-D-spectral formats and we have found the ^{15}N,^{13}C-CON spectrum, correlating the ^{13}C-carbonyl with the ^{15}N-amide of each peptide plane, to be an especially effective choice for studying IDPs.[16–19] This is illustrated through comparison of the ubiquitin ^{15}N,^{13}C-CON (Figure 2(c)) and the FCP1 ^{15}N,^{13}C-CON (Figure 2(d)), and the contrast with the respective ^{1}H,^{15}N-HSQC spectra of proteins in the same figure.[17] In addition to providing significantly enhanced chemical shift dispersion for FCP1, as compared to the ^{1}H-detected format, the ^{15}N,^{13}C-CON spectrum yields sharper line widths (at a given magnetic field strength), further enhancing the practical resolution of the experiment. Perhaps even more significantly for IDPs, which tend to be enriched in proline compared to cooperatively folding proteins, the ^{15}N,^{13}C-CON spectrum contains a resonance for each peptide bond involving a proline ^{15}N (easily identified by the downfield shift in the ^{15}N dimension) contrasting with the ^{1}H,^{15}N-HSQC, which does not.

Although current technology does limit the application of ^{13}C-detected spectroscopy to spectrometers equipped with cryogenic probes, excellent spectral quality is typically achievable on spectrometers operating at 11.7 or 14.1 T magnetic field strength, thus putting the technique within reach for a large number of NMR facilities. Of additional practical importance, acquisition of 2-D and 3-D spectra with ^{13}C-detection often requires twice the number of spectrometer hours as a ^{1}H-detected experiment acquired on the same instrument, but nonuniform sampling techniques and recent pulse program improvements offer attractive solutions to this limitation. New NMR techniques often require new practitioners to have sophistication with pulse programming, but example pulse programs for all measurements discussed in this article have been published in ^{13}C-detected formats and most pulse sequences are available in the standard libraries offered by the spectrometer manufacturer.

Structural Observables

Cooperatively folded protein domains typically yield large numbers of long-range NOEs that are used as the principal constraints in the determination of their structures; in contrast, IDPs typically yield very few nonlocal NOEs with which their structures can be defined. As a result, deviations from random-coil chemical shift and alternative geometric constraints, such as RDC measurements, have become critical data points from which IDP conformational ensembles are constrained.

Chemical Shift. The structural information encoded in the chemical shift itself has emerged as a key indicator of IDP ensemble properties.[20] The chemical shift is fundamentally determined by the local geometry and chemical environment of the nuclear spin. IDPs tend to be depleted in long-range tertiary contacts and so local structural constraints on the backbone are potent for defining the ensemble. It has been argued that inaccuracies can arise in ensembles constrained by population-weighted measures, such as the chemical shift. However, the length scale constrained by the chemical shift is likely to be well within the thermal blob length of the chain, where lower amplitude and more local conformational fluctuations should dominate, causing inaccuracies induced by the mean polymeric behavior of the chain as a whole to be minimal.[21]

Given that chemical shift can be used to constrain structure, it is important to develop best practices for inclusion of chemical shift in model building and for representing the data graphically, such that they may better assist investigators in understanding the properties of IDPs. The isotropic chemical shift of many backbone nuclei has been shown to depend explicitly on the presence of secondary structure (e.g., α-helix, PPII-helix, and β-strand), which causes deviation of the measured chemical shift from the expectation value for a random-coil reference state (also known as the secondary shift). Importantly, secondary-shift information from a single nucleus can be misleading, particularly where the population of secondary structure in the ensemble is small. In addition, the problem of identifying the 'correct' set of random-coil reference chemical shifts for IDPs has not been resolved to the complete satisfaction of the community. For these reasons, consensus measurements derived from multiple chemical shifts are highly recommended for IDP applications.

If regions of strong secondary chemical shift are identified in an IDP, it is clear that they will be a valuable source for structural constraints. In all cases, back-calculation of chemical shifts from models for assessment purposes is desirable. Such calculations require numerical evaluation of chemical shifts from structures, because the correlation between protein structure and chemical shift is not obvious, especially for IDPs. Whether for assessment purposes or for the generation of refinement inputs, computational methods designed to convert chemical shifts into secondary-structure propensities (SSP) are required. For assessment purposes, we find the SSP method to be particularly effective. SSP explicitly accounts for the differential sensitivity of the individual amino acid types to secondary structure and provides output in the form of a continuum measurement.[22] For the generation of heuristic inputs used in the preliminary stages of model building, we favor the δ2D method, which translates chemical shift lists into probabilities that each residue in the chain will sample secondary structure.[23] Both of these methods also offer the advantage of presenting the output data in an intuitive graphical format.

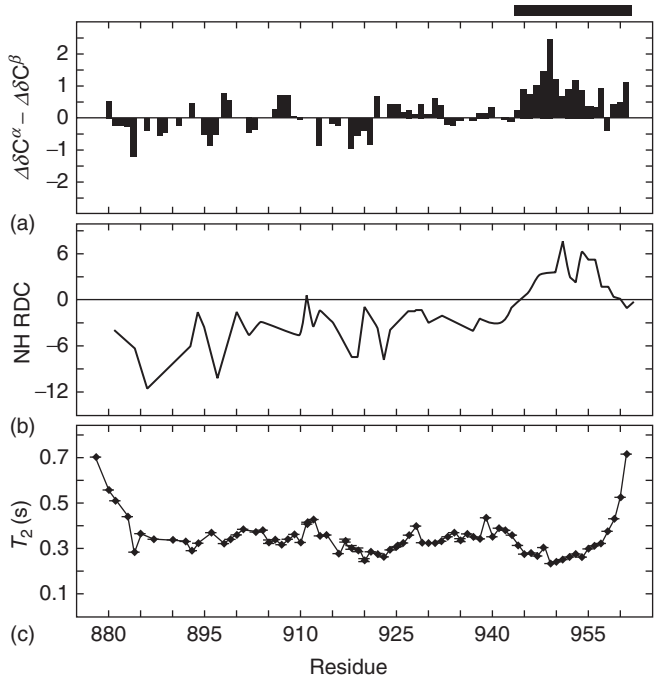

Figure 3. NMR parameters used to constrain the structure of FCP1. (a) Backbone chemical shift is represented as the difference between secondary shifts of C^α and C^β, which will be >2.0 ppm for α-helix and <−2.0 ppm for β-strand. (b) NH-RDC values measured in compressed polyacrylamide gels, which will be small and negative for coil conformations and both positive and larger in absolute magnitude for α-helix. (c) ^{15}N-T_2 relaxation times, which tend to be shorter for regions of high contact density. The bar above the figure indicates the region of FCP1 that becomes a stable α-helix in complex with Rap74

Finally, structure ensembles for IDPs are often difficult to parse, making presentation of the raw secondary chemical shift data valuable for developing the investigator's intuition for the nature of a given IDP's structure. We find the most effective visual representation of secondary chemical shift results to be plotting $\Delta\bar{\delta}C^\alpha - \Delta\bar{\delta}C^\beta$ (where Δ implies the deviation of the measured shift from its random-coil value, e.g., $\Delta\bar{\delta}C_i = \bar{\delta}C_{i,measured} - \bar{\delta}C_{i,coil}$). In general, methods featuring subtraction of a pair of chemical shifts smooth out potential errors caused by choosing a nonoptimal random-coil shift set. As an example, we present $\Delta\bar{\delta}C^\alpha - \Delta\bar{\delta}C^\beta$ results for FCP1 in Figure 3(a). In this figure, a fully formed α-helix would yield a region of strongly positive deviation from zero. The Rap74 binding region of FCP1, which is strongly helical in the bound state, is identified by the bar above the figure. It is clear that the secondary chemical shifts are consistent with a bias toward α-helical structure in this region of the IDP, whereas the majority of the chain shows no strong or systematic deviation from random-coil shifts.[18]

Residual Dipolar Couplings. The chemical shifts of backbone nuclei are powerful constraints on local structure, but are generally not sufficient to fully constrain the ensemble, because they are local in nature. Measurement of RDCs is an effective means of complementing chemical shift information, because RDCs are sensitive to long-range interactions. The theory and practice of RDC measurement have been extensively reviewed elsewhere.[24] What is most important here is that the strength and the sign of the dipolar coupling between two nuclear spins place a constraint on the orientation of their internuclear vector, with respect to a global or segmental alignment tensor. This is illustrated well by the partially formed α-helix at the C-terminus of FCP1, which is clearly evident in its NH-RDC profile (Figure 3(b)). The NH-RDC is particularly useful for identifying α-helix, because it changes sign in the helical context, compared with a random/extended conformation.

The dipolar coupling is sensitive to dynamic averaging, up to the millisecond timescale, and therefore is subject to limitations when employed as a structural constraint for highly flexible systems. Specifically, multiple degenerate solutions for the orientation of each internuclear vector exist, leading to the possibility of incorrect constraints being placed on the chain if the data set is large or the chain length is long. Individual chain conformations are sure to produce poor back-calculations of the ensemble averaged RDC values. Thus, use of RDCs as structure constraints is dependent on appropriate conformational averaging when the ensemble is numerically evaluated and to a lesser extent on the number of chains present in the computed ensemble. Further complicating the problem of using RDCs as structure constraints, it is not guaranteed that the entire IDP chain will be constrained by a unique and global alignment tensor. Solutions to this problem involving segmental or local alignment tensor approaches have been discussed and are used by default in most common software packages for ensemble refinement.[25] Even if the physical rationale for this choice is debatable, the procedure appears to remove most artifacts and is therefore useful.

Observing IDP Dynamics

Conformational dynamics are an integral component of the ensemble description of IDPs. Because the backbone of IDPs nearly always experiences significant conformational relaxation in solution, the assessment of backbone spin relaxation represents an especially potent means of improving the description of IDP structure ensembles. In addition, intrinsically disordered proteins are very nearly required to undergo a change in backbone conformation upon complex formation, even in the most dynamic complexes, making the assessment of backbone spin relaxation imperative.

Paramagnetic Relaxation Enhancement. Intentionally, studying proteins that harbor paramagnetic metal centers has long been a challenge, because the paramagnetic relaxation enhancement (PRE) induces deleterious increases in the linewidth of nearby nuclear spins. On the other hand, the deliberate incorporation of an exogenous paramagnetic probe for the purpose of introducing line broadening at strategically chosen locations in a protein's structure has proven to be a highly effective means of generating long-range structure constraints. Especially for IDPs, control experiments must be performed to confirm that incorporating the probe has not perturbed the structure of the protein. The relaxation enhancement caused by proximity to a paramagnetic center can extend up to 20 Å or more, under favorable conditions. Thus, PRE constraints are an attractive complement to those provided by chemical shifts, or what NOEs can be measured, for an IDP.

While incorporating PRE constraints is relatively straightforward for cooperatively folding domains, backbone and probe dynamics both heavily influence the intensity of induced effects, causing complications for IDP applications. Transient contacts that are otherwise difficult to establish can yield strong PRE signatures that synergize well with RDC measurements for structure constraint.[14] Conversely, as with RDC measurements, the presence of multiple transient contacts can yield incompatible constraints, or leave minor states undetected, and so caution must be exercised when evaluating ensembles constrained by PRE. Owing to the convolution of effects from dynamics and low population conformational states, interpretation of PRE data is often improved through concurrent analysis of backbone spin relaxation, which helps to define the amplitude of local conformational dynamics in the ensemble.

NMR Spin Relaxation. IDP structural ensembles feature both rapid local conformational fluctuations and the potential for slower relaxation between multiple conformations on what might be a glassy potential energy surface. To varying extents, all of the structural constraints discussed earlier are influenced by the presence of conformational dynamics, and so their utilization can be improved by identifying those sites subject to enhanced local dynamics through independent measurements. Spin relaxation of the ^{15}N-nucleus of backbone amides is an especially common probe to use for studies of fast (ps–ns) timescale conformational dynamics because of its relatively simple spin physics: relaxation is dominated by the dipolar interaction with the attached proton at most magnetic field strengths and, so long as samples are prepared with natural carbon isotope abundance, the amide nitrogen and its covalently attached hydrogen constitute an isolated spin-1/2 pair. In the presence of uniform ^{13}C-enrichment, as is needed for carbon-detected spectroscopy, special precautions are needed to minimize the influence of the additional spins-1/2 on measured relaxation times.[19] For both cooperatively folding and intrinsically disordered proteins, determination of longitudinal relaxation times (T_1), transverse relaxation times (T_2), and heteronuclear Overhauser effects(^{15}N,^{1}H-NOE)provides a comprehensive determination of ps–ns conformational dynamics. While extraordinarily valuable for IDPs, measurement of ^{15}N,^{1}H-NOE can be complicated because values yielding near-zero peak intensity in the 'saturation-on' experiment are typical.

In both tightly packed and disordered systems, local contact density tends to correlate strongly with measured relaxation times. The T_2 relaxation time is particularly dependent on contact density, and so decreased transverse relaxation times compared to baseline have been used to establish the presence of tertiary structure in IDP ensembles. As with most constraints on IDP structure, caution must be exercised in making such assignments, because α-helical structure also increases contact density and therefore has the same effect on measured T_2 times as the presence of tertiary structure.[26] Of course, the influence of helical structure can be independently verified through assessment of chemical shifts, thus demonstrating the utility of applying a variety of constraints in solving IDP ensembles. The influence of α-helical structure on T_2 can easily be seen in Figure 3(c), where the partially helical C-terminal end of FCP1 is seen to possess both the chemical shift signature of a helix and a depressed set of T_2 values, relative to baseline.[19] Conversely, regions of enhanced dynamics are seen in the N-terminal half of the protein (which is glycine rich) in the absence of strong secondary chemical shifts, suggestive of enhanced local dynamics in this region.

Ensemble Refinement and Computational Approaches

Spectroscopic data suggest that IDPs may adopt anything from a heterogeneous, but relatively compact, ensemble of structures to a denatured state highly enriched in native-like secondary structural features, reminiscent of molten globules.[27] Regardless of the details, the highly dynamic nature of IDPs requires a description of their native state that goes beyond the static structures associated with cooperatively folding proteins. Physically, IDPs are best described as accessing ensemble states where the protein is able to interconvert, on some timescale(s), between multiple conformations on a rugged potential energy landscape. Therefore, the IDP community has embraced the idea of publishing diverse sets of structures that collectively span the conformational space thermally accessible to the IDP ensemble. For example, the structural constraints provided in Figure 3 were used in conjunction with the ENSEMBLE software package (see section titled 'TRaDES and ENSEMBLE') to generate the ensemble structure model of FCP1 displayed in Figure 4. Describing the ensemble state of an IDP, such as FCP1, is more about defining which properties govern the

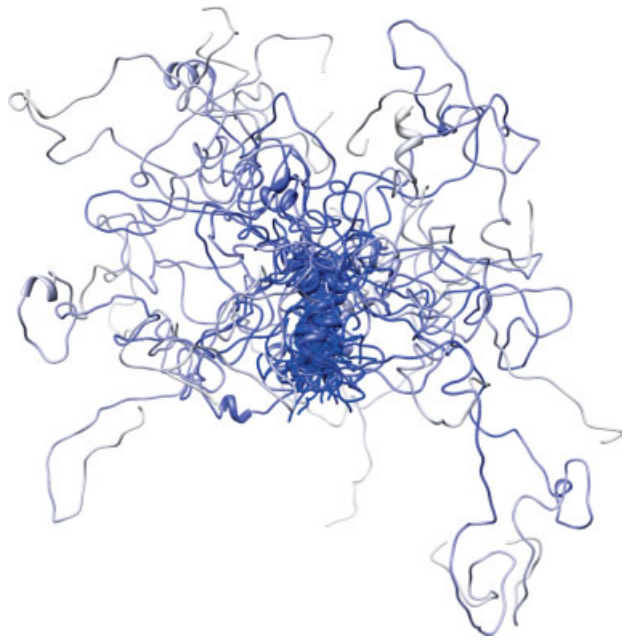

Figure 4. Conformational ensemble of FCP1 generated using the ENSEMBLE software package and refined using the structure constraints displayed in Figure 3. The ribbon is colored pale blue near the N-terminus and transitions to dark blue by the C-terminus, which is the region that superimposes more readily and harbors partial α-helical character

behavior of the ensemble and determining what areas in the conformational space will more likely be sampled than about achieving the highest nominal precision possible. The process of ensemble refinement is inherently computational and so the front-runner software that is currently publically available will be discussed in the following subsection.

Representative Software Packages

Two software packages for generating IDP structures and treating NMR constraints have emerged as the most general and commonly used within the community. The Flexible-Meccano approach, and its more recent evolution into the ASTEROIDS package, has proven highly successful for modeling IDPs.[14] Although not formally intended for NMR applications during its initial development, the TRaDES package provides an alternative to Flexible-Meccano for generating initial trial sets of structures.[28] In this capacity, TRaDES has been bundled with a set of algorithms collectively named ENSEMBLE that has also proven highly effective for IDP ensemble refinement.[25] The following subsections will present and discuss the features and uses of Flexible-Meccano/ASTEROIDS and TRaDES/ENSEMBLE. Continuing with the theme of using FCP1 as a model system, ensembles calculated using each of these two approaches will be shown, although for a shorter construct than previously discussed. These ensembles will be built from a 32-residue construct that was chosen to be of viable size for molecular dynamics (MD) calculations, which will be discussed in more detail next (see section titled 'The Potential of Molecular Mechanics Calculations') as an alternative to the refinement approaches presented in the following subsection.

Flexible-Meccano and ASTEROIDS. One approach to generating an accurate and semiquantitative collection of chain conformations representing the IDP ensemble is to initially generate statistical coil models and use them to back-calculate experimental observables. Where deviation is noted between the coil values and those measured experimentally, local or even long-range structure can be inferred. Flexible-Meccano uses a random-coil dihedral library constructed from dihedral angle distributions found in the loop regions of protein structures, deposited in the PDB, to produce a test set of statistical coils for a provided IDP sequence. Large numbers of structures can be rapidly generated with modern computing power and simple hard sphere clashes used to reject unreasonable structures. If deviations from the expectation values of, e.g., chemical shifts, residual dipolar couplings, or paramagnetic relaxation enhancement parameters are noted between experimental expectation values and those calculated from the statistical coil set, a heuristic approach can be employed to hypothesize ensemble biases. The Flexible-Meccano software then allows the user to iteratively recalculate the bank of IDP conformations, using the hypothesized structural biases as constraints. A representative set of FCP1 conformers generated using this procedure is presented in Figure 5(a), where the protein's bias toward α-helical character near its C-terminus can clearly be seen.

Recently, the ASTEROIDS software package has been developed, which automates the process of integrating NMR structural constraints into the iterative Flexible-Meccano refinement procedure. The NMR constraints most commonly employed in this procedure are chemical shift sets, RDCs, and PREs. As discussed in the section titled 'NMR Spectroscopy of Intrinsically Disordered Proteins', the averaging needed to treat RDC constraints converges very slowly if a global alignment tensor is used, so local alignment procedures are implemented in ASTEROIDS. In order to correctly account for long-range order, the user must therefore rely on corrective algorithms built into the software and/or provide PRE data capable of independently accounting for the influence of long-range interactions.

TRaDES and ENSEMBLE. Another highly successful method for efficiently sampling protein conformational space is implemented in the TRaDES software package. Like Flexible-Meccano, TRaDES can be used for initial generation of statistical coil conformation banks for IDP sequences, but the methodology employed is strongly different. TRaDES uses trajectory directed protein conformer generation, yielding structures that are sterically plausible, but random.[28] Atom placements are achieved by setting interatomic angles to values selected from an approximately normal distribution about mean values defined by residue type and motivated by PDB structures. This is in contrast to Flexible-Mecanno's direct use of monomer conformations extracted from the PDB. One especially useful additional feature of TRaDES conformer pool generation is that the prevalence of cis-proline peptide bonds present in the PDB is explicitly accounted for.

The suite of programs named ENSEMBLE has been paired with TRaDES to generate initial statistical coil conformer pools that are further refined into final IDP ensembles through

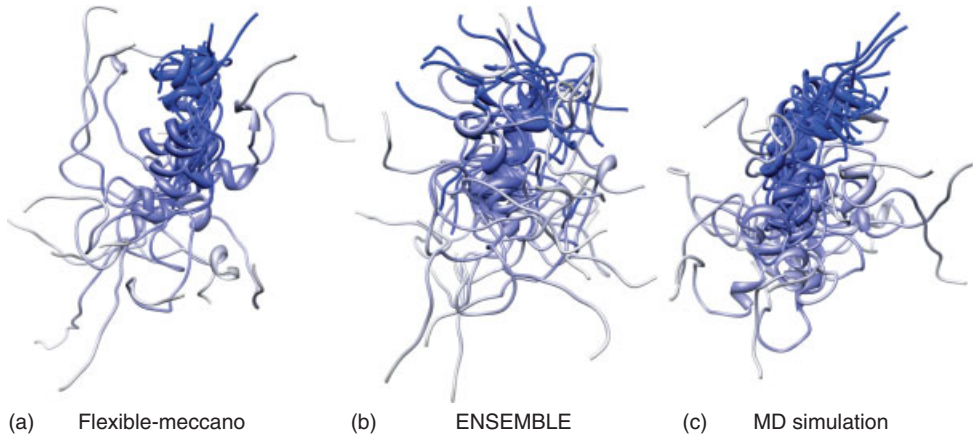

Figure 5. Conformational ensembles of FCP1 (930–961) generated using (a) Flexible-Meccano, (b) ENSEMBLE, and (c) MD calculations performed on the Anton supercomputer. All three bundles are represented as ribbon diagrams with a pale blue N-terminus transitioning to a dark-blue C-terminus. In each case, the structures in the bundle have been superimposed on their α-helical region

the introduction of NMR conformational constraints. Initial ensembles generated from the TRaDES conformer pool are optimized using a Monte Carlo procedure. In later iterations, structures are randomly modified in order to search conformational space for an ensemble that is optimally consistent with the NMR constraints. Similarly to Flexible-Meccano/ASTEROIDS, the ENSEMBLE procedure relies upon a local alignment tensor approach to incorporating RDC constraints, with known caveats that have already been discussed. In our hands, we have found the combination of diverse chemical shift sets, RDCs, and ^{15}N T_2 values to be a robust starting point for refinement in ENSEMBLE, as illustrated by the FCP1 ensemble represented in Figure 5(b). Clearly, the ENSEMBLE generated bundle is highly similar to the Flexible-Meccano ensemble, although with the α-helix more sharply defined.

The Potential of Molecular Mechanics Calculations

With the computing power available to researchers ever increasing, one attractive alternative to statistical coil-initiated refinement techniques is to generate IDP ensembles directly through extensive molecular mechanics calculations. MD computer simulations are useful because they represent the protein as a conformational ensemble that follows the laws of statistical thermodynamics, with the known limitation that the force field chosen for the calculation will influence the outcome. For applications to IDPs, force-field bias is potentially amplified, as modern force fields were not parameterized with the goal of sampling nonfolded conformations in mind. Furthermore, IDPs represent an extreme sampling problem, with very long trajectories likely being required to achieve converged sampling. Traditionally, MD simulations have been limited to the submicrosecond timescale, but recent hardware advances have shattered this ceiling. In our laboratory, we have computed multimicrosecond MD trajectories on the Anton machine, a purpose-built computer for the calculation of long-timescale trajectories.[29] Representative snapshots from the trajectory generated by simulating FCP1 in the AMBER99SB force field for 5 μs are shown in Figure 5(c). While the MD-generated bundle is qualitatively similar to those refined against NMR data using Flexible-Meccano and ENSEMBLE, it is also clear that the region extensively sampling α-helical conformations is extended closer to the N-terminus of the protein in the Anton trajectory. Further studies are required to determine whether this was an artifact induced by the choice of force field, or whether perhaps the conformational state in these few residues was underconstrained by the procedures used to generate the NMR-refined ensembles.

Outlook

The methods needed to generate NMR structural constraints and to measure conformational dynamics for intrinsically disordered systems have largely been established, although broader adoption of ^{13}C-detected techniques represents a potential growth area. Most future growth in IDP structure and dynamics measurement is likely to come in the form of additional advances in software to generate or validate IDP conformational ensembles. Especially promising is the possibility of MD simulations becoming a viable alternative to other methods of model generation. Recently, the blending of MD-generated conformations into the large conformer pools used for refinement against NMR data has been proposed, showing genuine potential.[30] The very existence of IDPs presents new challenges for the structural biology community and new opportunities for innovation by NMR spectroscopists that will likely last for many years to come.

Acknowledgments

This work was supported by an NSF-CAREER award (MCB-0953918) and startup funds from the Pennsylvania State University to SAS. The Anton machine at NRBSC/PSC was generously made available by D.E. Shaw Research. Anton computer time was provided by the National Resource for Biomedical Supercomputing (NRBSC) and the Pittsburgh Supercomputing Center (PSC) through Grant RC2GM093307 from the National Institutes of Health.

Biographical Sketch

Scott A. Showalter born in 1977. He obtained BSc degree in 1999, Cornell University; PhD in 2004, Washington University in St. Louis; 2005–2008, Postdoctoral Fellow, National High Magnetic Field Laboratory; 2008–present, Assistant Professor of Chemistry, The Pennsylvania State University. The Showalter Laboratory applies biophysical chemistry techniques to understand the function of partially disordered proteins and to define the features of protein–RNA interactions. Our research addresses the hypothesis that a conserved functional role for disordered biomolecules exhibiting heterogeneous conformational ensembles renders them uniquely capable of mediating cellular signaling, transcription, and translation.

Related Articles

Structure and Dynamics of Disordered Proteins

References

1. P. E. Wright and H. J. Dyson, *J. Mol. Biol.*, 1999, **293**, 321.
2. J. C. Wootton, *Comput. Chem.*, 1994, **18**, 269.
3. P. Romero, Z. Obradovic, X. H. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, *Protein. Struct. Funct. Dyn.*, 2001, **42**, 38.
4. A. Tantos, K. H. Han, and P. Tompa, *Mol. Cell. Endocrinol.*, 2012, **348**, 457.
5. P. E. Wright and H. J. Dyson, *Curr. Opin. Struct. Biol.*, 2009, **19**, 31.
6. B. Mészáros, P. Tompa, I. Simon, and Z. Dosztányi, *J. Mol. Biol.*, 2007, **372**, 549.
7. A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043.
8. P. Bernado and D. I. Svergun, *Mol. Biosyst.*, 2012, **8**, 151
9. K. N. Hu, R. H. Havlin, W. M. Yau, and R. Tycko, *J. Mol. Biol.*, 2009, **392**, 1055.
10. T. Mittag and J. D. Forman-Kay, *Curr. Opin. Struct. Biol.*, 2007, **17**, 3.
11. M. R. Jensen, R. W. Ruigrok, and M. Blackledge, *Curr. Opin. Struct. Biol.*, 2013, **23**, 426.
12. M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5762.
13. T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, and J. D. Forman-Kay, *Structure*, 2010, **18**, 494.
14. R. Schneider, J. R. Huang, M. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, M. R. Jensen, and M. Blackledge, *Mol. Biosys.*, 2012, **8**, 58.
15. W. Bermel, I. Bertini, I. C. Felli, M. Piccioli, and R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2006, **48**, 25.
16. B. O'Hare, A. J. Benesi, and S. A. Showalter, *J. Magn. Reson.*, 2009, **200**, 354.
17. S. A. Showalter, *Biomol. NMR Assigm.*, 2009, **3**, 179.
18. C. W. Lawrence, A. Bonny, and S. A. Showalter, *Biochem. Biophys. Res. Commun.*, 2011, **410**, 461.
19. C. W. Lawrence and S. A. Showalter, *J. Phys. Chem. Lett.*, 2012, **3**, 1409.
20. M. R. Jensen, L. Salmon, G. Nodet, and M. Blackledge, *J. Am. Chem. Soc.*, 2010, **132**, 1270.
21. M. Rubinstein and R. H. Colby, Polymer Physics, Oxford University Press: Oxford, New York, 2003.
22. J. A. Marsh, V. K. Singh, Z. Jia, and J. D. Forman-Kay, *Protein Sci.*, 2006, **15**, 2795.
23. C. Camilloni, A. De Simone, W. F. Vranken, and M. Vendruscolo, *Biochemistry*, 2012, **51**, 2224.
24. M. Blackledge, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2005, **46**, 23.
25. J. A. Marsh and J. D. Forman-Kay, *J. Mol. Biol.*, 2009, **391**, 359.
26. J. A. Marsh and J. D. Forman-Kay, *Proteins*, 2011.
27. V. N. Uversky, *Protein Sci.*, 2002, **11**, 739.
28. H. J. Feldman and C. W. Hogue, *Proteins*, 2000, **39**, 112.
29. D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. B. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, *Commun. ACM.*, 2008, **51**, 91.
30. K. A. Ball, A. H. Phillips, D. E. Wemmer, and T. Head-Gordon, *Biophys. J.*, 2013, **104**, 2714.