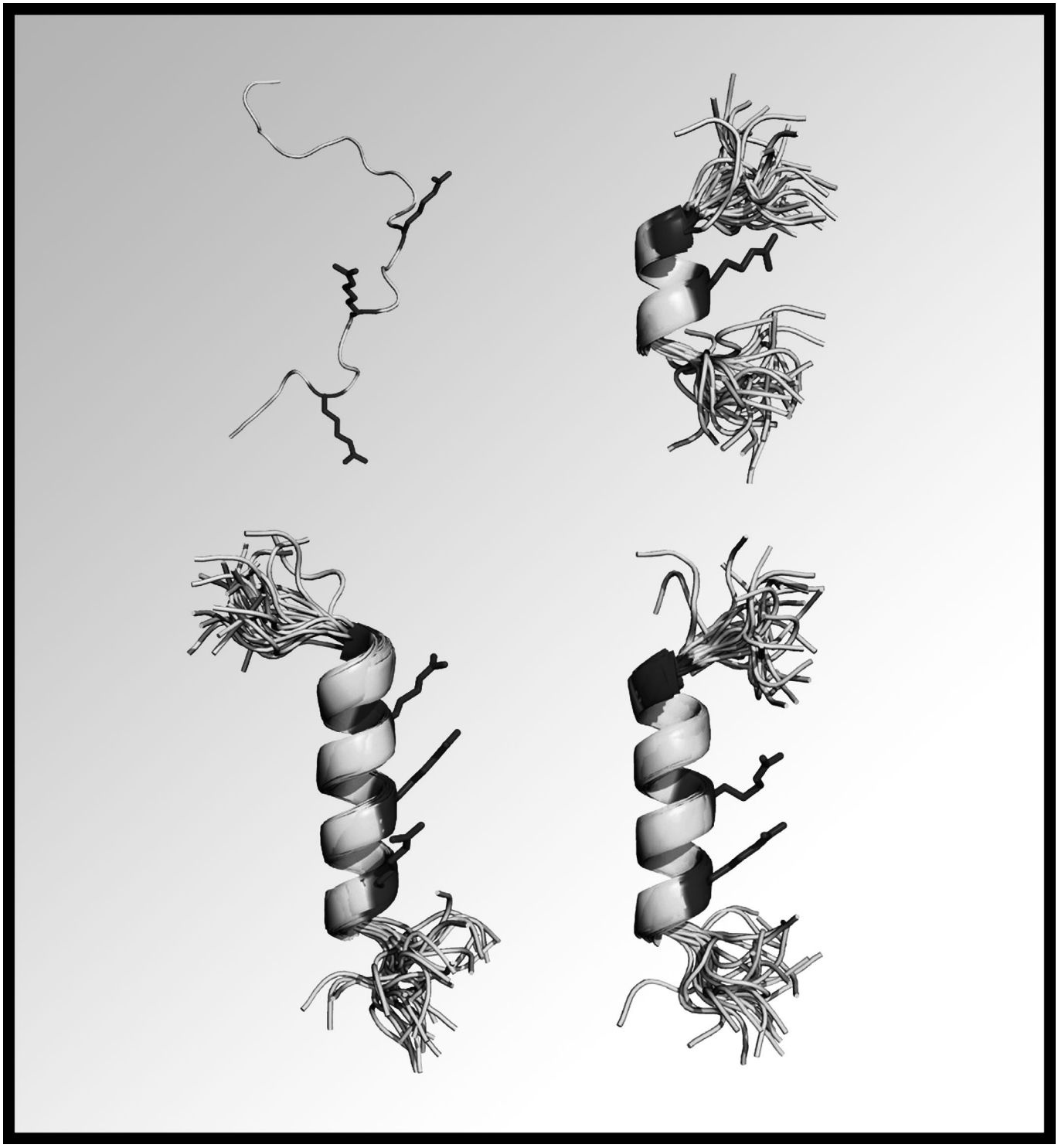
CHEMBIOCHEM

ChemPubSoc Europe

DOI: 10.1002/cbic.201200093

Intrinsically Disordered Proteins: From Sequence and Conformational Properties toward Drug Discovery

Nasrollah Rezaei-Ghaleh,*[a] Martin Blackledge,[b] and Markus Zweckstetter*[a]



Structural disorder of functional proteins under physiological conditions is widespread within eukaryotic proteomes. The lack of stable tertiary and secondary structure offers a variety of functional advantages to intrinsically disordered proteins (IDPs): their malleability of interaction with different partners, specific but low-affinity binding, and their fine modulation by post-translational modifications. IDPs are therefore central players in key processes such as cell-cycle control and signal-transduction pathways, and impairment of their function is associated with many disease states such as cancer and neurodegenerative disorders. Fascinating progress in the experimental characterization of IDPs has been made in the last decade, especially in NMR spectroscopy and small-angle X-ray scattering as well as in single-molecule techniques. It has been accompanied by the development of powerful computational tools to translate experimental results in explicit ensemble representations of IDPs. With the aid of bioinformatics tools, these advances have paved the way to targeting IDP interactions in rational drug-discovery projects.

Introduction

An ordered protein is characterized by a rigid 3D structure defined as a single set of backbone and side-chain atom positions and dihedral angles. This represents an average structure, around which small-amplitude fluctuations of position and angle occur. The singularity of the folded structure originates in the topology of the protein energy landscape, which exhibits a well-defined, funnel-shaped global-energy minimum that is kinetically accessible.[1] With the aid of physical (e.g., high/low temperature or pressure) and/or chemical (e.g., extreme pH or denaturants like urea) destabilizing agents, protein energy landscapes can be converted to relatively flat landscapes, so that the specific 3D structures are disrupted and replaced by either dynamic conformational ensembles of the monomeric protein or its insoluble aggregates.[2] Intrinsically disordered proteins (IDP) make up a category of proteins with weakly funneled or rugged energy landscapes in the absence of destabilizing agents.[1] As a consequence, they do not adopt any well-defined structure under near-physiological conditions, instead they exist as dynamic conformational ensembles in which atom positions and dihedral angles vary significantly over time with no specific equilibrium value. In the literature these proteins are also called "intrinsically unstructured", "natively unfolded", "natively denatured", and "natively disordered" proteins, among others. We prefer to use "intrinsically" rather than "natively" to highlight the possibility of their folding after binding to the partner, and "disordered" rather than "unstructured" to keep open the possible existence of transient local and long-range structures.[3]

It was known as early as the 1970s that some peptide hormones are partially or completely unstructured in their functional forms.[4] However, the occasional discovery of functional disordered proteins was largely overshadowed by the regular appearance of specific three-dimensional protein structures obtained through X-ray crystallography or NMR spectroscopy methods. The situation started to change in the 1990s when improved biochemical methods were increasingly employed. The classical methodology relied on detection of the activity of interest in some homogenate followed by protein purification for structural characterization. This gave IDPs fewer chances to be discovered, mainly due to their lower concentration and due to the fact that IDPs are more sensitive to proteases, which are released during homogenate formation. By contrast, the availability of gene-sequence data and gene-based functional analysis allows a given function to be mapped to a particular gene and transcription of the gene followed by production and purification of the protein. With no inherent bias against disordered proteins, the new methodology resulted in the discovery of many proteins that are disordered under physiological conditions.[5] The latest release (July 2011) of the DisProt database (http://www.disprot.org) contains 1388 disordered regions in 645 IDPs.[6]

Prediction of protein disorder based on amino acid-composition has revealed that a large fraction of eukaryotic proteins are either completely or partially disordered in solution.[7] These disordered proteins have been shown to be functionally important to the cell, and play crucial roles in cellular processes as diverse as signaling, cell-cycle control, molecular recognition, transcription, replication, and chaperoning.[8] The impairment of their physiological roles leads to a wide variety of disease states, including cancers, neurodegenerative diseases like Alzheimer's and Parkinson's disease, cardiovascular diseases, and type-II diabetes.[9] The growing number of discovered IDPs and their ubiquitous presence within eukaryotic proteomes, as well as the broad range of physiological and pathological functions associated with them, have fueled a great deal of enthusiasm for a detailed structural elucidation of IDPs.

Protein structure–function paradigm

Until the 1990s, the central tenet of biochemistry, the so-called structure–function paradigm, was that having a well-defined three-dimensional structure is a prerequisite for the function of proteins.[10] This was often justified by a sort of "specificity" argument. The first premise of the argument was that the complex network of enzymatic reactions and protein–protein or protein–nucleic acid interactions inside the cell will not work

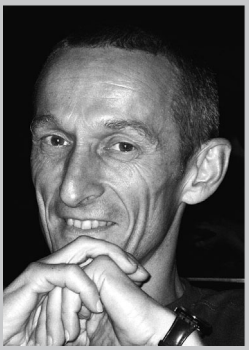
[a] *Dr. N. Rezaei-Ghaleh, Prof. Dr. M. Zweckstetter*
Department for NMR-Based Structural Biology
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11, 37077 Goettingen (Germany)
E-mail: nare@nmr.mpibpc.mpg.de
mzwecks@gwdg.de

[b] *Dr. M. Blackledge*
Protein Dynamics and Flexibility, Institut de Biologie Structurale
UMR 5075, CEA-CNRS-UJF
41 rue Jules Horowitz, 38027 Grenoble (France)

properly unless proteins act in a highly "specific" manner. The second premise, inspired by the "lock-and-key" hypothesis formulated by Emil Fischer to explain the remarkable specificity of enzyme actions, was that without a well-defined geometry and its implied rigidity, at least in the regions involved in molecular binding, the specificity of interaction will not be guaranteed. Experimental evidence, including hundreds of crystal structures for proteins and the in vitro refoldability of proteins into their functional states, was in support of this view.

Even after the discovery of many IDPs in the 1990s, the structure–function paradigm was protected against revision, by arguing that these proteins are incapable of finding their functional fold in the free form, but that with the help of their binding partners, which provide the missing elements of their folding, they will end up in their functional "folded" state. According to this view, binding to a partner plays the same role for IDPs as incorporation into a membrane does for membrane proteins. Although binding-induced folding definitely occurs for many IDPs,[11,12] there is an important subset of IDPs known as entropic chains that carry out their functions without becoming ordered; moreover, the disorder itself functions for them.[8] Several IDPs and regions fall into this category, with their roles described as "entropic springs", "bristles/spacers", "linkers", or "clocks". Furthermore, several IDPs were gradually found to be disordered even within the context of their so-called "fuzzy" complexes.[13] More generally, IDPs structured in the bound forms could also have benefited from functional advantages intrinsic to their disorder in the free form. For example, binding-induced folding provides the right combination of low affinity with high specificity, which is optimal for the protein–protein interactions involved in signal transduction pathways.[11] The higher binding promiscuity of IDPs prepares them to act as hubs in large protein–protein interaction networks, thereby facilitating the transfer of information across different subcellular systems.[14,15] With the continuous growth in the discovery of IDPs, along with a better structural and functional characterization of IDPs that has revealed the functional roles of protein disorder, a reappraisal of the structure–function paradigm was inevitable. The succeeding view seems to be less exclusive than the preceding one: both protein order and disorder are functional.[16,17]

Markus Zweckstetter studied physics at Ludwig Maximilians University in Munich. He wrote his PhD thesis under the supervision of Tad Holak at the Max Planck Institute for Biochemistry in Martinsried and received his PhD from the Technical University of Munich. From 1999 to 2001 he worked as a postdoctoral fellow in Dr. Adrian Bax's group at the National Institutes of Health in Bethesda, Maryland. He has headed the "NMR Protein Determination" Research Group at the Max Planck Institute for Biophysical Chemistry since 2001. Markus Zweckstetter has taught as Honorary Professor of Biology at the University of Göttingen since 2008.



Martin Blackledge is a Research Director of the Commisariat a l'Energie Atomique et aux Energies Alternative (CEA-EA) and Group Leader at the Institute of Structural Biology in Grenoble. He obtained his PhD in 1988 working on phosphorus NMR spectra in human organs under Prof. Sir George Radda (University of Oxford). His first post-doctoral position was in the group of Prof. Richard Ernst, where he used high-resolution NMR spectroscopy to study conformational dynamics in biomolecules. Since 1992, his research at the IBS has focused on the role of conformational dynamics and flexibility in biomolecular function.

Nasrollah Rezaei-Ghaleh graduated in medicine from the Tehran University of Medical Sciences, and received his PhD degree in 2007 from the Institute of Biochemistry and Biophysics of the University of Tehran for his studies on biophysical characterization of protein misfolding and aggregation. Since 2008, he has held a postdoctoral position in the Department for NMR-Based Structural Biology at the Max Planck Institute for Biophysical Chemistry in Göttingen. His research interests focus on the application of NMR methods in combination with computational tools to characterize the structure and dynamics of disordered proteins.

Intrinsic Disorder is Encoded in a Protein Amino Acid Sequence

Sequence signature of IDPs

Sequence analysis of ordered and disordered proteins reveals their distinct compositional features.[16] According to their relative occurrence, amino acid residues can be grouped into three classes: order-promoting residues—C, W, Y, I, F, V, L, and probably H, T, and N—disorder-promoting residues—E, P, Q, S, R, K, M and probably D—and neutral residues—A and G.[18] Clearly, bulky hydrophobic residues promote protein order, whereas a high proportion of polar and charged residues has the opposite effect. As early as 2000, it was noticed by Uversky et al. that IDPs can be distinguished from ordered proteins according to only two parameters: their average net charge (C) and hydropathy (H).[19] With these two parameters constituting a phase space, the line $H=(C+1.51)/2.785$ was shown to efficiently separate the two "phases", that is, ordered and disordered proteins. The amino acid composition of disordered proteins reveals further statistical differences depending on their length.[18,20]

In line with their unique sequence composition, IDPs and ID regions possess sequence patterns that are distinct from those of globular proteins. As shown by Lise and Jones, patterns rich in proline or (negatively or positively) charged residues are over-represented in disordered proteins.[21] Furthermore, it has been demonstrated that sequences corresponding to the non-globular regions of proteins have a local compositional complexity that is significantly lower than that of randomly shuffled sequences, whereas the sequences of globular regions are only slightly different from random sequences.[22] However, although low sequence complexity most often indicates protein disorder, or fibrous or extended structures, not all disordered proteins possess low sequence complexity.[23]

Evolutionary changes in protein sequence are limited by constraints on the residues involved in functional/structural interactions. The ratio of non-synonymous (K_A) to synonymous (K_S) mutations in the sequence is low in the ordered regions, of the order of 0.1–0.2.[24] As expected from the smaller number of structural constraints in disordered regions, their sequence evolution is significantly faster than that of ordered regions. In a comparative study, pairwise genetic distances within disordered and ordered regions of 26 protein families were investigated, and it was shown that disordered regions evolved significantly faster in 19 families and more slowly in two families only.[8] In the case of the sex-determining transcription factor, SRY, its unstructured Gln-rich trans-activator domain had a K_A/K_S of 0.4–0.8 (i.e., fast evolution), whereas that of the globular DNA binding domain was 0.1–0.2 (slow evolution).[25] In addition, the fraction of proteins with tandem repeated short segments is significantly higher in IDPs (39%) than for the proteins in the Swiss-Prot database overall (14%), or for yeast (18%) or human (28%) proteins.[26] This suggests a further evolutionary mechanism, expansion of internal regions, for disordered proteins and reflects their strong evolutionary activity.

Predictors of disordered structure

The sequence features of IDPs described above have been exploited to predict the presence of intrinsic disorder in proteins. Several ID predictors have been developed in recent years, including—but not limited to—PONDR models,[27,28] DisEMBL,[29] FoldIndex,[30] DISOPRED models,[31,32] GlobPlot,[33] IUPred,[34] and NORS.[35] Most of these ID predictors are available as web servers. The underlying algorithms of these predictors differ. For example, FoldIndex uses charge/hydrophobicity scores based on a sliding window to predict disordered regions,[30] this is similar to the suggestion made by Uversky et al. for whole proteins.[19] The PONDR methods are feed-forward neural networks based on amino acid composition, physicochemical properties, and sequence complexity.[27,28] GlobPlot uses the amino acids' propensities for disorder, calculated for each amino acid as the difference between its frequency of occurrence in regular secondary structures (α-helix or β-strand) and outside of them.[33] IUPred estimates the energy of pairwise interactions in a window around a residue,[34] while DISOPRED is based on sequence profiles.[31,32] The currently available disorder prediction tools have an approximate accuracy of 80%.[36] To achieve higher accuracies, different predictors have been combined into metapredictors, such as metaPrDOS[37] or PONDR-FIT.[38] While the current accuracy of ID prediction looks satisfying, it is important to recognize which (operational) definition of protein disorder has been implemented in the prediction procedure.[18]

Computational analyses of sequence data at the genome level indicate that IDPs and ID regions (IDRs) are highly abundant, especially in eukaryotic proteomes. It was found that 7–30% of prokaryotic proteins contain long ID regions of more than 30 consecutive residues. The prevalence of long ID regions among eukaryotic proteins was 45–50%.[39] With the inclusion of short ID regions, the prevalence of protein disorder is expected to be even higher. Interestingly, mitochondrial proteins are less frequently disordered; this is consistent with the prokaryotic origin of this organelle.[40] There could be several reasons for the higher occurrence of ID in the eukaryotic proteome. First, many of the functions associated with disordered regions, such as organization and biogenesis of the cytoskeleton, are unique to eukaryotes.[8] Second, IDPs are highly sensitive to proteolytic degradation and have relatively short lifetimes. In eukaryotes, a large proportion of disorder-containing proteins are located in cellular compartments, such as the nucleus, that are protease deficient and provide some physical protection against degradation. As they do not have cellular compartments, prokaryotes are less able to protect disordered proteins. In addition, there is a strong selective pressure on the biochemical efficiency of prokaryotes, so the cost of short protein life-times is far greater for them than for eukaryotes.[40]

Prediction of functional sites

IDPs often exert their function by binding to structured proteins. Upon interaction with ordered partners, many disordered regions undergo a disorder-to-order transition.[12] Computational studies of such complexes suggest that the binding interfaces of the disordered partners have a conformational propensity for the structures they take upon binding.[41] This leads to transient sampling of such conformations in the unbound IDPs. Experimental evidence has verified the existence of these so-called "preformed elements" in, for example, p53[42] and $p27^{Kip1}$.[43] The majority of preformed elements are α-helical, although β-strand and irregular secondary structures have also been found.[44] Irrespective of the exact structure of these elements in the complex, they can be detected by the disorder predictors described above as sharp dips in the disorder score profiles (Figure 1B).[45] Originally, such a dip was observed in the 4E binding protein (4EBP1),[46] which had been shown by NMR spectroscopy to be completely disordered.[47] Residues corresponding to the observed dip were found to form a short α-helix when 4EBP1 bound to the eukaryotic translation initiation factor 4E.[45,48] Such short fragments of around 20 residues within disordered regions that undergo a disorder-to-order transition upon binding are termed molecular recognition features (MoRFs) and, depending on the structure in their complexes, are classified as α-, β-, or ι-MoRF.[44] It is noteworthy

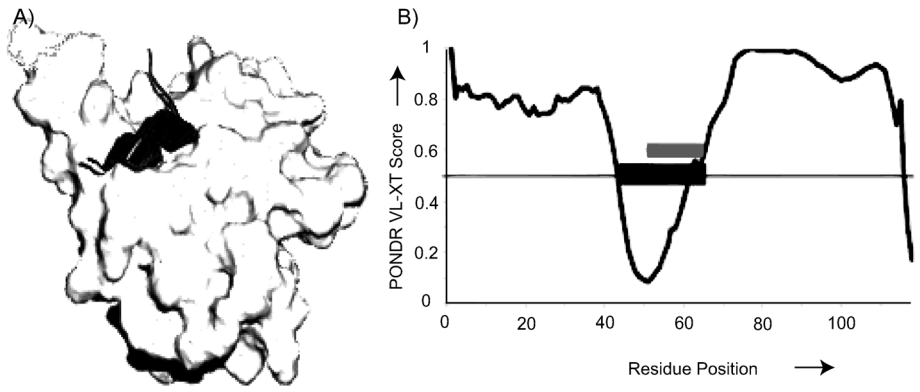


Figure 1. Example of a binding region detected by protein-order prediction. A) Eukaryotic initiation factor (light gray) and the binding region (black) of 4E-BP1. B) The PONDR VL-XT prediction for 4E-BP1. The binding region (gray bar) and the α-MoRF (black bar) are found as sharp dips in the protein-disorder score, spanning a short stretch of residues. Reprinted with permission from ref. [45]. Copyright: American Chemical Society, 2005.

that MoRFs can adopt different conformations when they bind to different partners. For example, the same short segment at the C terminus of p53 adopts four different conformations, one α-helical, one β-strand and two distinct coils, in complex with four different partners.[49]

More than 1000 MoRFs have been found through searching the protein data bank (PDB), and they have been clustered into about 400 families according to sequence similarity.[44] Detailed comparison of MoRF–partner interactions with those of two ordered partners revealed that MoRF–partner interfaces contained a higher fraction of hydrophobic side chains.[50] The higher abundance of hydrophobic groups in MoRFs along with their specific sequence patterns within intrinsically disordered regions has opened the way to developing MoRF predictors from sequence. One of the predictors, α-MoRF-PredII, has been able to predict α-helix-forming MoRFs with an accuracy of 87%.[51] As another predictor of potential binding sites within disordered regions, ANCHOR, picks up disordered segments that cannot form a sufficient number of contacts with their own local sequential neighborhood but are able to interact in an energetically favored manner with general globular partners.[52] From a different starting point, eukaryotic linear motifs (ELM) discovered through sequence analyses have been found commonly to mediate IDP interactions.[53] Since these motifs are short (3–15 residues), their prediction from sequence alone suffers from very high false-positive rates.[54] However, the more recent predictors, like SLiM-Finder, have significantly improved prediction accuracy by applying context filters.[55] The prediction of potential binding sites in IDPs has been proposed as a strategy to identify their structured partners and design short peptides or small molecules to inhibit IDP–partner interactions.[51,56]

Quantitative Methods for the Structural Characterization of IDPs

The structural characterization of IDPs aims to represent these highly flexible proteins as ensembles of interconverting structures. X-ray crystallography, the most common method of structural biology, is obviously incapable of catching the structural heterogeneity of IDPs in their isolated state. This is due to the fact that crystals of disordered molecules are extremely difficult to obtain, and in the rare cases when a crystal is formed, it will represent only one structure out of the thousands existing in solution. Similarly, biophysical techniques such as circular dichroism, or Raman and infrared spectroscopy, which are often used to investigate the secondary structure of ordered proteins, are of limited use for IDPs because these secondary structures are typically transient, weakly populated, and confined to short stretches of the protein sequence, therefore hardly detectable by those techniques.[57] General biochemical techniques such as limited proteolysis might provide clues to the presence of relatively persistent tertiary structure, or size-exclusion chromatography could give the overall dimensions of the studied protein.[17] However, small-angle X-ray scattering (SAXS), NMR spectroscopy, and single-molecule fluorescence combined with computational methods give particularly powerful techniques for the detailed structural characterization of IDPs.

Small-angle X-ray scattering

SAXS is a powerful method of addressing the overall size and shape of biological macromolecules in solution. In a SAXS experiment, a solution of the macromolecule of interest is illuminated by a monochromatic X-ray beam, and the intensity of the scattered beam is measured at different scattering angles. The angular dependence of the scattered intensity contains a wealth of structural information at various length scales. In particular, the low-angle portion of the scattering profile (typically below a few degrees) provides information about the large distances in the particle ($\gtrsim \sim 1$ nm). Traditionally, the Guinier's and Kratky's analyses of scattering data are used, respectively, to yield the average gyration radius of the macromolecule (R_g) and qualitatively to distinguish disordered states from globular particles. Furthermore, with the distance distribution function, $p(r)$, calculated from the scattering profile through a Fourier transformation, the maximum intraparticle distance, D_{max}, is obtained and a rough evaluation of particle shape is made possible. With the advent of novel computational approaches, the scattering profile can now be used to calculate the overall shape of particles through ab initio methods and rigid-body modeling. More strikingly, the applicability of this method is no longer limited to monodisperse solutions. In the presence of assembly (oligomeric mixtures) or conformational (completely or partially disordered macromolecules)

polydispersity, the scattering intensities could, in principle, be decomposed to their individual components, and the overall shape of each component could be reconstructed.[58] Below, we briefly describe the SAXS methods for the study of disordered proteins. More detailed information can be found in some recent reviews.[59,60]

Scattering properties of IDPs: By definition, IDPs do not exist in solution as a single, well-defined structure but extensively sample the conformational space and adopt a very large number of conformational states. The SAXS intensity of any IDP is therefore a linear combination of individual contributions arising from the various members of the conformational ensemble. The final SAXS profile, obtained after averaging over a large number of conformations, is smooth and essentially lacks specific features of the individual conformers. However, there are a few key features in the SAXS data, for example, a (relatively) large R_g, an expanded distance distribution function, $p(r)$, with very large D_{max} values, and a monotonically increasing Kratky plot, which can be used to identify disordered proteins.[61]

The IDRs of proteins are attached to the folded domains as tails, or located as flexible linkers between them. Kratky analysis of SAXS data can reveal the presence of disordered regions in an otherwise folded protein. Globular proteins show a bell-shaped Kratky plot with a well-defined maximum, whereas those of unfolded proteins present a plateau followed by a slow monotonic increase. The presence of an IDR attached to globular domains induces a dual behavior with a clear maximum coming from the folded part followed by a continuous rise due to the flexible part.[61] This behavior was observed for example in replication protein A (RPA). This protein is the main single-stranded (ss) DNA binding protein among eukaryotes. RPA contains seven globular domains connected by flexible linkers, and undergoes substantial interdomain motions that are functionally essential. The dual character of RPA, that is, the simultaneous presence of folded and unfolded parts, is clearly seen in its SAXS profile, especially through Kratky analysis.[62]

Ensemble representation of IDPs by using SAXS: If ab initio or rigid-body-modeling methods are used to reconstruct a "single" conformation for IDPs based on their SAXS intensities, highly elongated shapes will result,[61] although these visualize their relatively extended structure, they can mask the functionally significant presence of more compact conformations in their ensembles. To overcome this limitation, SAXS data alone or in combination with other techniques have been exploited to produce IDP ensembles. Generally, computational methods are used to generate a large "pool" of random conformations (in the order of thousands) for an IDP of a given sequence, then different subensembles of conformations (consisting of tens or hundreds of conformations) are selected to calculate their collective SAXS intensities, and finally the one that optimally fits the experimental SAXS profile is selected as the ensemble representation of the IDP.[63–65]

As an example, this approach (implemented as the ensemble optimization method, EOM[63]) was used to provide an ensemble representation of the N-terminal region of the vesicular somatitis virus phosphoprotein (VSV-P60). The interaction of this protein with the nucleoprotein constitutes a critical step in the encapsidation of VSV.[66] This 68-residue domain is an IDR, as indicated by its CD spectrum, poor resonance dispersion in ^{1}H NMR spectroscopy, and large R_g and D_{max} obtained from SAXS. However, analysis of the chemical shift and relaxation rates by NMR spectroscopy identified two regions with transient helical conformations. The SAXS-based ensemble representation of the domain provided by EOM showed a bimodal R_g distribution containing two subpopulations of compact and extended conformations. The abundance of the compact subpopulation showed the expected variation upon addition of helix-stabilizing (1 M trimethylamine-*N*-oxide) or -destabilizing (6 M guanidinium chloride) agents. Such a bimodal behavior has been observed for some other IDPs such as the cytoplasmic domain of human neuroligin 3[67] or the proline-rich salivary proteins;[68] this indicates that the presence of relatively structured subpopulations might be a generic feature of IDPs.

Similarly, SAXS-based ensembles of flexible multidomain proteins contain relatively compact conformers. The DNA binding protein HMGB1, which consists of two HMG-box domains (A and B boxes) joined by a linker and has an additional acidic tail, exemplifies this feature.[69] It has been suggested that this protein is in a dynamic equilibrium between a collapsed tail-bound and an open tail-free state (Figure 2A). In the collapsed state, the acidic tail interacts with the DNA binding surfaces of the two HMG-box domains, whereas in the open state, these surfaces are exposed. In line with this hypothesis, EOM analysis of the SAXS curve of HMGB1 at low ionic strength revealed a narrow R_g distribution shifted toward more compact structures than the random pool. At higher ionic strength, which mimics the vicinity to DNA, the R_g distribution became wider and shifted toward the larger values (Figure 2B, top). This is probably caused by weakening the electrostatic contacts between the acidic tail and the two HMG-box domains. Interestingly, mutant proteins, in which residues that are involved in this intramolecular interaction were deleted, showed an R_g distribution indistinguishable from that of the random pool (Figure 2B, bottom).

Although the ensemble representation of IDPs provided by SAXS data provides invaluable insight into this class of proteins, SAXS data lack high-resolution information. SAXS-based ensembles of IDPs are therefore only accurate in terms of their distribution for R_g, D_{max}, and anisotropy, but not for their individual molecular configurations.[61] This level of accuracy requires higher-resolution techniques such as NMR spectroscopy. The combined use of SAXS and NMR spectroscopy can benefit from the complementary features of these two techniques in monitoring structural the properties of disordered proteins on various scales. The program ENSEMBLE has been developed to integrate the use of SAXS and several NMR parameters in ensemble descriptions of flexible proteins.[70,71] The structural characterization of Sic1, the cyclin-dependent kinase (CDK) inhibitor, and its hexaphosphorylated form, pSic1, both in the free form and in a flexible complex with Cdc4, is a remarkable example of the success of this approach.[72]

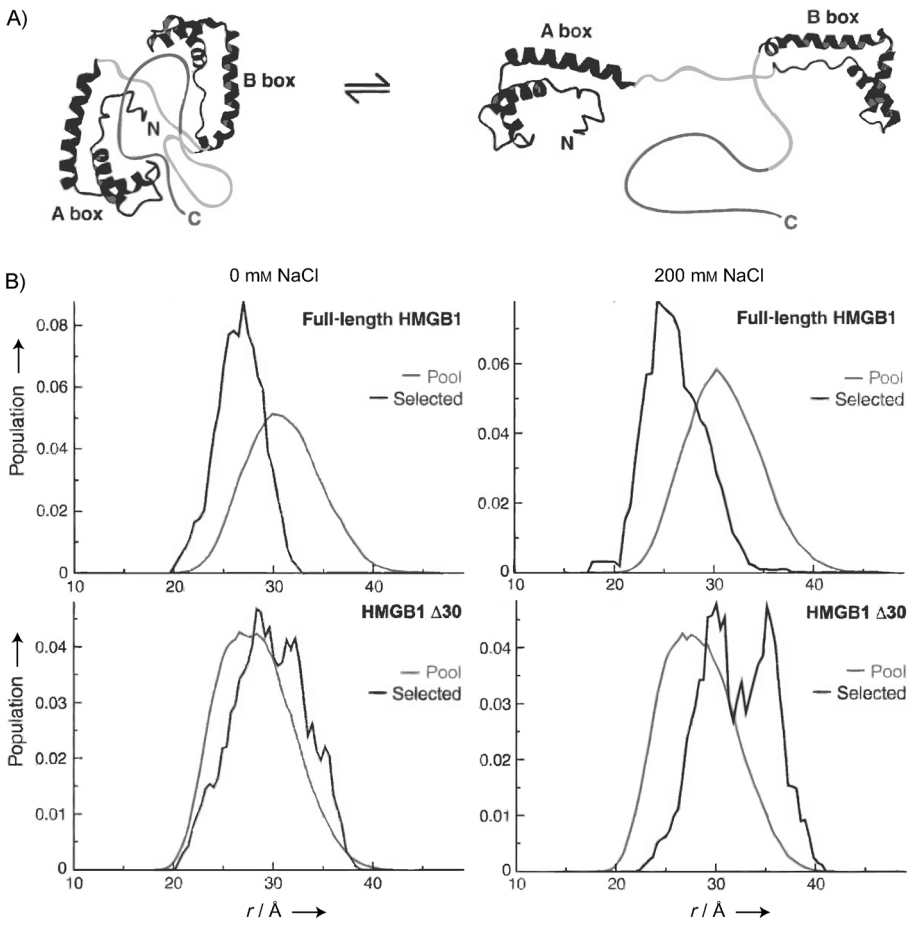


Figure 2. Analysis of SAXS data by EOM for the DNA binding protein HMGB1. A) Schematic representations of a dynamic equilibrium between a hypothetical compact, "tail-bound" form of HMGB1 and a more open form of the protein. In the compact form, the DNA-binding faces are occluded. Vicinity to DNA supposedly shifts the equilibrium toward the open state. B) Distribution of radius of gyration for the selected conformers (black) and the initial pool of structures (gray) obtained through analysis of SAXS data by EOM. Compared to the pool, the selected ensemble has a narrow R_g distribution shifted toward more compact states. The higher ionic strength, mimicking the vicinity of the protein to DNA, causes a shift toward larger radii. See the text for more details. Reprinted with permission from ref. [69]. Copyright: Elsevier, 2010.

Nuclear magnetic resonance spectroscopy

NMR spectroscopy and IDPs: In the past unfolded proteins were not amenable to NMR spectroscopy, mainly due to the poor dispersion of their proton resonances, which led to severe signal overlap in 1D and higher-dimensional proton NMR spectra.[73] On the other hand, NMR experiments recorded on IDPs benefit from higher sensitivities because their NMR relaxation rates (and resultant signal broadenings) are not determined by the overall tumbling rate but by the internal motions, which occur on faster timescales. Several advances in NMR technology, methodology, and sample-preparation techniques, including the use of higher-field NMR spectrometers, isotope-edited and triple-resonance pulse sequences, as well as uniform- and selective-labeling schemes helped to circumvent the signal-overlap problem, thereby making NMR spectroscopy one of the key techniques in the characterization of disordered proteins.[73]

Assignment of NMR resonances: The chemical shifts of ^{13}C and ^{15}N nuclei of proteins are more sensitive to the amino acid type and sequence than those of ^{1}H. Accordingly, the resonances of these nuclei are more dispersed than those of protons even in the absence of rigid tertiary or secondary structure. The standard approach to assigning the NMR resonances of IDPs is to take advantage of heteronuclear dispersion through a combination of triple-resonance experiments, which provide sequential information through matching $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{15}N chemical shifts.[74,75] ^{15}N chemical shift correlations are particularly useful owing to the good signal dispersion along the ^{15}N dimension. However, in case of highly repetitive sequences and when the IDP is very large (more than 200 residues), the NMR spectra will still be highly crowded. One way to tackle this problem is to employ a divide-and-conquer strategy, that is, divide the IDP into several small segments, assign the individual fragments, and finally match their frequencies in the context of the full-length IDP. This strategy was used to assign the backbone resonances of the 441-residue tau protein.[76] When the amino acid sequence of an IDP contains highly repetitive motifs and the backbone chemical shifts show severe degeneracy, more-sophisticated approaches are required. The use of sequential connectivity information provided by C^{α} and C^{β} or N and H^{N} chemical shifts, combined with high-dimensional experiments (in which up to seven different chemical shifts are correlated) and automated data processing, constitutes a powerful method to overcome the poor resolution of IDP resonances.[77,78] When this approach is combined with ^{13}C direct detection and non-uniform sampling, the achievable resolution will be even further increased.[79] In addition, ^{13}C direct detection can reduce the impact of line broadening due to chemical exchange.[80,81] With the current power of assignment tools, protein disorder with low sequence complexity and the resultant spectral crowding no longer presents a formidable obstacle to NMR studies.

Chemical shifts and scalar couplings: Chemical shifts probe the local physicochemical environment of the nuclei of interest. Deviations of experimentally observed chemical shifts from their expected random-coil values, the so-called secondary chemical shifts, are sensitive to secondary structure.[82] Residues of an α-helix show positive $^{13}C^{\alpha}$ and negative $^{13}C^{\beta}$ secondary

shifts, while those of β-strands have negative $^{13}C^{\alpha}$ and positive $^{13}C^{\beta}$ secondary shifts.[82] Typically, a value of ±2 ppm for a combined $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shift ($\Delta\delta$ $C^{\alpha}-C^{\beta}$) over a few consecutive residues indicates fully formed secondary structures, whereas smaller values reflect a transient adoption of corresponding conformations.[83]

Secondary chemical shifts have been widely utilized to detect the residual secondary-structure propensities of IDPs. The combined use of the chemical shifts of different nuclei in a single "secondary-structure-propensity" (SSP) score can improve the accuracy of secondary-structure detection.[83] The secondary chemical shifts of several nuclei consistently denote that residues located in the A and H helices of the fully folded apomyoglobin possess residual α-helical propensity in the acid unfolded form.[84] Similarly, the human lysozyme and its disease-related mutants were shown to contain residual α-helical structures in the unfolded state.[85] As another example, comparison of the SSP scores between α- and γ-synuclein revealed a remarkable difference in the amyloid-forming region, with an increased α-helical propensity observed for γ-synuclein.[86] It has been suggested that the presence of relatively stable α-helical structure causes the lower aggregation propensity of γ-synuclein. The utilization of secondary $^{13}C^{\alpha}$ shifts along with some other NMR parameters showed that unfolded α-synuclein in solution is well correlated with the amyloid fibrils of α-synuclein, that is, the same regions exhibit a propensity for the β-strand conformation.[87]

In addition to chemical shifts, three-bond scalar couplings can be used to detect and characterize residual secondary structure. The coupling between H^N and H^{α} [$^3J(H^NH^{\alpha})$], obtained in a straightforward manner through a variety of NMR experiments[88,89] depends on the backbone angle ϕ. Positive deviation of experimental couplings from the random-coil values reflects β-sheet-like structures;[90] negative deviations are observed in regions of (transient) α-helix and polyproline helix type II (PPII) that can be further distinguished by taking chemical shifts into account.[76] As indicated above, the detection of transient elements of secondary structure requires accurate values for the random-coil state (see, for example, refs. [91]–[93]) as any uncertainty in chemical-shift referencing leads to systematic errors in the observed secondary shifts. The simultaneous use of secondary $^{13}C^{\alpha}$ and $^{13}C'$ or $^{13}C^{\beta}$ shifts can show up inconsistencies in referencing.[94] In a previous study on tau protein,[76] secondary H^NH^{α} scalar couplings appeared to overestimate the propensity for secondary structure slightly. We think that this is caused by two problems: 1) there is only a small database of random values for scalar couplings.[95] 2) chemical-shift values are generally more accurate than scalar couplings. In addition, it is important to keep in mind that the $^{13}C^{\alpha}$ chemical shift of a residue also depends on the conformation of its side chain.[96]

Nuclear Overhauser enhancement: In studies of well-ordered globular proteins, the most structurally informative NMR parameter is homonuclear proton NOE, which reports on the spatial proximity of the two protons (typically within a distance of 5 Å). In the case of IDPs, there are always sequential and some medium-range NOEs but long-range NOEs are generally missing from the NOESY spectra.[73] The absence of long-range NOE peaks does not necessarily exclude the transient formation of long-range contacts in IDPs, first because these contacts are present only in a small fraction of the conformational ensemble, and second, the very rapid dynamics of internuclear proton–proton vectors diminishes the NOE intensities.[73] Even if present, however, the unambiguous assignment of NOE peaks in IDPs is severely hindered by the high degeneracy of side-chain proton chemical shifts. Nevertheless, with advances in labeling schemes (e.g., perdeuteration), the sensitivity of NOESY experiments has increased, and the use of longer mixing times (ca. 1 s) and therefore the transfer of magnetization over longer distances (up to 10 Å) are now possible. By using this approach, several long- and medium-range NOEs have been observed for the unfolded state of the drkN SH3 domain.[97]

Residual dipolar coupling: Residual dipolar coupling (RDC) is a relatively new NMR method of studying disordered states. The dipolar coupling between two nuclear spins *i* and *j* depend on the geometry of the internuclear vector as follows [Eqs. (1) and (2)]:

$$D_{ij} = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2 r^3}\left\langle\frac{3\cos^2\theta(t)-1}{2}\right\rangle = D_{\max}\langle P_2(\cos\theta)\rangle \quad (1)$$

with

$$D_{\max} = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2 r^3} \quad (2)$$

Here θ is the angle of the internuclear vector relative to the external magnetic field, *r* is the internuclear distance and γ_i is the gyromagnetic ratio of spin *i*. The angular parentheses define an average over all conformations exchanging on a sub-millisecond timescale.[98]

The intrinsic dipolar coupling for an amide ^{15}N,^{1}H spin pair is very strong, around 23 kHz; however, with isotropic motion of N–H vectors occurring in free solutions, all possible orientations (θ) are sampled with equal probability, and the measured coupling vanishes through averaging. If proteins are dissolved in weakly aligning media, a slight bias will be introduced that favor some orientations over others, and a small fraction of dipolar coupling will survive after averaging. The media commonly used to introduce weak alignment are lipid bicelles, filamentous phages, lyotropic ethylene glycol/alcohol phases, or strained polyacrylamide gels, in which alignment originates from steric and/or electrostatic interactions between the protein and the medium.[98] In general, RDC values report on the orientation of internuclear vectors relative to the so-called alignment tensor, which describes the net alignment of the protein relative to the magnetic field in terms of an order matrix. In folded proteins there is a global alignment uncoupled from their local fluctuations, therefore RDCs can represent different orientations of the internuclear vector relative to a common molecular frame. However, with the high heterogeneity of the ensemble of IDPs in terms of size and shape, individual conformers will not experience the same alignment. In-

stead, the RDCs of IDPs are often interpreted in terms of local alignment.[99] With secondary-structure elements preferentially aligned in parallel to the magnetic field, for example in elongated cavities, the N–H vectors of α-helix and β-sheet residues will be parallel and orthogonal to the external field, respectively; this leads to the positive and negative N–H RDC values.

N–H RDCs have been measured for several chemically denatured proteins. They were generally found to have negative signs, with the maximal values measured in the center of the protein and a decaying trend toward zero at the termini.[98] This bell-shaped profile of RDCs was theoretically explained by modeling denatured proteins as random flight chains.[100] More specifically, the uniformly negative sign of RDCs in denatured proteins is attributed to the alignment of individual peptide planes and the prevalence of polyproline II (PPII) and extended conformations. On top of the overall bell-shaped profile, the experimental RDCs of unfolded proteins reveal a significant site-to-site variation, as found for example in acid-denatured acyl-CoA binding protein[101] or the unfolded states of apomyoglobin.[102] The site-to-site variation is due to the different degrees of motion that are possible for residues with side chains of different volume: more-bulky residues are more restricted in their motion and have more-negative N–H RDCs.[103] The RDC profile can be further modified by the presence of residual structures.

More recently, direct approaches have been developed (and are still under development) to reproduce RDC patterns in a more quantitative manner by using an explicit ensemble description of disordered proteins. The explicit ensemble models are often generated by using amino-acid-specific coil libraries that define the propensity of each residue for coil conformation. Such libraries have been constructed by taking the non-α-helical and non-β-sheet conformations of a large number of existing high-resolution X-ray structures.[104] In a next step, RDCs are predicted for each individual conformer of the ensemble and averaged over the entire ensemble. The prediction of RDCs relies on shape-based alignment algorithms, which are implemented in programs like PALES.[105] This approach has been applied to a two-domain viral protein, protein X, from the Sendai virus phosphoprotein.[106] This protein contains a natively disordered N-terminal domain followed by an ordered domain. The RDC patterns of the two domains were well reproduced by this approach. The RDCs of the intrinsically disordered Δ131Δ mutant of staphylococcal nuclease (residues 10–140) was also accurately predicted by this approach; this indicated that local conformational propensity, without any residual tertiary fold, is sufficient to explain RDCs. Further success of the statistical coil approach has established its use to predict random-coil RDCs, as implied by the local conformational propensities, in order to constitute a disordered "background" line.[98] The RDC background that is predicted on the basis of statistical coil ensembles is indeed very similar to the side-chain-bulkiness profile, thus highlighting the importance of steric interactions for the composition of Ramachandran plots.[103] Similar to chemical shifts and scalar couplings, any deviation of experimentally measured RDCs from the calculated random-coil values can be termed "secondary RDCs" and used to detect residual structures. This approach was used to detect a long-range contact in α-synuclein, between the N- and C-terminal regions.[107] In the four-repeat domain of the tau protein (K18), the presence of four stable turn conformations was identified by using a similar approach.[108]

Paramagnetic relaxation enhancement: Despite the absence of any rigid tertiary structure in IDPs, it is becoming increasingly clear that a wide range of transient long-range interactions are present.[57] Paramagnetic relaxation enhancement (PRE) is perhaps the most useful NMR parameter for detecting these long-range contacts. Normally, PRE experiments target the quantity $R_2^{paramag}$, the transverse relaxation rate caused by dipolar coupling between the nuclear spin of interest, *I*, and the unpaired electron of the paramagnetic center, *S*. This rate is described by Equations (3) and (4):

$$R_2^{paramag} = {}^1/_{20} D_{IS}^2 (4J(0) + 3J(\omega_I)) \quad (3)$$

where

$$D_{IS} = \left(\frac{\mu_0}{4\pi}\right) \frac{\mu_B g_s \gamma_I}{r_{IS}^3} \quad (4)$$

$J(\omega)$ is the spectral density function, g_s is the electron spin *g* factor, and γ_I is the gyromagnetic ratio of the nuclear spin. The PRE effect shows a strong dependence on the distance between the electron and nuclear spin (r_{IS}), and therefore carries distance information.[109] To measure the PRE, a nitroxide spin label such as *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl)-methyl-methanesulfonothioate (MTSL) is attached to a single cysteine in the protein. In the oxidized (paramagnetic) state, the dipolar interaction between the unpaired electron of this spin label and nuclear spins will increase the relaxation rate of nuclear magnetization, that is, broaden the signal. This PRE effect will be absent in the reduced (diamagnetic) state, which can therefore be used as a reference. Differences in peak intensity between, for example, the two two-dimensional ^1H,^{15}N HSQC spectra obtained in the paramagnetic and diamagnetic states can be used to calculate the PRE and estimate the distance between the paramagnetic center and any given amide.[73] Due to the larger gyromagnetic ratio of electrons compared to protons, the dipolar coupling between an electron spin and a proton spin is much stronger than between two proton spins. Accordingly, the PRE effect can reach farther proton spins (up to 25 Å from the electron center), and shorter-lived contacts will also be detectable.[98]

Residues, that are close (up to 10–15 residues away) to the site of spin-label attachment in the protein sequence always show a significant PRE effect, as they are inevitably placed in spatial proximity to the unpaired electron. For an ideal random chain, the PRE is expected to monotonically decrease along the sequence.[110] Any deviation from this ideal behavior points to the presence of long-range stabilizing interactions between two parts of the protein. Introducing a cysteine residue and its conjugated spin label at several positions in the protein sequence and measuring PRE will then unravel the possible network of long-range contacts. By using this approach, the wild-

type htau40 (with 441 residues including two cysteines at positions 291 and 322) and ten single-cysteine mutants were labeled with MTSL.[76,111] The obtained PRE profiles revealed an intricate network of long-range contacts between different parts of the protein (Figure 3). For example, with the spin label attached to position 15 at the N terminus, intensity ratios of

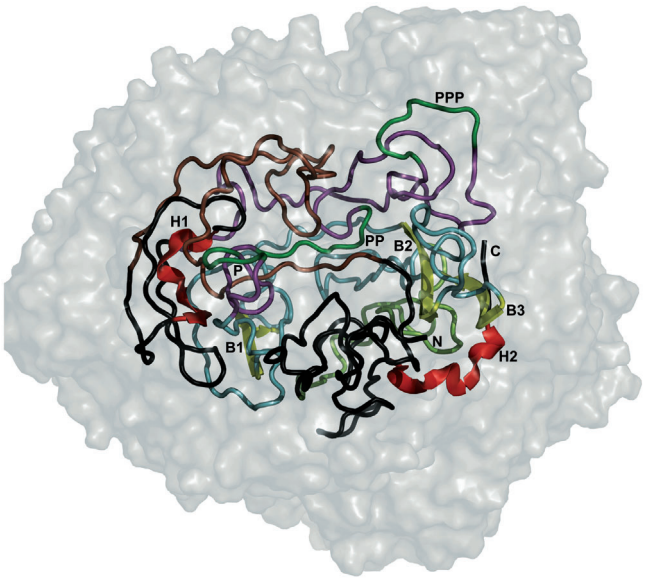


Figure 3. Representation of the conformations of full-length human tau protein (htau40, 441 residues) calculated from a large number of PRE-based distances restraints. Stretches with α-helical, β-structure, and polyproline-II propensity are depicted in red, yellow and green, respectively. In the background, an ensemble of conformations in agreement with the experimental distance restraints is shown in light gray. Reprinted with permission from ref. [76]. Copyright: Mukrasch et al., 2009.

approximately 0.8 were observed for residues 130–200, the proline-rich domain P2 (residues 200–240), and residues in the repeat region up to residue 330. On the other hand, PRE profiles of C291/C322- and C384-MTSL-labeled htau40 indicated that the C-terminal domain (residues 360–441) transiently contacts the repeat region and the 40 N-terminal residues.

Although PRE remains one of the most effective NMR tools for studying IDPs, some caution should be exercised in its use and in interpreting PRE results. The structure and dynamics of a protein might be perturbed by the mutation itself and/or the hydrophobic moiety of the spin label. Therefore, the correct choice of mutation site is important, as is the measurement of several different MTSL positions to probe for consistency.[112] In addition, quantitative interpretation of PRE in disordered proteins is not trivial. Owing to the r^{-6} dependence of the PRE effect, the ensemble average of PRE is dominated by contributions from compact members. This helps to detect lowly populated compact states in IDPs, but on the other hand, complicates its quantitative structural interpretation because the PRE-derived distance will be much smaller than the (linear) average distances in the ensemble.[112] Moreover, in IDPs, the distance between the spin label and the nuclear spins of interest is subjected to rapid fluctuations caused by segmental motion. With the proper timescale, this distance fluctuation provides an additional relaxation mechanism.[109] The flexibility of the spin label constitutes a further challenge, as it introduces a large experimental uncertainty into the PRE-derived distances. However, a recent study has shown that explicit modeling of spin-label mobility can significantly improve the reproduction of experimental PRE data.[113]

Relaxation studies: The backbone and side-chain dynamics of disordered proteins can be studied through NMR relaxation measurements. The conventional NMR techniques for characterizing backbone dynamics involve measuring ^{15}N R_1, R_2 and steady-state ^{1}H,^{15}N heteronuclear NOE. The model-free formalism[114] is not usually applicable to highly disordered proteins because the assumption of a single global correlation time is not valid, and temporal deconvolution of internal motions and molecular tumbling is not possible. Instead, the reduced-spectral-density mapping analysis of relaxation rates provides valuable insight into protein backbone motions occurring on the ps–ns timescale.[115] Simple models of polypeptide chains, in which the dynamics of backbone amide groups are dominated by unrestrained segmental motions, predict that the relaxation rates reach a plateau in the central region of the protein sequence.[116] By using such a model, the ^{15}N relaxation-rate profiles of the unfolded lysozyme in water and 8 M urea were approximately reproduced and yielded a similar persistence length of seven residues.[116,117] However, the observation of site-to-site variation in relaxation rates indicates that the simple approximation of uniform backbone-dependent interactions between neighboring residues is not sufficient to explain the relaxation rates of disordered proteins.

Of the ^{15}N relaxation rates, the exchange-mediated transverse relaxation rate (R_{ex}) represents backbone motions on the μs–ms timescale. This rate can be estimated by comparing the apparent R_2 and $R_{1\rho}$ (longitudinal relaxation rate in a rotating frame),[118] or alternatively, it can be derived as the difference between apparent R_2 and exchange-free R_2 calculated from cross-correlated relaxation rates (η_{xy}).[119] As an example, with this approach, a μs–ms motion was detected in the middle residues (23–29) of unstructured amyloid β (Aβ) peptide. It has been suggested that this motion is linked to the transient formation of a hairpin structure that is stabilized in the aggregated states of the peptide.[119] Moreover, side-chain dynamics measured through ^{2}H and ^{13}C relaxation rates can be used to probe hydrophobic interactions in disordered states.[120] Recently, the NMR relaxation–dispersion method has been exploited to derive chemical shifts of a weakly populated disordered state of the FF domain and calculate its structure on the basis of chemical shifts.[121] This method offers a fascinating way to detect and characterize the weakly populated disordered state of proteins that are globular under native conditions.

Ensemble description of IDPs from NMR data: The deviation of measured NMR parameters such as chemical shifts, scalar couplings, RDCs, and PREs from the predicted random-coil values reveals the local and long-range structural propensities of IDPs.[57] The next step is to translate this information into a quantitative description of IDPs as represented by an ensem-

ble of conformations. Remarkable progress has been made toward this aim.[122,123] In one approach, restrained molecular-dynamics (MD) simulations are run in parallel over multiple replicas of the studied IDP in order to guide the protein ensemble to regions of conformational space that agree with experimental data. This bias in conformational sampling is achieved through the introduction of additional terms into the physical potential of simulation. These terms disfavor conformations for which the expected NMR observables deviate from the measured ones. A number of programs are used to predict NMR observables for individual conformers, and then the ensemble is averaged. Chemical shifts are usually estimated by using programs such as SHIFTX,[124] SHIFTS,[125] SPARTA[126] and CamShift,[127] and RDCs are typically predicted by PALES.[105]

In a recent study, experimental RDCs were used to carry out an ensemble-restrained MD simulation of chemically denatured ubiquitin. Refinement by RDCs caused significant changes in the starting statistical coil models of the protein, and the final ensemble contained several long-range orders including an interaction between strands β1 and β2 of the native protein.[128] This interaction had previously been detected in other types of experimental data.[129,130] In another example, Allison et al. studied α-synuclein by using a large number of PREs in combination with the R_g obtained from SAXS.[131] The generated ensemble was validated by demonstrating an excellent agreement with distance distributions obtained through electron-transfer (ET) measurements.[132]

Despite its advantages and popularity, the general applicability of ensemble-restrained MD simulation is arguable. The major concern is the efficiency of conformational-space sampling and probable limitations caused by the nonphysical terms added to the force field of simulation.[122] To mitigate this drawback, an entirely different approach has been devised to construct ensembles for IDPs. In the first step of this method, conformational space is extensively sampled without any experimental constraints. This sampling is performed either through enhanced MD techniques such as replica exchange,[133] accelerated MD[134] or quenched MD,[135] or alternatively with statistical coil models like Flexible-Meccano[136] or TraDES[71] to produce a very large pool of the order of tens of thousands of conformers. In a second step, a subensemble (of the order of hundreds of conformers) is selected to minimize the difference between the predicted and experimental data. Different tools have been developed for the selection step, for example ENSEMBLE[71] and ASTEROID.[113]

Using the Flexible-Meccano/ASTEROID method, Salmon et al. generated an ensemble representation of α-synuclein.[113] They used a combination of RDC and PRE data to define both the local and long range structural features of this highly flexible protein. The final ensemble contained long-range order between distant regions of α-synuclein, as was previously shown and suggested to play a key role in the protein's aggregation behavior.[137] A similar approach augmented with extensive measurement of RDCs ($^1D_{HN}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, $^2D_{HNC'}$) was utilized to provide an explicit ensemble description of N_{TAIL}, the intrinsically disordered C-terminal domain of the Sendai virus nucleoprotein (Figure 4). The molecular recognition element of N_{TAIL}

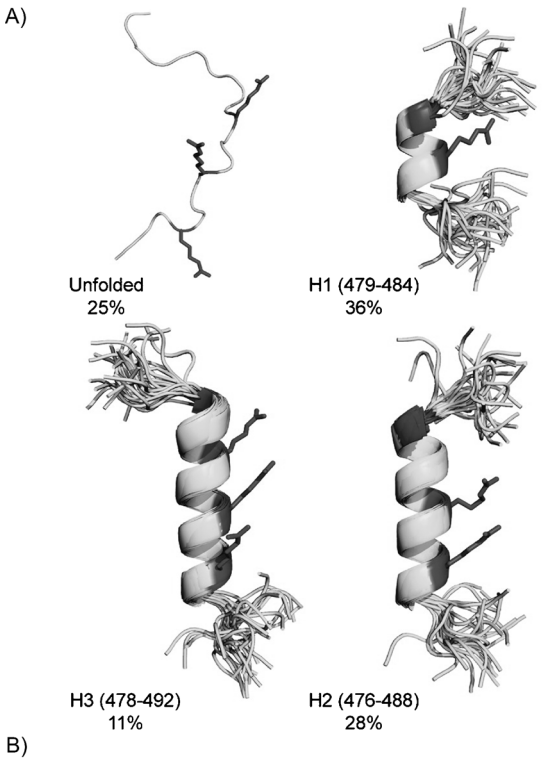


Figure 4. A representation of the conformational ensemble of N_{TAIL} in solution, obtained by the Flexible-Meccano/ASTEROID method using multiple RDCs. A) Four conformations are illustrated with their populations: a single structure representing unfolded conformers, and H1–H3, which contain helical elements of increasing length. Note the residues that form the helix and the N-capping residues (marked as gray). The side chains of three arginines that make up the molecular recognition site are also shown. B) Primary sequence of the central region of N_{TAIL} showing the positions of the arginines responsible for the interaction with the phosphoprotein and the N-capping amino acids. Reprinted with permission from ref. [138]. Copyright: American Chemical Society, 2008.

populated three specific overlapping helical conformers within the ensemble. The use of multiple RDCs improved the quality of the ensemble descriptions.[138] In a later study by the same group, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ and ^{15}N chemical shifts were used to construct an ensemble for the α-MoRF element of N_{TAIL}.[139] The ensemble showed an excellent agreement with the experimental shifts and was well cross-validated against RDCs. This striking result suggests that, despite their lower conformational information content than RDCs, chemical shifts provide a convenient way to probe the conformational behavior of disordered proteins with reasonable accuracy.

Although the new computational methods exemplified above have pushed the modeling of IDPs a big step forward, it should be noted that the atomic-scale ensemble description of IDPs inherently represents an ill-posed problem, that is, the number of experimental data that can possibly be measured for any system is much smaller than the number of degrees of conformational freedom. This inevitably leads to degeneracy,

that is, more than one ensemble exists that could fulfill a single set of experimental data. As a consequence, agreement with the experimental data does not imply that this ensemble is the only possible ensemble to describe the IDP of interest. One possible way to mitigate this problem is to generate several different ensembles that are in agreement with the experimental data, and then try to identify the common structural elements in them.[71,135] The other general limitation of these methods is related to the prediction of NMR observables for the conformational ensembles. Any NMR experimental parameter has its own characteristic averaging timescale, hence its value is not only determined by the conformational properties of the ensemble, but also by the rate of interconversion within the ensemble.[122] Thirdly, experimental data such as PRE that strongly depend on the relevant distances (r^{-6}) can be almost entirely contributed by a small number of compact members of the ensemble, thus leaving the rest of the ensemble severely under-restrained.[112] This type of limitation can be overcome by using experimental parameters with various size dependences, for example, combining NMR data with R_g obtained through SAXS.[131]

Single-molecule FRET

IDPs exist as extremely heterogeneous ensembles of conformations. Their structural/dynamic properties will therefore follow relatively wide distributions. Quantities measured by many biophysical techniques are bulk properties averaged over the ensemble of conformers; information on their actual distribution is lacking. With bulk measurements, different distributions will be indistinguishable if their averages are identical. Single-molecule measurements have the advantage of providing distributions as such, without any loss of information through averaging. Single-molecule (sm) FRET—the nonradiative transfer of singlet excitation energy from donor to acceptor chromophores—has been used to detect the conformational heterogeneity of IDPs. The transfer efficiency is strongly distance dependence, that is, proportional to r^{-6}. To measure a specific distance, the two fluorescence probes (donor and acceptor) are covalently linked to existing or introduced cysteine residues at the positions of interested, and then the FRET efficiency between them is measured. Single-molecule resolution in FRET measurements can be obtained in two ways: experiments can be carried out on "freely diffusing" or "immobilized" molecules. In the case of "freely diffusing" molecules, a large number of molecules in solution are observed one at a time, from these a histogram of FRET efficiency is generated. This reveals how the studied intramolecular distance varies within the ensemble. On the other hand, single "immobilized" molecules are observed for long times (seconds) so as to monitor their conformational fluctuations.[140]

Single-molecule fluorescence techniques were exploited to unravel the structure and dynamics of the natively unfolded yeast prion protein Sup35.[141] The prion-determining domain of Sup35, NM, consists of two distinct regions: an amyloidogenic N-terminal region (residues 1–123) that is abundant in uncharged polar residues (Asn, Gln and Tyr), and a charged solubilizing middle region, M (residues 124–250). The sm-FRET experiment on a dual-labeled amyloid core variant (N21C/S121C) showed a compact Gaussian distribution of FRET efficiencies centered around $0.8E$; this is indicative of a relatively uniform population of conformations with an average distance of $\approx$43 Å between residues 21 and 121 (Figure 5 top). Upon increasing the GdmCl concentration from 0 to 6 M, the single

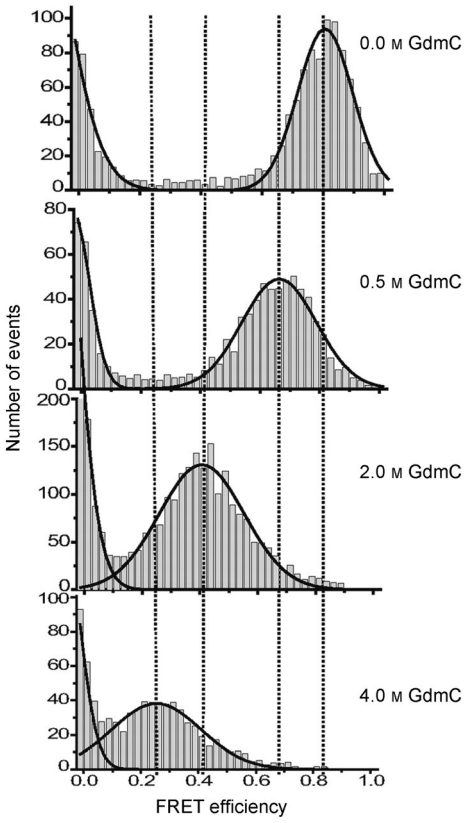


Figure 5. FRET-efficiency histograms of NM at various concentrations of GdmCl. Upon increasing the concentration of the denaturant, the single non-zero peak centered around $E = 0.8$ is gradually shifted to lower E values; this indicates that the distance between the two dyes increases. Reprinted with permission from ref. [141]. Copyright: The National Academy of Sciences, 2007.

non-zero FRET peak was gradually shifted from higher to lower FRET efficiencies; this reflects an increase in the distance between the two dyes from the native 43 Å to the unfolded 63 Å value. The non-cooperative nature of the unfolding transition suggested that the N region of NM domain constituted a multitude of interconverting species, rather then a small number of discrete conformers. Further study with sm fluorescence correlation spectroscopy (FCS) experiments revealed fast conformational fluctuations on the 20 to 300 ns timescale. The membrane-induced folding of α-synuclein was also studied by sm-FRET.[142] It was shown that two folded conformations are encoded in the protein sequence, one with extended and the other with broken α-helical structures; their relative population is fine-tuned by the lipid composition of the membrane. Multiple preferred conformations were also found in the N-terminal

domain of p53.[143] Interestingly, only some of those conformations were able to interact with the DNA binding domain.

Structural Characteristics of IDPs

Overall dimensions

One of the most unambiguous features of the different conformational states of proteins is their hydrodynamic dimension. For instance, the hydrodynamic radius of globular proteins increases by 15–20% when they transform to a molten globular conformation. Pre-molten-globular and unfolded states possess even larger hydrodynamic volumes. The hydrodynamic radii of proteins can be obtained through various techniques such as dynamic light scattering (DLS) and pulsed-field gradient NMR. Early hydrodynamic measurements of IDPs revealed that they can be clustered into two subgroups with different molecular-mass dependence of their hydrodynamic volumes. One group of IDPs behaves as random coils, whereas the other group resembles pre-molten-globular states. Interestingly, the simultaneous use of far-UV CD spectroscopy and hydrodynamic methods indicated that the more compact proteins of the second group (pre-molten-globule-like) contained higher amounts of secondary structure than the less compact coil-like proteins of the first group.[17]

The radius of gyration of an excluded-volume random-coil polymer is expected to follow Flory's equation,

$$R_g = R_0 N^{\nu} \quad (5)$$

Here N is the number of residues in the polymeric chain, and ν is theoretically estimated to be ~0.6. The SAXS-based measurement of R_g for a series of chemically denatured proteins yielded a ν value of 0.598 ± 0.028 in close agreement with the theoretical expectation, indicating the random coil nature of denatured proteins.[144] It is not clear whether IDPs obey the same length dependence as denatured proteins. Indeed, some theoretical and experimental considerations suggest that presence of denaturants perturb conformational sampling behavior of protein residues favoring more extended conformations.[145,146] Generating coil library-based conformational ensembles for 23 proteins, predicting SAXS curves and then deriving R_g through Guinier's analysis, it was shown that,

$$R_g = (2.54 \pm 0.01) N^{(0.522 \pm 0.01)} \quad (6)$$

with a ν value (0.522) that is markedly smaller than that of unfolded proteins (0.589) and closer to the ideal Gaussian polymeric chain.[147] This indicates that, in the absence of any secondary-structure propensity, IDPs are more compact than chemically denatured proteins. Experimentally obtained R_g values for several IDPs are close to the predicted ones through Equation (6), although some significant discrepancies are also observed.[61] Interestingly, any departure from predicted behavior is usually in line with their predicted or independently observed secondary- or tertiary-structure propensities. For example, a number of tau protein constructs have relatively large R_g values; these are consistent with their high number of proline residues, which form extensive PPII structures, and the presence of partially structured regions as shown by NMR.[76,108] On the other hand, α-synuclein has an R_g lower than expected; this reflects the presence of transient long-range contacts and the resultant chain compaction.[107]

Residual secondary structure

As previously discussed, the amino acid composition of IDPs is distinguished by a low number of bulky hydrophobic residues. As a consequence, hydrophobic collapse does not occur for IDPs. A decrease in the local activity of water accompanied by hydrophobic collapse is crucial to strengthening hydrogen bonds in secondary structures.[148] Therefore, without hydrophobic collapse, any possibly existing secondary structures will be fluctuating without stabilizing forces. The presence of such transient secondary structures is difficult to detect by conventional techniques such as far-UV CD spectroscopy, although some earlier studies were able to distinguish between the two subgroups of IDPs based on their secondary-structure content.[17] Much more knowledge about the secondary structure of IDPs has been obtained in recent years. In particular, the utilization of NMR parameters that are sensitive to the local structure, such as secondary chemical shifts and scalar couplings as well as RDCs, revealed that IDPs are far from random coils and contain a significant amount of transient secondary structure.[57] These flickering elements of secondary structure are often observed in regions where the IDPs adopt a similar structure when bound to an interaction partner[41] or when they are in their aggregated state.[87,149] This is similar to what is commonly observed in chemically denatured states: the secondary-structure elements of native proteins might partially exist in the unfolded state (hence the term "residual").[84,150] These so-called preformed elements or MoRFs facilitate folding of the IDPs upon their binding to their partners. The presence of these elements can also be associated with aggregation, as shown in the case of amyloid β (Aβ) peptides. Aggregation of this intrinsically disordered peptide into an oligomeric and/or fibrillar state is believed to play a pivotal role in Alzheimer's disease pathogenesis. It was shown that the same regions of Aβ sequence that are converted into β-strands in Aβ oligomers and fibrils tend to adopt an extended, β-strand like conformation in the monomeric state in solution.[119,149] It has been demonstrated that binding to an inhibitory peptide or zinc ion modulates Aβ's transient secondary structures and influences its aggregation.[119,149] The transient formation of local structures in IDPs is well illustrated in the full-length tau protein: a combination of NMR chemical shifts and scalar coupling constants reveals distinct stretches of 7–11 residues that transiently adopt α-helical, β-strand, and polyproline-II (PPII) conformations. In addition, RDC data support the secondary-structure propensities of these residues and detect stable turns at specific sites.[76,108]

Although the transient formation of secondary structure in most IDPs is well known, the cooperativity of these structures within IDPs is less clear. For instance, if a stretch of residues

possess negative secondary C^{α} and positive C^{β} chemical shifts, does it imply that these residues adopt β-strand-like conformations simultaneously (within the same members of the ensemble) or at different times (spread within the ensemble)? In globular proteins, the formation of α-helices and β-sheets is highly cooperative. However, as mentioned in relation to the stability of secondary structures in the absence of hydrophobic collapse, the cooperativity of secondary structures in IDPs could be very different from that in globular proteins. It has been suggested that the PPII helix is an abundant conformation of disordered proteins, and the individual residues spend a considerable amount of time in this conformation;[151] this results in the observed CD spectra and NMR ^3J couplings. However, experimental and modeling approaches have shown that the PPII conformation in disordered proteins is not cooperative and consecutive residues do not adopt the PPII conformation simultaneously (the persistence length can be as short as two residues),[152] therefore disordered proteins cannot be represented as a chain of extended and long PPII helices.

Transient long-range contacts

When the first NMR-based characterization of denatured proteins was published in the early 1990s, it was found that disordered proteins contain a significant amount of residual structure.[153] This was in apparent contrast to the size and hydrodynamic behavior of denatured proteins and indicated that they were principally random coils ("reconciliation problem").[154] The experimental parameter used to probe the long-range contacts was mainly interproton NOE, which, as described above, was of limited sensitivity and could only detect contacts within a short distance and with a relatively long persistence time. As the characterization of unfolded and partially folded states lies at the heart of the "protein folding" problem, more-powerful techniques were gradually devised to characterize the tertiary structure of disordered proteins. In NMR, PRE and RDC were progressively employed to monitor the transient formation of long-range contacts. On the other hand, sm-FRET provided a nanoscopic ruler to measure intramolecular distances in the individual protein molecules. Utilization of NMR probes led to the detection of a rich network of long-range contacts in disordered proteins, for instance in the denatured Δ131Δ fragment of staphylococcal nuclease[155] and the unfolded state of the drkN SH3 domain.[156] IDPs were shown to contain an even more intricate networks of long-range contacts, as exemplified by full-length tau protein.[76]

The apparent inconsistency between the hydrodynamic and structural behavior of disordered proteins can generally be reconciled if the different distance dependences of the experimental parameters are taken into consideration. Structural parameters such as NOE and PRE or FRET are strongly dependent on the distance (r^{-6}), therefore a small fraction of compact conformers could lead to a measurable effect. However, such a small fraction might not cause a detectable difference in the R_g or R_h of the disordered protein because they are differently dependent on distance and their ensemble average will be nearly totally determined by the majority of the coil-like conformers. The other important point concerns the simultaneous presence of tertiary contacts. As demonstrated for the denatured Δ131Δ fragment, if a "single" representative structure is forced to fulfill all the PRE-based distances simultaneously, the calculated structure will possess a native-like topology with an overly compact shape,[155] which is obviously incorrect. Alternatively, the distance restraints can be satisfied by separate members of the ensemble, or equivalently, the corresponding contacts might form at different times. Indeed, the ensemble MD simulation of the denatured Δ131Δ fragment has shown that, although several native, long-range contacts are transiently formed in this protein, their simultaneous existence occurs very rarely.[157] The full-length tau protein seems to follow a similar scenario. The large number of tertiary contacts observed in the PRE experiments do not exist in a single compact structure, instead they are spread over an ensemble of relatively extended conformations.[76]

Increasing evidence suggest that the presence of residual secondary and tertiary structure is of functional or pathological importance. Residual structures can modulate the binding properties of IDPs, post-translational modifications, susceptibility to proteolytic degradation or aggregation propensity. For example, a temporary contact between the N- and C-terminal regions of α-synuclein is present, and removal of the C-terminal region increases protein aggregation.[137] In the tau protein, heteronuclear NMR spectroscopy has been successfully employed to identify all the phosphorylation sites and provide kinetic data for the enzymatic modification of each site.[158,159] The long-range contacts present in solution in the wild-type monomeric tau are modified by mutations that mimic phosphorylation, and the mutated protein shows enhanced aggregation.[111] Upon aggregation, the network of long-range interactions in tau is further modulated, with two discontinuous epitopes of a conformation-specific antibody coming into closer contact; this provides a rationale for the increased affinity of the antibody to tau aggregates when compared to monomeric protein.[160]

Protein quartet versus ensemble view

As mentioned above, reappraisal of the structure–function paradigm has led to its being succeeded by a less-exclusive view that the functional state of proteins is not limited to the well-ordered state but also includes disordered forms. It was initially formulated in the "protein trinity" model,[16] soon replaced with the "protein quartet" model.[17] According to the latter model, a broad range of structural features exist in the equilibrium conformations of proteins, classified as well-ordered forms, molten globules, pre-molten globules, and random coils. The well-ordered state is characterized by a densely packed hydrophobic core that is the origin of the cooperative conformational changes in folded proteins. Before cooperative disruption of the core occurs, there are only small-amplitude thermal fluctuations around the well-defined average structure of the protein with or without small cooperative conformational changes in the local sequence neighborhood. The other three states, on the other hand, lack a hydrophobic

core and consequentially undergo random non-cooperative conformational changes with no single equilibrium structure.[18]

Although they adopt a well-ordered structure, folded proteins are still subjected to motions of different amplitudes occurring over a large range of timescales. Protein dynamics is crucial to biological functions, including enzymatic catalysis, recognition events and transport processes. There is a growing tendency among structural biologists to represent proteins not as a single "static" conformation but as an ensemble of conformations describing protein dynamics on various timescales. This is accompanied by the realization that the weakly populated members of the protein ensemble might play important roles in protein function,[161] thereby implying that a complete ensemble representation of proteins is essential to capture all their functional potentialities. Conformational ensembles of the globular protein ubiquitin have been calculated by using NMR and molecular dynamics simulations.[162,163] The conformational heterogeneity of ubiquitin within the ensemble was taken to support a mechanism of conformational selection for binding.[162]

The necessity to view a protein structurally as an ensemble of conformations is more evident for IDPs. With their characteristic amino acid sequence properties, IDPs and regions are not able to form a hydrophobic core over which proteins could fold into a single, well-ordered structure. In general, IDPs and regions can be in any of three disordered states: they can be either collapsed (as a molten globule) or extended (as a premolten globule or a coil) or can contain various amounts of flickering secondary and tertiary structure.[9] The protein-quartet model tries to bring all the various features of IDPs (such as their size, and local and long-range structural orders) into a single frame. Alternatively, the ensemble view highlights the heterogeneity of the conformational ensembles of IDPs, which fundamentally cannot be reduced to snapshot descriptions.[83] All the experimental parameters are averaged over such heterogeneous ensembles. The experimental parameters are differently dependent on the timescales of interconverison between ensemble members,[98] and are sensitive to different degrees to the presence of outlier conformations. For example, if a single structure is forced to fulfill the measured SAXS data, the result will be a highly elongated structure, whereas the same procedure for the NMR data will lead to an overly compact structure. This reveals that conformational heterogeneity of IDPs cannot be reduced to any single average structure.

According to the ensemble view, not only the average but also the distribution of the conformational characteristics should be provided in any accurate description of IDPs. With bulk measurements such as SAXS or NMR, this goal is achieved by constructing explicit ensembles that best fit the experimental data. On the other hand, single-molecule techniques such as sm-FRET provide a direct way to observe distributions of the related parameters within the ensemble. Both types of technique clearly demonstrate that IDPs exist as a very heterogeneous ensemble, with a significant presence of outlier conformations that might be of functional importance.

Functional Aspects of Intrinsic Disorder in Proteins

Coupled folding and binding: Specificity versus affinity

Protein–protein and protein–nucleic acid interactions play pivotal roles in the regulation of many molecular biology processes. When signaling is switched on, regulatory interactions occur in a specific manner. When it is switched off, the complexes quickly dissociate. In many interactions, at least one of the partners is disordered in the free form, but undergoes a disorder-to-order transition upon binding. As first proposed by Dyson and Wright,[12] large conformational entropy penalty paid for disorder-to-order transitions couples the high specificity of these interactions with low affinities. This is an example of the enthalpy–entropy compensation principle.[5] Notably, IDPs expose a larger interface area for binding than folded proteins of the same size, thus providing the basis for the high specificity of their interactions.[40] As mentioned above, this combination of high specificity and low affinity is exploited in regulatory processes. The kinetics of protein–protein interactions is also influenced by coupled folding and binding. IDPs might have a bigger capture radius than a compact, folded protein, as suggested by Shoemaker et al.[164] The unstructured protein binds weakly to its target at relatively large distances, but folding upon binding will bring the target into the final binding site: a "fly-casting" mechanism. This facilitates the diffusive search for the target and increases the association rate constant without necessarily affecting the equilibrium constant. This has been experimentally shown to occur in the binding of the intrinsically disordered inhibitor IA3 to the yeast protease YPrA.[165] In addition, a simulation study has suggested that complex formation between the Ets domain of SAP-1 and its specific DNA sequence is promoted by this mechanism.[166]

Computational studies have shown that IDPs often contain a conformational preference for the structure they will take upon binding.[41] This is observed as transient local and long-range structure in the free form of IDPs. These preformed elements are often α-helices. The presence of preformed elements in the free form facilitates binding-induced folding, but, limits their previous kinetic/thermodynamic advantages. Recent studies have shown that the presence of preformed elements is not a general feature of induced folding, but at least some disordered regions display fully template-dependent folding.[49]

Several recent observations suggest that IDPs might not be fully ordered in the bound state.[54] It has been shown for Sic1 binding to Cdc4,[72] inhibitor 2 binding to PP1c,[167] and UPF2 binding to UPF1[168] that the IDP also has a structural ambiguity in the bound state. The existence of protein disorder in the so-called fuzzy complexes may be important for function.[13] A very interesting example is presented by the Sic1 protein, which has a central role in cell-cycle regulation.[72] After phosphorylation at exactly six sites, this IDP gains the ability to bind Cdc4, which has only one binding pocket for a phosphate group. The Sic1–Cdc4 complex is highly dynamic and fuzzy,

thus making it possible for each of the six phosphate groups in Sic1 to occupy the binding pocket one after the other.

In addition to hetero-association with a binding partner, induced folding also occurs through the self-association of IDPs to form aggregates. As with binding, the large conformational penalty from folding upon aggregation uncouples the remarkable specificity of protein aggregation from its affinity. The resemblance between binding and aggregation extends further to the presence of preformed elements in monomeric IDPs. This has been demonstrated, for example, for Aβ,[119] α-synuclein,[87] and tau.[160] Despite a large conformational ordering during aggregation, it is believed that IDPs keep part of their structural ambiguity within the aggregates, as recently highlighted for filaments of the tau protein.[160] The fuzziness of aggregate structures is likely to be more prominent in the early-forming oligomers and play a role in their higher cytotoxicity.

Regulating protein levels of IDPs

It is well established that most disordered proteins are present in cells at relatively low levels compared to those of well-ordered proteins.[169] This is due to the shorter half lives of IDPs, which in turn are connected to their increased susceptibility to degradation by proteases and the proteasome system. The remarkable sensitivity of IDPs to proteases is a specific example of their higher propensity for post-translational modifications. This is because the chemically modified sites of IDPs are better displayed to the modifying enzymes.[8] Moreover, disordered proteins possess a greater proportion of ubiquitylation sites than structured proteins, thus making them accessible targets for ubiquitin-dependent proteasome systems.[170] This occurs, for example, in the IDPs securin and cyclin B.[171] The IDPs are also more vulnerable to ubiquitin-independent proteasome activity, as shown, for example, for tau[172] and $p21^{Cip1}$.[173] The rapid turnover of IDPs provides a functional advantage for those that are involved in cellular signaling and transcriptional processes. In addition to IDPs' higher susceptibility to protein degradation, the expressed level of IDPs is also regulated at the post-transcriptional level. A higher proportion of miRNA target sites is found in the transcripts that encode IDPs, and the mRNA decay rate is also higher for IDPs.[170] This phenomenon seems to be evolutionarily conserved, thus suggesting that fine tuning of the expression level of IDPs in a timely manner is crucial for the cell. On the other hand, the misregulation of the level of IDPs can result in pathological states, for example enhanced expression of α-synuclein can contribute to the pathogenesis of Parkinson's disease.[174]

Whereas the majority of IDPs have a low basal availability inside the cell, some of them can be required at high levels or for long times, for example during specific phases of the cell cycle. The susceptibility of IDPs to degradation pathways should then be tuned in response to specific circumstances. This can be achieved by incorporation into stable protein complexes (e.g., ribosomal proteins), or binding of the so-called nanny proteins to the disordered segments that facilitate degradation.[175] The latter mechanism is exemplified in the tumor-suppressor protein p53.[176] Post-translational modifications of IDPs can also modulate their binding affinity and/or kinetics, thereby modulating their accessibility to degradation mechanisms.

Intrinsic Disorder and Drug Discovery

The role of intrinsically disordered proteins in disease

Cellular housekeeping roles such as enzymatic catalysis and transport processes are often played by globular proteins. Disordered proteins are more closely linked to regulation, signal transduction and recognition processes in which their ability to bind multiple partners, along with their characteristic high-specificity/low-affinity interactions, plays an advantageous role. As a consequence, structural disorder is enriched in proteins that are involved in diseases such as cancer, cardiovascular disease, diabetes and neurodegenerative diseases. This has been summarized in the D^2 (disorder in disorders) concept.[9] A 2002 study showed that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted ID regions of 30 residues or more.[177] For cardiovascular-disease-related proteins, the abundance of long disordered regions was found to be 61%. Similar abundances were also found for the proteins associated with neurodegenerative diseases and diabetes,[9] and this high prevalence is without including short disordered regions.

A remarkable example of an IDP linked to signaling and cancer is p53. The p53 molecule has three domains, the N-terminal transcription-activation domain, the middle DNA binding domain (DBD), and the C-terminal oligomerization domain. The DBD is structured, but the terminal domains are intrinsically unstructured.[178] p53 plays a central role in signaling networks, regulating the expression of the genes involved in cell-cycle progression, apoptosis induction, DNA repair, and cellular response to stress.[179] Loss of p53 function often leads to malignancy in cells.[180] Among the IDPs involved in neurodegenerative diseases, the prion protein plays the key role in transmissible spongiform encephalopathies (TSE).[181] Aβ and tau proteins are at the center of Alzheimer's pathogenesis,[182,183] while α-synuclein and tau are central to synucleinopathies including Parkinson's disease.[184,185] With regard to the last example, it shold be mentioned that a recent publication suggests that α-synuclein is not disordered in vivo, but exists as an α-helical tetramer.[186] The results of this study were, however, questioned by a more recent publication that showed that human and rodent α-synucleins expressed in the nervous system exist predominantly as an unfolded monomer.[187] In vivo, lipids, detergents or other modifiers bound to α-synuclein can shift the dynamic equilibrium to more helical and more oligomeric species depending on the local concentration and environment. The different α-synuclein species are expected to be important for aggregation and α-synuclein-related neurotoxicity.[188]

Why is structural disorder more abundant in disease-associated proteins? It has been shown that protein–protein interaction networks have "small-world" architecture, with a few proteins having many partners (acting like hubs) and many proteins having only a few partners. In such networks, the typical distance between two randomly chosen nodes is small, that is,

they are connected through a small number of edges. This interaction network facilitates the cross-talk between proteins and the transfer of information through the network. The loss of hub proteins is more likely to cause cellular death.[189] Protein disorder facilitates binding to multiple patterns, and has been shown to be especially common within the hub proteins.[14] The remarkable plasticity of binding is represented by the IDP Cdk inhibitor $p21^{Cip1}$, which interacts with CycA–Cdk2, CycE–Cdk2, CycD–Cdk4 complexes,[190] the Rho kinase,[191] and apoptosis-signal-regulating kinase.[192] Strikingly, interactions with different partners can be mediated by the same region of the IDP, as was shown for the C terminus of p53, which adopts different conformations in complex with different partners.[49]

IDPs as targets for small-molecule drugs

The ID regions of proteins play key roles in pathologically crucial protein–protein interactions. These interactions are therefore attractive targets for drug-development projects. There are several strategies to interfere with these intermolecular interactions. One approach is to block the binding site on the ordered binding partner, which most often has pockets to accommodate the relatively short fragments of IDPs that are involved in the interaction. Such cavities are promising targets for small molecules. These small-molecule inhibitors of protein–protein interactions can be found through random screening of compound libraries, or if a 3D structure of the partner is available, by rational drug-design approaches. Inhibition of the p53 interaction with Mdm2 exemplifies this approach.[193] The unstructured N-terminal transcription-activation domain of p53 interacts with Mdm2 through a short stretch of residues (13–29). NMR studies have shown that, in the unbound state, the N terminus of p53, including residues 13–29, lacks stable structure and is highly flexible, but transiently takes on helical conformation.[42,178] In complex with Mdm2, this region of p53 forms an amphipathic α-helix that binds into a deep groove on the surface of Mdm2.[194] As a consequence of binding to Mdm2, p53 goes silent in terms of transcription regulation, undergoes ubiquitylation, which directs it toward degradation, and through a nuclear export signal present in Mdm2, is transported out of the nucleus. Interruption of p53–Mdm2 binding can restore p53 function and lead to the induction of apoptosis in the hosting cancerous cells. Several peptides as well as small molecules have been created to mimic the α-MoRF of the N terminus of p53 and inhibit its binding to Mdm2.[195] The success of these trials promises to bring a new drug-design strategy, that is, the detection of MoRFs in the IDPs, identification of their binding sites over structured proteins, structure determination for the complex, and peptide or small-molecule design to inhibit protein–protein interaction.[196]

The alternative approach is to target the IDPs themselves. This approach is even more fascinating than the previous one because, considering the promiscuity of IDP interactions, it might be feasible to block multiple protein–protein interactions with a single small molecule. When a protein binds to small molecules, the disorder-to-order transition of the involved region is inhibited. This inhibition can be caused by direct promotion of the disordered state or by its misfolding. Several small-molecule inhibitors have been found against clinically attractive disordered targets, like the oncoproteins c-Myc[197] and EWS-FLI1[198] or the Alzheimer's disease-related peptide Aβ.[56,199] The c-Myc protein is overexpressed in most human cancers. In order to be active, c-Myc has to heterodimerize with its partner, Max. Both c-Myc and Max are disordered in isolation, but upon binding they mutually fold to form a helical coiled coil. A systematic screening for small-molecule inhibitors of c-Myc–Max interactions has led to the discovery of several specific inhibitors. It is worth noting that the interaction sites of c-Myc have been located at three short stretches of residues: 366–375, 375–385 and 402–409.[197]

Aβ is generated through proteolysis of the amyloid precursor protein (APP). APP is first cleaved by β-secretase to yield the 99-residue C-terminal domain (C99), which forms a substrate for a second enzyme, called γ-secretase, to produce Aβ peptides of various lengths and with various aggregation properties.[200] Small molecules that modulate γ-secretase activity (GSMs) are of therapeutic interest; however, their clinical application is limited by the fact that γ-secretase is involved in the processing of several other physiologically important substrates. A viable way around this is to target the disordered substrate C99 (or its part, Aβ), rather than the enzyme itself. Indeed, it has been suggested that certain GSMs, namely fenofibrate, tarenflurbil and sulindac sulfide, bind to C99/Aβ in a specific manner.[199,201] However, further studies demonstrated that the interaction of these GSMs with the C99/Aβ was nonspecific.[202,203] The contrasting results were attributed to the presence of protein and/or GSM aggregates, which provide platforms for their nonspecific associations. Since both IDPs and small molecules are frequently prone to aggregation, it is important to take critical consideration of the assembly state of both partners in small-molecule–IDP interaction studies.

As fibrillar aggregation is a key event in the pathology of neurodegenerative diseases, tremendous effort has been made to develop anti-aggregation drugs. Several small molecules have been found to block amyloid aggregation, with some of them proven to protect against the toxicity of aggregates. Polyaromatic compounds such as phthalocyanines tetrasulfonated (PcTS) show a strong anti-amyloidogenic effect. As an inhibitor of α-synuclein aggregation, PcTS binds to a specific site in the N terminus of α-synuclein (Tyr39 is the anchoring residue; Figure 6).[204] NMR and electronic absorption studies have demonstrated that the binding capacity and anti-amyloid activity of phthalocyanines are determined by their relative ability to self-stack through π–π interactions.[205] In another study, the catechol-containing compounds entacapone and tolcapone were found to be potent inhibitors of oligomerization and fibrillogenesis of both Aβ and α-synuclein.[206] The interaction between an anti-aggregation small molecule, Congo red, and α-synuclein has been recently characterized by NMR spectroscopy and scattering techniques.[207] It was shown that, even at the low concentrations used, Congo red existed as micelles. Although only a small fraction of free monomeric α-synuclein (~2%) was bound to Congo red micelles, the relatively fast

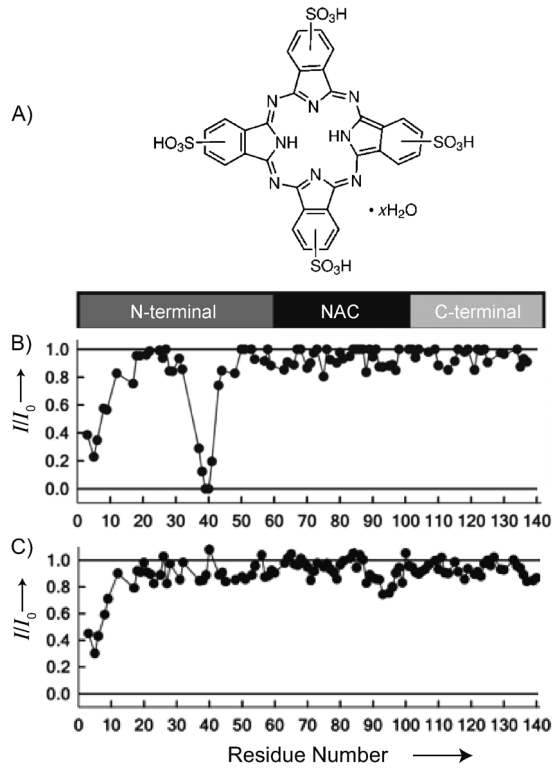


Figure 6. NMR investigation of PcTS (A) binding to α-synuclein that reveals the role of Tyr39 in binding. B) The intensity profile of wild-type α-synuclein shows a severe broadening at residues 37–41 after PcTS is added. C) Upon mutation of Tyr39 to Ala, no signal broadening is observed in this region. The aromatic residue Tyr39 plays a key role in α-synuclein aggregation, and its targeting by PcTS inhibits amyloid fibrillar formation. Reprinted with permission from ref. [204]. Copyright: The National Academy of Sciences, 2009.

exchange between the free and Congo red-bound states of α-synuclein led to strong attenuation of its NMR signals.

Another approach to inhibiting aggregation is to redesign the self-recognition interface of proteins in such a manner that the modified molecule is still capable of interacting with the native form, but its further assembly into larger aggregates is inhibited. This approach was successfully exploited to block diabetes-related islet amyloid polypeptide (IAPP) aggregation by creating an IAPP analogue that contained two N-methyl groups on the same side of the β-strand in the amyloid core while keeping the NH groups of the other side.[208] Interestingly, this IAPP analogue was also proven to be effective at blocking cytotoxic assembly of Aβ.[209]

The binding of IDPs to small molecules seems to have the same properties as their interaction with other proteins, in particular, a combination of high specificity with low affinity.[56] The low affinity is a serious drawback for the clinical application of possible drugs because significant effects will be observed only at very high concentrations at which the occurrence of nonspecific interactions will be problematic. Interestingly, the plasticity of IDP binding to small molecules might open a way to circumvent this problem. As was shown in the case of c-Myc, three close short regions are involved in binding to different small molecules;[197] this suggests that the use of bivalent molecules could improve binding affinities.

Summary and Outlook

The last decade has witnessed a sharp surge in the number of experimentally identified intrinsically disordered proteins and regions, along with a steady advance in the accuracy of tools that predict protein disorder. It is now well established that protein disorder is not an eccentric feature of the protein universe, but is present all over the eukaryotic proteomes. It has been convincingly demonstrated in a large number of examples that protein disorder is indeed beneficial to protein function. Novel experimental methods, especially in NMR spectroscopy and SAXS, have been devised and successfully employed to characterize IDPs. In parallel, tremendous effort has been devoted to developing computational tools to transform experimental data into conformational ensemble descriptions of IDPs. Although serious challenges are still left, the hope of generating atomic-scale ensemble models for IDPs is no longer unrealistic. To this end, single-molecule measurements will play an essential role. With a better understanding of the conformational properties of IDPs and resultant functional insights, drug discovery against clinically related IDPs will be based on more-rational strategies.

Acknowledgements

N.R.-G. thanks Francesca Munari for critical reading of the manuscript. M.B. would like to acknowledge the Agence National de Recherche for financial support from TAUSTRUCT-ANR MALZ 2010. This work was supported by the Max Planck Society and the DFG (ZW 71/2-2 and 3-2).

[1] G. A. Papoian, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14237 – 14238.
[2] K. A. Dill, *Protein Sci.* **1999**, *8*, 1166 – 1180.
[3] N. Rezaei-Ghaleh, M. Zweckstetter, *eLS*, Wiley, Chichester, **2011**.
[4] C. Boesch, A. Bundi, M. Oppliger, K. Wüthrich, *Eur. J. Biochem.* **1978**, *91*, 209 – 214.
[5] H. J. Dyson, P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197 – 208.
[6] M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, A. K. Dunker, *Nucleic Acids Res.* **2007**, *35*, D786 – 793.
[7] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, A. K. Dunker, *Biochemistry* **2005**, *44*, 1989 – 2000.
[8] P. Tompa, *FEBS Lett.* **2005**, *579*, 3346 – 3354.
[9] V. N. Uversky, C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biophys. Bioeng.* **2008**, *37*, 215 – 246.
[10] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321 – 331.
[11] P. E. Wright, H. J. Dyson, *Curr. Opin. Struct. Biol.* **2009**, *19*, 31 – 38.
[12] H. J. Dyson, P. E. Wright, *Curr. Opin. Struct. Biol.* **2002**, *12*, 54 – 60.
[13] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2 – 8.
[14] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, V. N. Uversky, *FEBS J.* **2005**, *272*, 5129 – 5148.
[15] L. de La Cruz, R. Bajaj, S. Becker, M. Zweckstetter, *Protein Sci.* **2010**, *19*, 2045 – 2054.
[16] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, *J. Mol. Graph. Model.* **2001**, *19*, 26 – 59.
[17] V. N. Uversky, *Protein Sci.* **2002**, *11*, 739 – 756.
[18] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, A. K. Dunker, *Biophys. J.* **2007**, *92*, 1439 – 1456.
[19] V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins Struct. Funct. Bioinf. Proteins* **2000**, *41*, 415 – 427.

[20] P. Radivojac, Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, A. K. Dunker, *Protein Sci.* **2004**, *13*, 71 – 80.
[21] S. Lise, D. T. Jones, *Proteins Struct. Funct. Bioinf.* **2005**, *58*, 144 – 150.
[22] J. C. Wootton, *Comput. Chem.* **1993**, *17*, 149 – 163.
[23] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, A. K. Dunker, *Proteins Struct. Funct. Bioinf.* **2001**, *42*, 38 – 48.
[24] L. D. Hurst, *Trends Genet.* **2002**, *18*, 486.
[25] L. S. Whitfield, R. Lovell-Badge, P. N. Goodfellow, *Nature* **1993**, *364*, 713 – 715.
[26] P. Tompa, *Bioessays* **2003**, *25*, 847 – 855.
[27] Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, *Proteins Struct. Funct. Bioinf.* **2003**, *53 Suppl. 6*, 566 – 572.
[28] K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker, Z. Obradovic, *BMC Bioinform.* **2006**, *7*, 208.
[29] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, R. B. Russell, *Structure* **2003**, *11*, 1453 – 1459.
[30] J. Prilusky, C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, J. L. Sussman, *Bioinformatics* **2005**, *21*, 3435 – 3438.
[31] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, D. T. Jones, *Bioinformatics* **2004**, *20*, 2138 – 2139.
[32] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, *J. Mol. Biol.* **2004**, *337*, 635 – 645.
[33] R. Linding, R. B. Russell, V. Neduva, T. J. Gibson, *Nucleic Acids Res.* **2003**, *31*, 3701 – 3708.
[34] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, *J. Mol. Biol.* **2005**, *347*, 827 – 839.
[35] J. Liu, B. Rost, *Nucleic Acids Res.* **2003**, *31*, 3833 – 3835.
[36] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, A. K. Dunker, *Cell. Res.* **2009**, *19*, 929 – 949.
[37] T. Ishida, K. Kinoshita, *Bioinformatics* **2008**, *24*, 1344 – 1348.
[38] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 996 – 1010.
[39] B. Xue, R. W. Williams, C. J. Oldfield, A. K. Dunker, V. N. Uversky, *BMC Syst. Biol.* **2010**, *4 Suppl. 1*, S1.
[40] A. L. Fink, *Curr. Opin. Struct. Biol.* **2005**, *15*, 35 – 41.
[41] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, *J. Mol. Biol.* **2004**, *338*, 1015 – 1026.
[42] H. Lee, K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi, K. H. Han, *J. Biol. Chem.* **2000**, *275*, 29426 – 29432.
[43] E. R. Lacy, I. Filippov, W. S. Lewis, S. Otieno, L. Xiao, S. Weiss, L. Hengst, R. W. Kriwacki, *Nat. Struct. Mol. Biol.* **2004**, *11*, 358 – 364.
[44] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, *J. Mol. Biol.* **2006**, *362*, 1043 – 1059.
[45] C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, A. K. Dunker, *Biochemistry* **2005**, *44*, 12454 – 12470.
[46] E. Garner, P. Romero, A. K. Dunker, C. Brown, Z. Obradovic, *Genome Inf. Ser.* **1999**, *10*, 41 – 50.
[47] C. M. Fletcher, G. Wagner, *Protein Sci.* **1998**, *7*, 1639 – 1642.
[48] S. Mader, H. Lee, A. Pause, N. Sonenberg, *Mol. Cell. Biol.* **1995**, *15*, 4990 – 4997.
[49] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, *BMC Genomics* **2008**, *9 Suppl. 1*, S1.
[50] V. Vacic, C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, V. N. Uversky, A. K. Dunker, *J. Proteome Res.* **2007**, *6*, 2351 – 2366.
[51] Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky, A. K. Dunker, *Biochemistry* **2007**, *46*, 13468 – 13477.
[52] B. Meszaros, I. Simon, Z. Dosztanyi, *PLoS Comput. Biol.* **2009**, *5*, e1000376.
[53] C. M. Gould, F. Diella, A. Via, P. Puntervoll, C. Gemund, S. Chabanis-Davidson, S. Michael, A. Sayadi, J. C. Bryne, C. Chica, M. Seiler, N. E. Davey, N. Haslam, R. J. Weatheritt, A. Budd, T. Hughes, J. Pas, L. Rychlewski, G. Trave, R. Aasland, M. Helmer-Citterich, R. Linding, T. J. Gibson, *Nucleic Acids Res.* **2010**, *38*, D167 – 180.
[54] P. Tompa, *Curr. Opin. Struct. Biol.* **2011**, *21*, 419 – 425.
[55] N. E. Davey, N. J. Haslam, D. C. Shields, R. J. Edwards, *Nucleic Acids Res.* **2010**, *38*, W534 – 539.
[56] S. J. Metallo, *Curr. Opin. Chem. Biol.* **2010**, *14*, 481 – 488.
[57] D. Eliezer, *Curr. Opin. Struct. Biol.* **2009**, *19*, 23 – 30.
[58] L. Giehm, D. I. Svergun, D. E. Otzen, B. Vestergaard, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3246 – 3251.
[59] D. A. Jacques, J. Trewhella, *Protein Sci.* **2010**, *19*, 642 – 657.
[60] H. D. Mertens, D. I. Svergun, *J. Struct. Biol.* **2010**, *172*, 128 – 141.
[61] P. Bernado, D. I. Svergun, *Mol. Biosyst.* **2012**, *8*, 151 – 167.
[62] D. I. Pretto, S. Tsutakawa, C. A. Brosey, A. Castillo, M. E. Chagot, J. A. Smith, J. A. Tainer, W. J. Chazin, *Biochemistry* **2010**, *49*, 2880 – 2889.
[63] P. Bernado, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, *J. Am. Chem. Soc.* **2007**, *129*, 5656 – 5664.
[64] S. Yang, L. Blachowicz, L. Makowski, B. Roux, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15757 – 15762.
[65] B. Rozycki, Y. C. Kim, G. Hummer, *Structure* **2011**, *19*, 109 – 116.
[66] C. Leyrat, M. R. Jensen, E. A. Ribeiro, Jr., F. C. Gerard, R. W. Ruigrok, M. Blackledge, M. Jamin, *Protein Sci.* **2011**, *20*, 542 – 556.
[67] A. Paz, T. Zeev-Ben-Mordehai, M. Lundqvist, E. Sherman, E. Mylonas, L. Weiner, G. Haran, D. I. Svergun, F. A. Mulder, J. L. Sussman, I. Silman, *Biophys. J.* **2008**, *95*, 1928 – 1944.
[68] H. Boze, T. Marlin, D. Durand, J. Perez, A. Vernhet, F. Canon, P. Sarni-Manchado, V. Cheynier, B. Cabane, *Biophys. J.* **2010**, *99*, 656 – 665.
[69] K. Stott, M. Watson, F. S. Howe, J. G. Grossmann, J. O. Thomas, *J. Mol. Biol.* **2010**, *403*, 706 – 722.
[70] J. A. Marsh, C. Neale, F. E. Jack, W. Y. Choy, A. Y. Lee, K. A. Crowhurst, J. D. Forman-Kay, *J. Mol. Biol.* **2007**, *367*, 1494 – 1510.
[71] J. A. Marsh, J. D. Forman-Kay, *J. Mol. Biol.* **2009**, *391*, 359 – 374.
[72] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J. D. Forman-Kay, *Structure* **2010**, *18*, 494 – 506.
[73] H. J. Dyson, P. E. Wright, *Chem. Rev.* **2004**, *104*, 3607 – 3622.
[74] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93 – 158.
[75] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2001**, *123*, 9490 – 9491.
[76] M. D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, *PLoS Biol.* **2009**, *7*, e34.
[77] S. Hiller, C. Wasmer, G. Wider, K. Wüthrich, *J. Am. Chem. Soc.* **2007**, *129*, 10823 – 10828.
[78] R. L. Narayanan, U. H. Durr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 11906 – 11907.
[79] J. Novacek, A. Zawadzka-Kazimierczuk, V. Papouskova, L. Zidek, H. Sanderova, L. Krasny, W. Kozminski, V. Sklenar, *J. Biomol. NMR* **2011**, *50*, 1 – 11.
[80] L. Skora, S. Becker, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 9223 – 9225.
[81] W. Bermel, I. Bertini, I. C. Felli, M. Piccioli, R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25 – 45.
[82] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171 – 180.
[83] T. Mittag, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2007**, *17*, 3 – 14.
[84] J. Yao, J. Chung, D. Eliezer, P. E. Wright, H. J. Dyson, *Biochemistry* **2001**, *40*, 3561 – 3571.
[85] F. Sziegat, J. Wirmer-Bartoschek, H. Schwalbe, *Angew. Chem.* **2011**, *123*, 5628 – 5632; *Angew. Chem. Int. Ed.* **2011**, *50*, 5514 – 5518.
[86] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795 – 2804.
[87] H. Y. Kim, H. Heise, C. O. Fernandez, M. Baldus, M. Zweckstetter, *ChemBioChem* **2007**, *8*, 1671 – 1674.
[88] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772 – 7777.
[89] P. Permi, I. Kilpelainen, A. Annila, S. Heikkinen, *J. Biomol. NMR* **2000**, *16*, 29 – 37.
[90] L. J. Smith, K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton, C. M. Dobson, *J. Mol. Biol.* **1996**, *255*, 494 – 506.
[91] K. Tamiola, B. Acar, F. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 18000 – 18003.
[92] M. Kjaergaard, S. Brander, F. M. Poulsen, *J. Biomol. NMR* **2011**, *49*, 139 – 149.
[93] S. Schwarzinger, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, H. J. Dyson, *J. Am. Chem. Soc.* **2001**, *123*, 2970 – 2978.
[94] L. Wang, H. R. Eghbalnia, A. Bahrami, J. L. Markley, *J. Biomol. NMR* **2005**, *32*, 13 – 22.
[95] K. W. Plaxco, C. J. Morton, S. B. Grimshaw, J. A. Jones, M. Pitkeathly, I. D. Campbell, C. M. Dobson, *J. Biomol. NMR* **1997**, *10*, 221 – 230.
[96] D. S. Wishart, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 62 – 87.
[97] K. A. Crowhurst, J. D. Forman-Kay, *Biochemistry* **2003**, *42*, 8687 – 8695.

[98] M. R. Jensen, P. R. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernado, M. Blackledge, *Structure* **2009**, *17*, 1169 – 1185.
[99] J. A. Marsh, J. M. Baker, M. Tollinger, J. D. Forman-Kay, *J. Am. Chem. Soc.* **2008**, *130*, 7804 – 7805.
[100] O. I. Obolensky, K. Schlepckow, H. Schwalbe, A. V. Solov'yov, *J. Biomol. NMR* **2007**, *39*, 1 – 16.
[101] W. Fieber, S. Kristjansdottir, F. M. Poulsen, *J. Mol. Biol.* **2004**, *339*, 1191 – 1199.
[102] R. Mohana-Borges, N. K. Goto, G. J. Kroon, H. J. Dyson, P. E. Wright, *J. Mol. Biol.* **2004**, *340*, 1131 – 1142.
[103] M. K. Cho, H. Y. Kim, P. Bernado, C. O. Fernandez, M. Blackledge, M. Zweckstetter, *J. Am. Chem. Soc.* **2007**, *129*, 3032 – 3033.
[104] A. K. Jha, A. Colubri, K. F. Freed, T. R. Sosnick, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13099 – 13104.
[105] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 3791 – 3792.
[106] P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002 – 17007.
[107] P. Bernado, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2005**, *127*, 17968 – 17969.
[108] M. D. Mukrasch, P. Markwick, J. Biernat, M. Bergen, P. Bernado, C. Griesinger, E. Mandelkow, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2007**, *129*, 5235 – 5243.
[109] Y. Xue, I. S. Podkorytov, D. K. Rao, N. Benjamin, H. Sun, N. R. Skrynnikov, *Protein Sci.* **2009**, *18*, 1401 – 1424.
[110] K. Teilum, B. B. Kragelund, F. M. Poulsen, *J. Mol. Biol.* **2002**, *324*, 349 – 357.
[111] S. Bibow, V. Ozenne, J. Biernat, M. Blackledge, E. Mandelkow, M. Zweckstetter, *J. Am. Chem. Soc.* **2011**, *133*, 15842 – 15845.
[112] D. Ganguly, J. Chen, *J. Mol. Biol.* **2009**, *390*, 467 – 477.
[113] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407 – 8418.
[114] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546 – 4559.
[115] A. G. Palmer III, *Chem. Rev.* **2004**, *104*, 3623 – 3640.
[116] H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, C. M. Dobson, *Biochemistry* **1997**, *36*, 8977 – 8991.
[117] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, H. Schwalbe, *Science* **2002**, *295*, 1719 – 1722.
[118] K. E. Paleologou, A. W. Schmid, C. C. Rospigliosi, H. Y. Kim, G. R. Lamberto, R. A. Fredenburg, P. T. Lansbury, Jr., C. O. Fernandez, D. Eliezer, M. Zweckstetter, H. A. Lashuel, *J. Biol. Chem.* **2008**, *283*, 16895 – 16905.
[119] N. Rezaei-Ghaleh, K. Giller, S. Becker, M. Zweckstetter, *Biophys. J.* **2011**, *101*, 1202 – 1211.
[120] D. Yang, A. Mittermaier, Y. K. Mok, L. E. Kay, *J. Mol. Biol.* **1998**, *276*, 939 – 954.
[121] D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, L. E. Kay, *Science* **2010**, *329*, 1312 – 1316.
[122] R. Schneider, J. R. Huang, M. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, M. Ringkjobing Jensen, M. Blackledge, *Mol. Biosyst.* **2012**, *8*, 58 – 68.
[123] C. K. Fisher, C. M. Stultz, *Curr. Opin. Struct. Biol.* **2011**, *21*, 426 – 431.
[124] S. Neal, A. M. Nip, H. Zhang, D. S. Wishart, *J. Biomol. NMR* **2003**, *26*, 215 – 240.
[125] X. P. Xu, D. A. Case, *J. Biomol. NMR* **2001**, *21*, 321 – 333.
[126] Y. Shen, A. Bax, *J. Biomol. NMR* **2007**, *38*, 289 – 302.
[127] K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, *J. Am. Chem. Soc.* **2009**, *131*, 13894 – 13895.
[128] S. Esteban-Martin, R. B. Fenwick, X. Salvatella, *J. Am. Chem. Soc.* **2010**, *132*, 4626 – 4632.
[129] J. R. Huang, S. Grzesiek, *J. Am. Chem. Soc.* **2010**, *132*, 694 – 705.
[130] S. Meier, M. Strohmeier, M. Blackledge, S. Grzesiek, *J. Am. Chem. Soc.* **2007**, *129*, 754 – 755.
[131] J. R. Allison, P. Varnai, C. M. Dobson, M. Vendruscolo, *J. Am. Chem. Soc.* **2009**, *131*, 18314 – 18326.
[132] J. C. Lee, H. B. Gray, J. R. Winkler, *J. Am. Chem. Soc.* **2005**, *127*, 16388 – 16389.
[133] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141 – 151.
[134] L. Y. Chen, N. J. Horing, *J. Chem. Phys.* **2007**, *126*, 224103.
[135] A. Huang, C. M. Stultz, *PLoS Comput. Biol.* **2008**, *4*, e1000155.
[136] G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2009**, *131*, 17908 – 17918.
[137] C. W. Bertoncini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, M. Zweckstetter, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1430 – 1435.
[138] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. Ruigrok, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 8055 – 8061.
[139] M. R. Jensen, L. Salmon, G. Nodet, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 1270 – 1272.
[140] A. C. Ferreon, A. A. Deniz, *Biochim. Biophys. Acta Proteins Proteomics* **2011**, *1814*, 1021 – 1029.
[141] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2649 – 2654.
[142] A. C. Ferreon, Y. Gambin, E. A. Lemke, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5645 – 5650.
[143] F. Huang, S. Rajagopalan, G. Settanni, R. J. Marsh, D. A. Armoogum, N. Nicolaou, A. J. Bain, E. Lerner, E. Haas, L. Ying, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20758 – 20763.
[144] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12491 – 12496.
[145] M. C. Stumpe, H. Grubmüller, *J. Am. Chem. Soc.* **2007**, *129*, 16126 – 16131.
[146] S. Meier, S. Grzesiek, M. Blackledge, *J. Am. Chem. Soc.* **2007**, *129*, 9799 – 9807.
[147] P. Bernado, M. Blackledge, *Biophys. J.* **2009**, *97*, 2839 – 2845.
[148] G. Nemethy, I. Z. Steinberg, H. A. Scheraga, *Biopolymers* **1963**, *1*, 43 – 69.
[149] N. Rezaei-Ghaleh, E. Andreetto, L. M. Yan, A. Kapurniotu, M. Zweckstetter, *PLoS One* **2011**, *6*, e20289.
[150] O. Zhang, J. D. Forman-Kay, *Biochemistry* **1997**, *36*, 3959 – 3970.
[151] S. J. Whittington, B. W. Chellgren, V. M. Hermann, T. P. Creamer, *Biochemistry* **2005**, *44*, 6269 – 6275.
[152] J. Makowska, S. Rodziewicz-Motowidlo, K. Baginska, J. A. Vila, A. Liwo, L. Chmurzynski, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1744 – 1749.
[153] D. Neri, M. Billeter, G. Wider, K. Wüthrich, *Science* **1992**, *257*, 1559 – 1563.
[154] I. S. Millet, L. E. Townsley, F. Chiti, S. Doniach, K. W. Plaxco, *Biochemistry* **2002**, *41*, 321 – 325.
[155] J. R. Gillespie, D. Shortle, *J. Mol. Biol.* **1997**, *268*, 170 – 184.
[156] Y. K. Mok, C. M. Kay, L. E. Kay, J. Forman-Kay, *J. Mol. Biol.* **1999**, *289*, 619 – 638.
[157] C. J. Francis, K. Lindorff-Larsen, R. B. Best, M. Vendruscolo, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 145 – 152.
[158] G. Lippens, J. M. Wieruszeski, A. Leroy, C. Smet, A. Sillen, L. Buee, I. Landrieu, *ChemBioChem* **2004**, *5*, 73 – 78.
[159] I. Landrieu, L. Lacosse, A. Leroy, J. M. Wieruszeski, X. Trivelli, A. Sillen, N. Sibille, H. Schwalbe, K. Saxena, T. Langer, G. Lippens, *J. Am. Chem. Soc.* **2006**, *128*, 3575 – 3583.
[160] S. Bibow, M. D. Mukrasch, S. Chinnathambi, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, *Angew. Chem.* **2011**, *123*, 11723 – 11727; *Angew. Chem. Int. Ed.* **2011**, *50*, 11520 – 11524.
[161] A. J. Baldwin, L. E. Kay, *Nat. Chem. Biol.* **2009**, *5*, 808 – 814.
[162] O. F. Lange, N. A. Lakomek, C. Farès, G. F. Schröder, K. F. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B. L. de Groot, *Science* **2008**, *320*, 1471 – 1475.
[163] B. Vogeli, T. F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren, R. Riek, *J. Am. Chem. Soc.* **2009**, *131*, 17215 – 17225.
[164] B. A. Shoemaker, J. J. Portman, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8868 – 8873.
[165] R. Narayanan, O. K. Ganesh, A. S. Edison, S. J. Hagen, *J. Am. Chem. Soc.* **2008**, *130*, 11477 – 11485.
[166] Y. Levy, J. N. Onuchic, P. G. Wolynes, *J. Am. Chem. Soc.* **2007**, *129*, 738 – 739.
[167] J. A. Marsh, B. Dancheck, M. J. Ragusa, M. Allaire, J. D. Forman-Kay, W. Peti, *Structure* **2010**, *18*, 1094 – 1103.
[168] M. Clerici, A. Mourao, I. Gutsche, N. H. Gehring, M. W. Hentze, A. Kulozik, J. Kadlec, M. Sattler, S. Cusack, *EMBO J.* **2009**, *28*, 2293 – 2306.

[169] M. M. Babu, R. van der Lee, N. S. de Groot, J. Gsponer, *Curr. Opin. Struct. Biol.* **2011**, *21*, 432 – 440.
[170] Y. J. Edwards, A. E. Lobley, M. M. Pentony, D. T. Jones, *Genome Biol.* **2009**, *10*, R50.
[171] C. J. Cox, K. Dutta, E. T. Petri, W. C. Hwang, Y. Lin, S. M. Pascal, R. Basavappa, *FEBS Lett.* **2002**, *527*, 303 – 308.
[172] D. C. David, R. Layfield, L. Serpell, Y. Narain, M. Goedert, M. G. Spillantini, *J. Neurochem.* **2002**, *83*, 176 – 185.
[173] R. J. Sheaff, J. D. Singer, J. Swanger, M. Smitherman, J. M. Roberts, B. E. Clurman, *Mol. Cell* **2000**, *5*, 403 – 410.
[174] A. L. McCormack, D. A. Di Monte, *Curr. Protein Pept. Sci.* **2009**, *10*, 476 – 482.
[175] P. Tsvetkov, N. Reuven, Y. Shaul, *Nat. Chem. Biol.* **2009**, *5*, 778 – 781.
[176] P. Tsvetkov, N. Reuven, C. Prives, Y. Shaul, *J. Biol. Chem.* **2009**, *284*, 26234 – 26242.
[177] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradović, A. K. Dunker, *J. Mol. Biol.* **2002**, *323*, 573 – 584.
[178] R. Dawson, L. Müller, A. Dehner, C. Klein, H. Kessler, J. Buchner, *J. Mol. Biol.* **2003**, *332*, 1131 – 1141.
[179] K. H. Vousden, X. Lu, *Nat. Rev. Cancer* **2002**, *2*, 594 – 604.
[180] M. Hollstein, D. Sidransky, B. Vogelstein, C. C. Harris, *Science* **1991**, *253*, 49 – 53.
[181] R. Riek, S. Hornemann, G. Wider, R. Glockshuber, K. Wüthrich, *FEBS Lett.* **1997**, *413*, 282 – 288.
[182] R. Riek, P. Guntert, H. Dobeli, B. Wipf, K. Wüthrich, *Eur. J. Biochem.* **2001**, *268*, 5930 – 5936.
[183] M. von Bergen, S. Barghorn, S. Jeganathan, E. M. Mandelkow, E. Mandelkow, *Neurodegener. Dis.* **2006**, *3*, 197 – 206.
[184] D. Eliezer, E. Kutluay, R. Bussell, Jr., G. Browne, *J. Mol. Biol.* **2001**, *307*, 1061 – 1073.
[185] N. Badiola, R. M. de Oliveira, F. Herrera, C. Guardia-Laguarta, S. A. Goncalves, M. Pera, M. Suarez-Calvet, J. Clarimon, T. F. Outeiro, A. Lleo, *PLoS One* **2011**, *6*, e26609.
[186] T. Bartels, J. G. Choi, D. J. Selkoe, *Nature* **2011**, *477*, 107 – 110.
[187] B. Fauvet, M. K. Mbefo, M.-B. Fares, C. Desobry, S. Michael, M. T. Ardah, E. Tsika, P. Coune, M. Prudent, N. Lion, D. Eliezer, D. J. Moore, B. Schneider, P. Aebischer, O. M. El-Agnaf, E. Masliah, H. A. Lashuel, *J. Biol. Chem.* **2012**; DOI: 10.1074/jbc.M111.318949.
[188] I. Dikiy, D. Eliezer, *Biochim. Biophys. Acta Biomembr.* **2012**, *1818*, 1013 – 1018.
[189] P. Bork, L. J. Jensen, C. von Mering, A. K. Ramani, I. Lee, E. M. Marcotte, *Curr. Opin. Struct. Biol.* **2004**, *14*, 292 – 299.
[190] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, P. E. Wright, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11504 – 11509.
[191] H. Tanaka, T. Yamashita, M. Asada, S. Mizutani, H. Yoshikawa, M. Tohyama, *J. Cell Biol.* **2002**, *158*, 321 – 329.
[192] M. Asada, T. Yamada, H. Ichijo, D. Delia, K. Miyazono, K. Fukumuro, S. Mizutani, *EMBO J.* **1999**, *18*, 1223 – 1234.
[193] L. T. Vassilev, *Cell Cycle* **2004**, *3*, 417 – 421.
[194] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, N. P. Pavletich, *Science* **1996**, *274*, 948 – 953.
[195] L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, E. A. Liu, *Science* **2004**, *303*, 844 – 848.
[196] A. K. Dunker, V. N. Uversky, *Curr. Opin. Pharmacol.* **2010**, *10*, 782 – 788.
[197] D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, *J. Am. Chem. Soc.* **2009**, *131*, 7390 – 7401.
[198] H. V. Erkizan, Y. Kong, M. Merchant, S. Schlottmann, J. S. Barber-Rotenberg, L. Yuan, O. D. Abaan, T. H. Chou, S. Dakshanamurthy, M. L. Brown, A. Uren, J. A. Toretsky, *Nat. Med.* **2009**, *15*, 750 – 756.
[199] T. L. Kukar, T. B. Ladd, M. A. Bann, P. C. Fraering, R. Narlawar, G. M. Maharvi, B. Healy, R. Chapman, A. T. Welzel, R. W. Price, B. Moore, V. Rangachari, B. Cusack, J. Eriksen, K. Jansen-West, C. Verbeeck, D. Yager, C. Eckman, W. Ye, S. Sagi, B. A. Cottrell, J. Torpey, T. L. Rosenberry, A. Fauq, M. S. Wolfe, B. Schmidt, D. M. Walsh, E. H. Koo, T. E. Golde, *Nature* **2008**, *453*, 925 – 929.
[200] R. E. Tanzi, L. Bertram, *Cell* **2005**, *120*, 545 – 555.
[201] L. Richter, L. M. Munter, J. Ness, P. W. Hildebrand, M. Dasari, S. Unterreitmeier, B. Bulic, M. Beyermann, R. Gust, B. Reif, S. Weggen, D. Langosch, G. Multhaup, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14597 – 14602.
[202] A. J. Beel, P. Barrett, P. D. Schnier, S. A. Hitchcock, D. Bagal, C. R. Sanders, J. B. Jordan, *Biochemistry* **2009**, *48*, 11837 – 11839.
[203] P. J. Barrett, C. R. Sanders, S. A. Kaufman, K. Michelsen, J. B. Jordan, *Biochemistry* **2011**, *50*, 10328 – 10342.
[204] G. R. Lamberto, A. Binolfi, M. L. Orcellet, C. W. Bertoncini, M. Zweckstetter, C. Griesinger, C. O. Fernandez, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21057 – 21062.
[205] G. R. Lamberto, V. Torres-Monserrat, C. W. Bertoncini, X. Salvatella, M. Zweckstetter, C. Griesinger, C. O. Fernandez, *J. Biol. Chem.* **2011**, *286*, 32036 – 32044.
[206] S. Di Giovanni, S. Eleuteri, K. E. Paleologou, G. Yin, M. Zweckstetter, P. A. Carrupt, H. A. Lashuel, *J. Biol. Chem.* **2010**, *285*, 14941 – 14954.
[207] A. S. Maltsev, A. Grishaev, A. Bax, *Biochemistry* **2012**, *51*, 631 – 642.
[208] L. M. Yan, M. Tatarek-Nossol, A. Velkova, A. Kazantzis, A. Kapurniotu, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2046 – 2051.
[209] L. M. Yan, A. Velkova, M. Tatarek-Nossol, E. Andreetto, A. Kapurniotu, *Angew. Chem.* **2007**, *119*, 1268 – 1274; *Angew. Chem. Int. Ed.* **2007**, *46*, 1246 – 1252.

Received: February 7, 2012
Published online on April 13, 2012

