Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy

Malene Ringkjøbing Jensen,[†,‡,§] Markus Zweckstetter,[∥,⊥,#] Jie-rong Huang,[†,‡,§] and Martin Blackledge*,[†,‡,§]

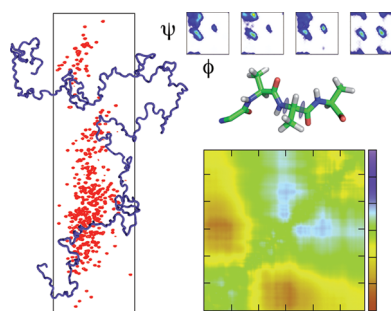[†]Université Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38027 Grenoble, France

[‡]CEA, DSV, IBS, F-38027 Grenoble, France

[§]CNRS, IBS, F-38027 Grenoble, France

[∥]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[⊥]German Center for Neurodegenerative Diseases (DZNE), 37077 Göttingen, Germany

[#]Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), University Medical Center, 37073 Göttingen, Germany

CONTENTS

1. INTRODUCTION

The last 15 years have seen a paradigm shift in our understanding of protein biochemistry, with the realization that an unexpectedly high fraction of the human genome codes for functional proteins, or domains of proteins that lack a stable tertiary fold.[1−4] Bioinformatic analyses of diverse proteomes predict that intrinsically disordered proteins (IDPs) are prevalent throughout living organisms, with particular abundance in eukaryotic proteomes where they play crucial roles in many biochemical processes, including signal transfer, regulation, transcription, and replication and in many human pathologies. Rather than coding for an energetically stable three-dimensional fold, associated with conformational fluctuations that allow for biological function,[5] the primary sequence of IDPs results in a much flatter free-energy surface, spanning multiple different conformations.[6] This structural flexibility allows IDPs to exploit functional mechanisms inaccessible to folded proteins. IDPs have been proposed to participate in highly specific but promiscuous interaction networks, exhibiting rapid dissociation rates, dynamic, or "fuzzy" interactions
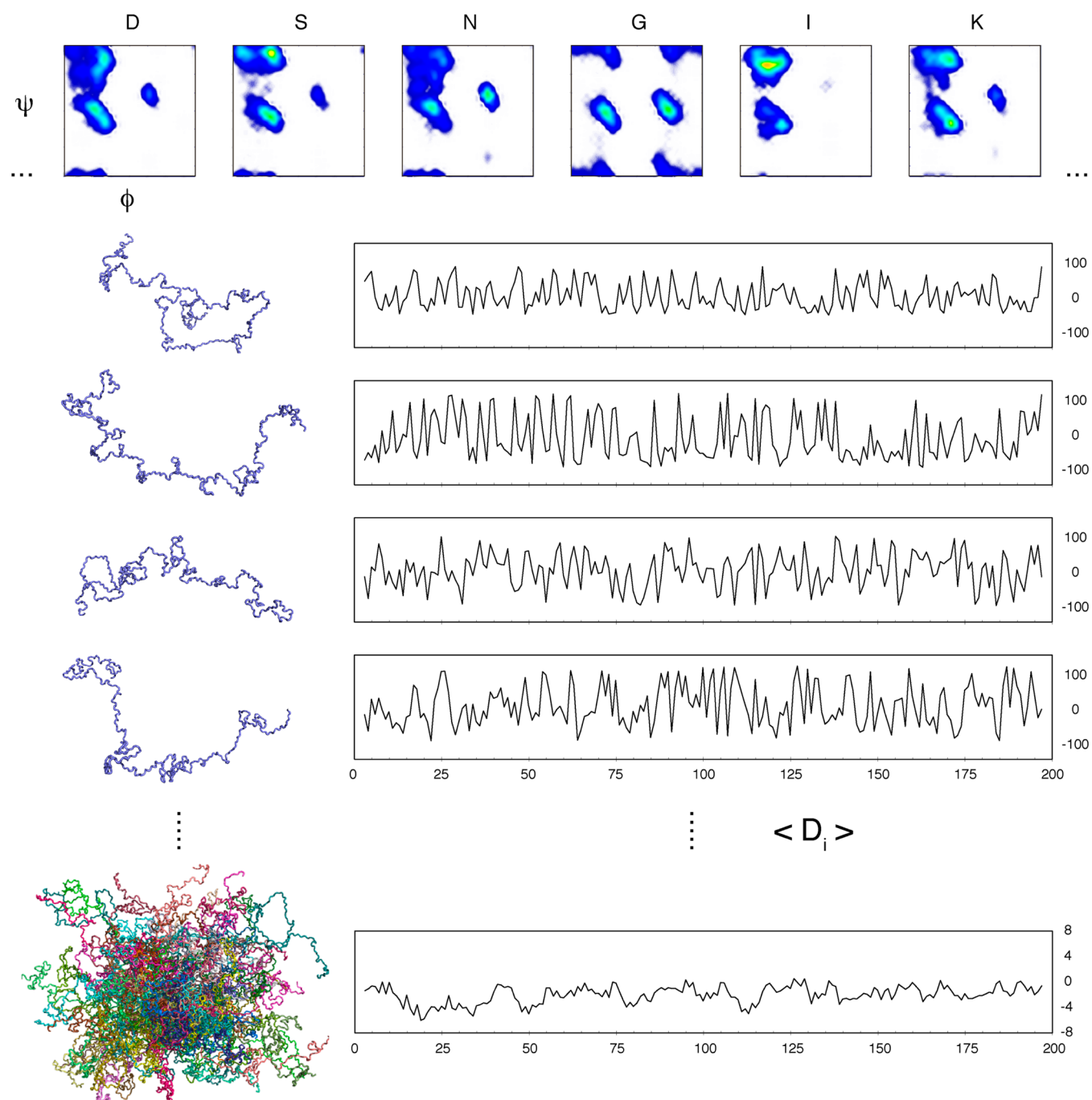
Figure 1. The use of the statistical coil generator *flexible-meccano* for predicting observables of IDPs. *Flexible-meccano* relies on an amino acid specific coil library derived from loop regions of high-resolution crystal structures (top). PDB files of the protein are constructed by randomly sampling the coil library according to the primary sequence in conjunction with a simple steric exclusion principle (left). Observables are calculated for each PDB file, for example, RDCs (right), which can be averaged over a large ensemble to yield expected values for IDPs.

involving local folding-upon-binding into partner-specific bound conformations, or nonspecific and transient fly-casting complexes.[7−13]

To better understand the link between primary sequence and the free-energy surface explored by IDPs, and to determine how primary sequence encodes the diverse mechanisms that characterize function in these proteins, it is essential to develop atomic resolution descriptions of the molecular behavior of IDPs in their free and bound conformations.[14−18] IDPs exhibit high levels of flexibility, even in crowded environments such as the cell,[19] placing them outside the realm of classical structural

biology. All classical approaches to structural biology, in particular crystallographic methods, attempt to determine the three-dimensional structure of a single conformational substate, an approach which has little relevance for an IDP, where such a state represents only one out of a continuum of significantly populated substates. Substantial efforts have therefore been devoted to the development of molecular descriptions that account for the conformational heterogeneity inherent to IDPs. Among the available solution-state techniques, nuclear magnetic resonance (NMR) spectroscopy is now established as one of the most powerful experimental tools for the study of

IDPs, reporting on atomic resolution, ensemble-averaged conformational descriptors under conditions that are close to physiological.[14,20,21]

The development of experimental and analytical techniques to study IDPs has evolved in parallel over recent years, and here we present a general overview of recent progress in the interpretation of solution-state experimental data, in particular from NMR, to describe the free-energy landscape explored by IDPs. We illustrate this general progress with examples from our own and other laboratories active in this field, describing analysis of first-order interactions such as chemical shifts and dipolar couplings, as well as relaxation effects reporting on dynamics and transient contacts. Because of the insufficient independent experimental measurements as compared to the number of degrees of conformational freedom, molecular ensemble descriptions of IDPs are likely to be highly underdetermined. As the development of such ensemble descriptions is gaining popularity, we therefore pay particular attention to the calibration and validation of representative ensembles developed using different approaches, to assess their ability to accurately delineate the conformational space sampled by IDPs in their free and bound forms.
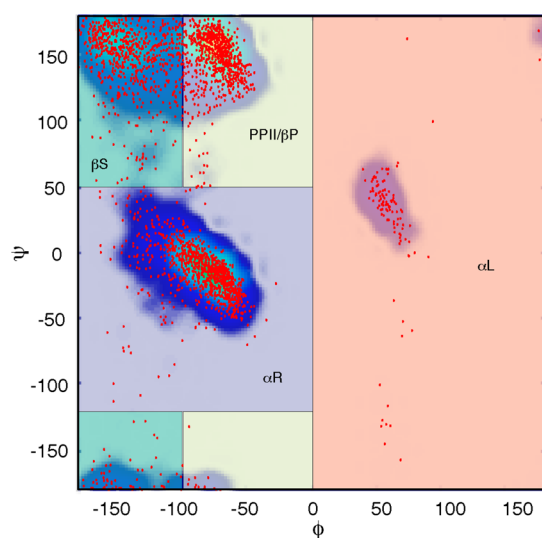

Figure 2. Ramachandran plot showing the statistical coil sampling of threonine (red points and density map from low, blue, to high, red population). The Ramachandran plot is divided into four regions: αR (purple), αL (salmon), βS (green), and βP (yellow). The populations of these regions will be discussed throughout this Review.

2. NMR METHODS FOR PROBING THE CONFORMATIONAL BEHAVIOR OF IDPs

2.1. NMR and the Ensemble Average

Although amide proton NMR resonances from IDPs exhibit a comparatively restricted chemical shift dispersion as compared to globular, folded proteins, the high flexibility of the chain results in spectroscopic properties that are characteristic of small molecules, allowing for backbone chemical shift assign-

ment of ^{1}H, ^{15}N, and ^{13}C resonances, even for very large intrinsically disordered proteins.[22] The dynamic properties of the chain yield spectra that are only limited by the spectral overlap, rather than the molecular weight restrictions that apply to folded proteins. This property also allows for the development and application of multidimensional experiments for the assignment of highly overlapped IDP spectra. The limited spectral width in the ^{1}H^N dimension has led to the development of a number of multidimensional experiments using selective excitation to enhance longitudinal relaxation rates of the ^{1}H^N nuclei via dipole–dipole interactions with aliphatic protons, thereby allowing rapid pulsing.[23–26] Sparse sampling techniques can also be exploited to allow higher dimensional spectra to be recorded efficiently.[27–30]

Each NMR resonance reports on an ensemble average over all equivalent nuclei present in the sample and on an average over all conformations that interconvert on rates faster than the so-called chemical shift time scale (in the millisecond range for solution-state biomolecular NMR). This vast conformational range therefore provides well-behaved statistical averaging, which is particularly appropriate for the investigation of highly dynamic molecules such as IDPs. The obvious disadvantage is that no unique solution exists that can give rise to a single experimental parameter (for example, a chemical shift), highlighting the importance of a second, equally powerful aspect of NMR, and that is the number of experimental parameters that have distinct and complementary conformational dependences. This richness raises the prospect of measuring sufficient experimental data to reduce conformational ambiguity and better define the free-energy landscapes of IDPs, even at amino acid resolution. NMR is rich in both short-range and long-range structural information, distinguishing this technique from other spectroscopies such as circular dichroism (CD),[31–33] Raman optical activity (ROA),[34,35] Fourier transform infrared (FTIR)[36] or Förster resonance energy transfer (FRET),[37–39] and small-angle X-ray scattering SAXS[40,41] that are sensitive to either local or long-range conformational properties but not directly to both.

Conformational averaging of NMR observables can be modeled in terms of structural ensembles under certain assumptions about the interconversion rates of different conformers. For first-order interactions, such as chemical shifts and scalar or dipolar couplings, this effectively means that interconversion rates must be faster than the chemical shift time scale, an assumption that in general is well-founded for IDPs in solution, where significant line-broadening or evidence for slow exchange (multiple peaks) is relatively rare. Interactions that involve nuclear or electron spin relaxation are more complex, because in these cases the measured relaxation rate is also dependent on the time scales characterizing the angular correlation function of the relaxation active interaction (often corresponding to an internuclear vector). Overall, however, the averaging properties of NMR parameters, and the established theoretical basis associated with calculation of this average, make it an extremely powerful tool for investigating the dynamic behavior of IDPs.

2.2. Chemical Shifts

The chemical shift (CS) of a nucleus depends on its local physicochemical environment and is exquisitely sensitive to the presence of secondary structure in both folded and disordered proteins.[42–47] Wishart and co-workers proposed, already more than 20 years ago, the chemical shift index (or the secondary
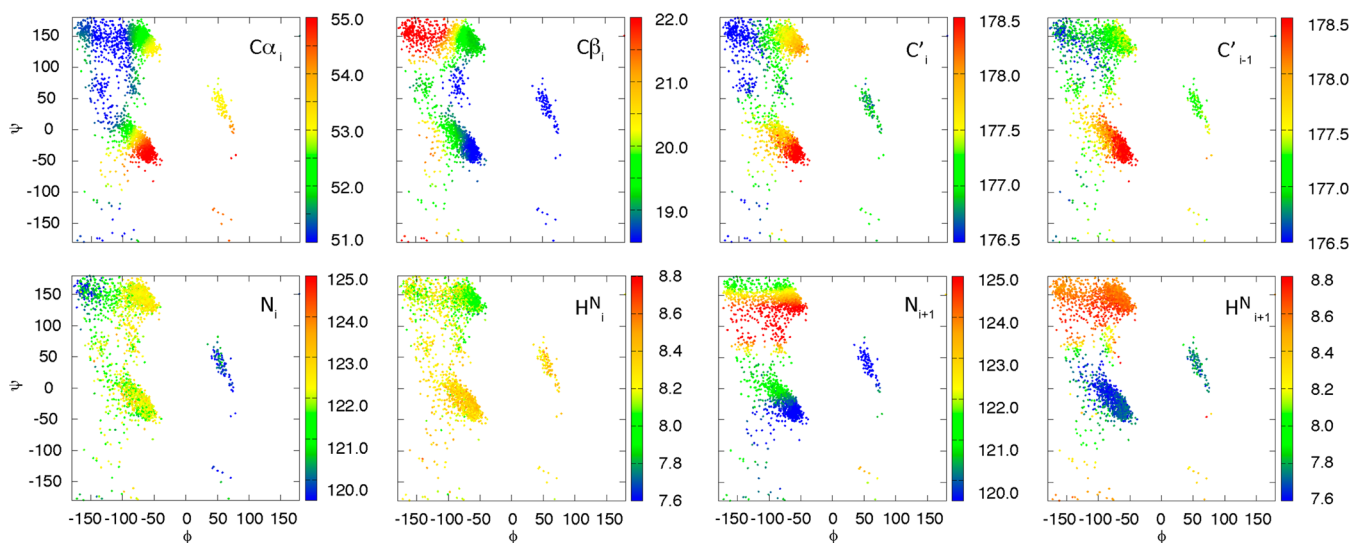
Figure 3. Dependence of different types of chemical shifts of an -Ala-Ala$_{i-1}$-Ala$_i$-Ala$_{i+1}$-Ala- peptide on the backbone dihedral angles of residue i. An ensemble comprising 50 000 conformers of the peptide was constructed using *flexible-meccano*, and the chemical shifts were predicted using SPARTA for each conformation. The conformations were clustered into bins with a radius of 1° according to the dihedral angles of residue i, and the chemical shifts within each bin were subsequently averaged and plotted against ϕ and ψ of residue i. Reprinted in part with permission from ref 82. Copyright 2012 American Chemical Society.
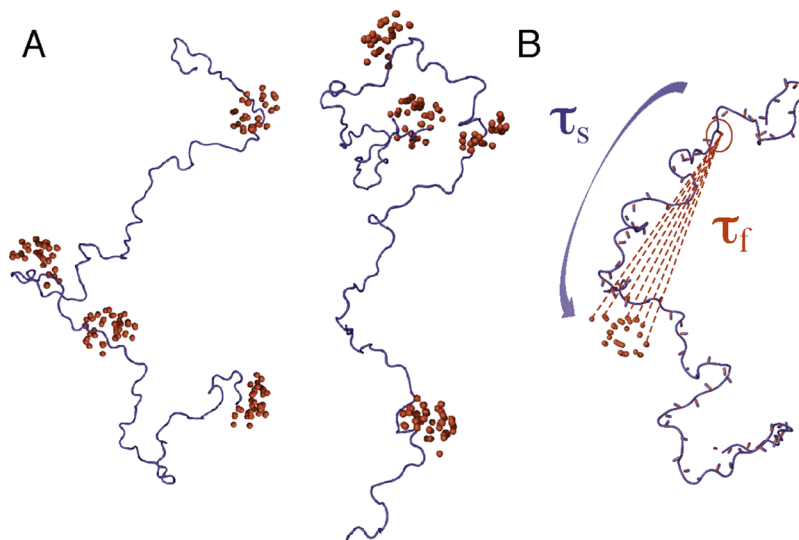

Figure 4. Analysis of PREs in IDPs by explicitly accounting for the dynamics of the MTSL side chain. (A) Two *flexible-meccano* conformers (blue) are shown with distributions of the side-chain MTSL spin labels (red) for four different spin label positions. Previously proposed MTSL rotameric libraries were randomly sampled for a total of 600 conformers for each site. Each position was retained and included in the averaging procedure, if no steric clashes were found with the given backbone conformation. (B) The theoretical framework for predicting PREs for a single IDP conformer takes into account two correlation times accounting for the local dynamics of the MTSL spin label (τ_f) and the overall reconfiguration of the interaction vector (τ_s). Reprinted in part with permission from ref 112. Copyright 2011 American Chemical Society.

chemical shift) as a fast and reliable way of identifying secondary structure types in proteins directly from experimental ^{1}H and ^{13}C CSs.[48−50] The secondary chemical shift (SCS) probes the deviation of experimental CSs with respect to so-called random coil (RC) shifts:

$$\delta_{SCS} = \delta_{EXP} - \delta_{RC} \qquad (1)$$

The random coil values depend on the amino acid type and reflect expected CSs for a protein in the absence of secondary and tertiary structure. The CSs of backbone ^{13}C nuclei are the most sensitive to the presence of secondary structure, with ^{13}Cα and ^{13}C' SCSs that are positive in α-helices and negative in β-sheets, while ^{13}Cβ SCSs follow the inverse dependence. If

quantitative statements are to be made about secondary structural content, it is important that the δ_{RC} values used to calculate the SCSs match the experimental conditions as closely as possible. For this reason, different sets of δ_{RC} values have been derived under various conditions. In most cases, short unfolded, glycine-rich peptides have been used, where CSs are measured for the central residue when substituted for each of the 20 amino acid types. Different data sets have been obtained under nondenaturing,[51] mildly denaturing[52,53] or denaturing conditions,[54−56] and nearest neighbor effects on the CSs have been investigated. More recently, a systematic study of the temperature and pH dependence of random coil values has
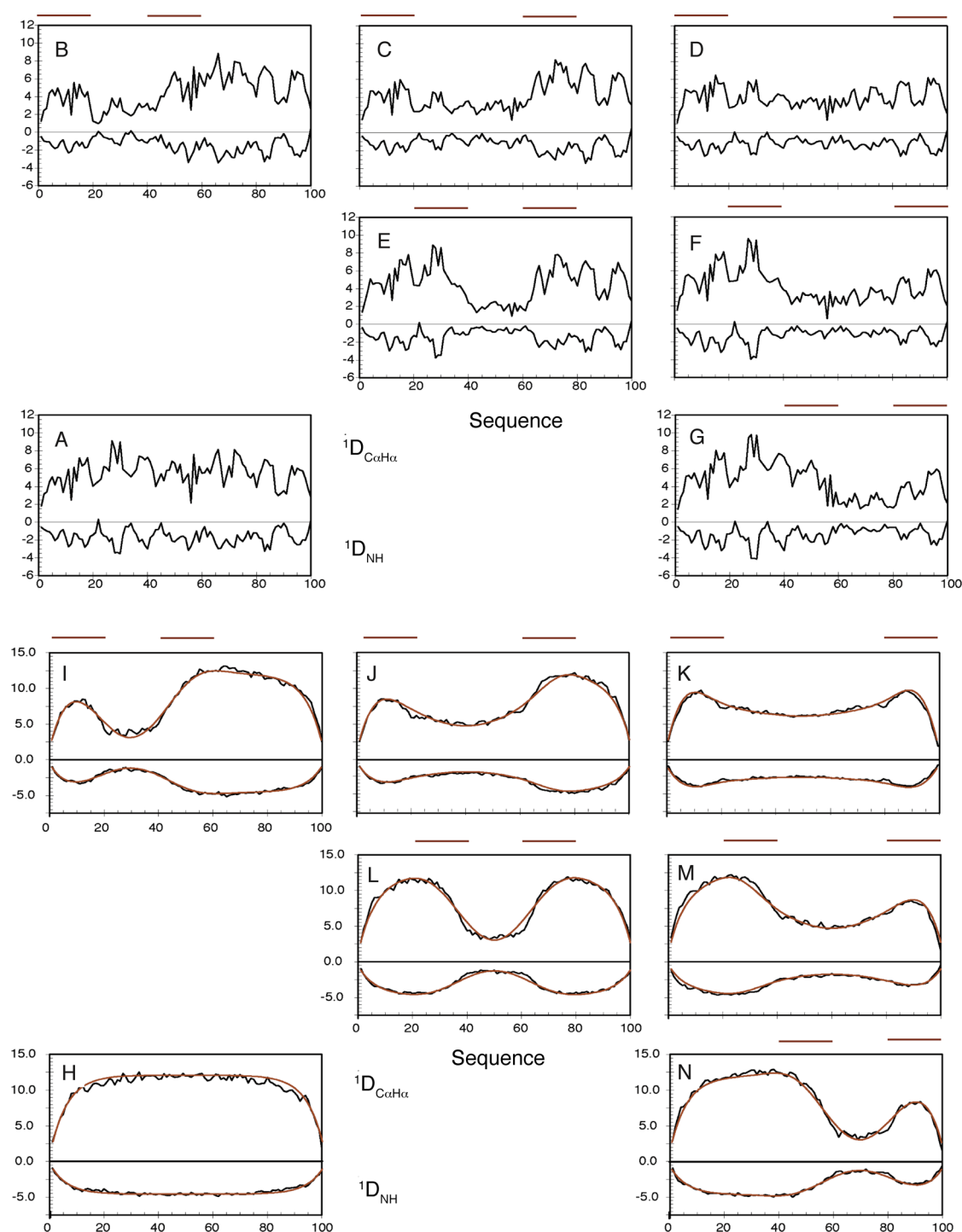
Figure 5. RDCs are modulated by the presence of persistent long-range contacts in IDPs. (Top) *Flexible-meccano* simulations of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs (black lines) in a model protein of 100 amino acids with arbitrary primary sequence in the absence (A) and the presence (B–G) of different persistent long-range contacts: 1–20 and 41–60 (B), 1–20 and 61–80 (C), 1–20 and 81–100 (D), 21–40 and 61–80 (E), 21–40 and 81–100 (F), 41–60 and 81–100 (G). (Bottom) *Flexible-meccano* simulations of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs (black lines) in a poly valine chain of 100 amino acids in the absence (H) and the presence (I–N) of the same long-range contacts. The red lines correspond to the parametrization of the baseline via a simple analytical expression that depends on the length of the chain and the positions of the contacting regions. In both panels, RDCs were averaged over 100 000 conformers, and the red lines on top of each plot indicate the positions of the two contacting regions. Reprinted with permission from ref 112. Copyright 2011 American Chemical Society.

been carried out providing a useful addition to already existing data sets.[53]

Another way of obtaining δ_{RC} values is to employ the large database (for example, the BMRB[57]) of experimentally assigned CSs of proteins of known three-dimensional structure.[58–61]

Characteristic CSs of α-helix and β-sheet can be derived directly from the database. Random coil values are derived from CSs measured in loop regions of proteins that are also present in high-resolution structural databases. Statistical coil ensembles are generally assembled by randomly sampling amino-acid

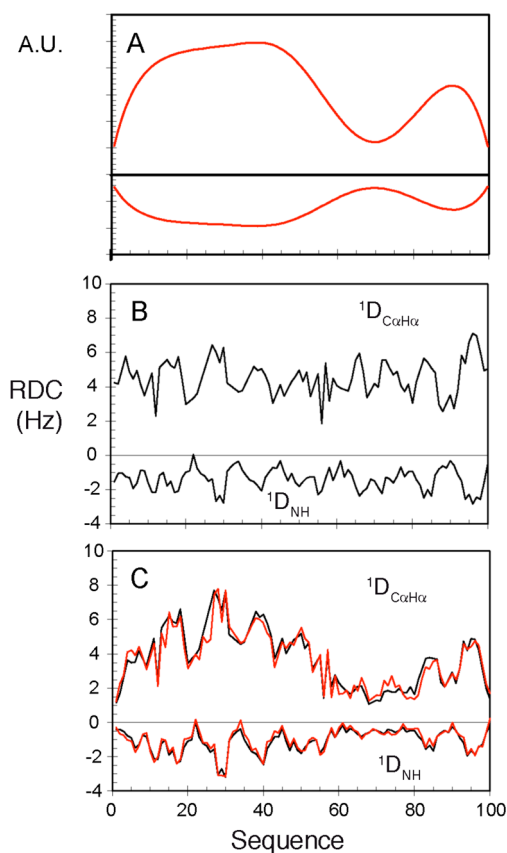**Figure 6.** Example of the combination of analytically calculated baselines and RDCs averaged using local alignment windows for a model protein of 100 amino acids of arbitrary sequence. (A) Baseline contribution calculated analytically for contacts between regions 41–60 and 81–100. (B) RDCs calculated using local alignment windows of 15 amino acids in length. RDCs were averaged over 200 structures. (C) Combination of the baseline from (A) and the local RDCs from (B) (red curves) as compared to the RDCs calculated using a global alignment tensor over 100 000 structures carrying a contact between regions 41–60 and 81–100 (black curves). Reprinted in part with permission from ref 112. Copyright 2011 American Chemical Society.

specific potential energy basins, on the basis of conformations extracted from loop regions of high-resolution three-dimensional structures of proteins, so that random coil and statistical coil descriptions are often equivalent. One of these descriptions, *flexible-meccano* (Figure 1),[62,63] follows these sampling procedures, combined with simple volume exclusion (amino acid-specific hard-spheres placed on the $C\beta$), to construct statistical ensembles of conformers of unfolded or partially folded proteins, and is used throughout this Review to illustrate expected features of NMR parameters measured in IDPs. Statistical coil descriptions have been developed in this way by a number of different groups, with notably similar potentials covering accessible Ramachandran space (Figure 2),[62,64−71] divided into four regions αR, αL, βS, and βP. The populations of these regions will be discussed throughout this Review.

For IDPs that typically adopt conformations close to random coil, CSs have been used to pick up propensities for forming secondary structure.[72,73] Forman-Kay and co-workers have proposed the secondary structure propensity (SSP) algorithm that estimates propensities to form α-helix or β-extended conformations along the primary sequence of IDPs.[74] The

algorithm employs a set of random coil values together with expected CSs for fully formed α-helices and β-sheets. The concerted use of CSs of multiple nuclei allows a more accurate determination of the propensities as well as a simultaneous estimation of potential reference offsets on the experimental CSs. A similar approach is the so-called neighbor-corrected structural propensity calculator (ncSPC) that derives structural propensities by employing the random coil values and their corresponding neighbor correction factors derived from a small IDP chemical shift database in conjunction with singular value decomposition.[75,76]

SSP and ncSPC focus solely on deriving propensities to form α-helix and β-extended conformations; however, growing evidence from vibrational spectroscopy,[77,78] CD,[79] and more recently NMR[80−82] suggests that poly proline II (βP) conformations are highly abundant in IDPs. The algorithm δ2D proposed by Vendruscolo and co-workers provides propensities to form αR, βS, and βP conformations from experimental CSs.[83] The algorithm employs previously derived characteristic CSs of αR, βS, and βP and solves a set of linear equations for the populations in these regions, together with a population having a random coil shift.

A potential problem inherent to all approaches that include a random coil shift as part of a fitting procedure is that the exact definition of random coil is rather ill-defined. If random coil is equivalent to statistical coil (vide supra), then βP, αR, and βS should already have finite populations within this continuous distribution. Fitted populations should then be added to these underlying populations, to determine the true population of the region of Ramachandran space. If however random coil is defined as having zero population of βP, αR, and βS, then the hypersurface is not a continuous statistical coil distribution, and values could not be easily compared to experimental values derived from peptides or under denaturing conditions where these regions are expected to have a finite population. This question is particularly pertinent for $^1H\alpha$, $^{13}C\alpha$, and $^{13}C\beta$ CSs that are not expected to depend strongly on the presence or absence of the hydrogen bonds that stabilize cooperative secondary structure.

The use of secondary CSs as a means to identify secondary structure, or the resolution of a set of linear equations to determine populations of a limited number of regions of Ramachandran space, are both low resolution attempts to capture the main features of the complex hyper-surfaces relating observed CSs to backbone dihedral angles. These hyper-surfaces are convoluted with additional contributions from neighboring residues, side-chain conformations, hydrogen bonding, ring-current effects, protonation states, as well as effects from different ionic strengths and temperatures, which are inherently difficult to predict theoretically. However, as the number of proteins of known three-dimensional structure with associated experimental CSs increases, we are gaining more insight into the shape and appearance of these CS hyper-surfaces. This insight has formed the basis for the development of a number of CS predictors that allow the calculation of CSs directly from the three-dimensional structure of a protein. Common to the most successful predictors such as SHIFTX,[84] SHIFTX2,[85] SPARTA,[86] SPARTA+,[87] and CamShift[88] is that they employ assigned CSs for proteins of known structure to parametrize their CS prediction method. All of these methods rely on a static view of protein structures and do not account for the potential impact of protein dynamics on the experimental (and in principle already ensemble-averaged)
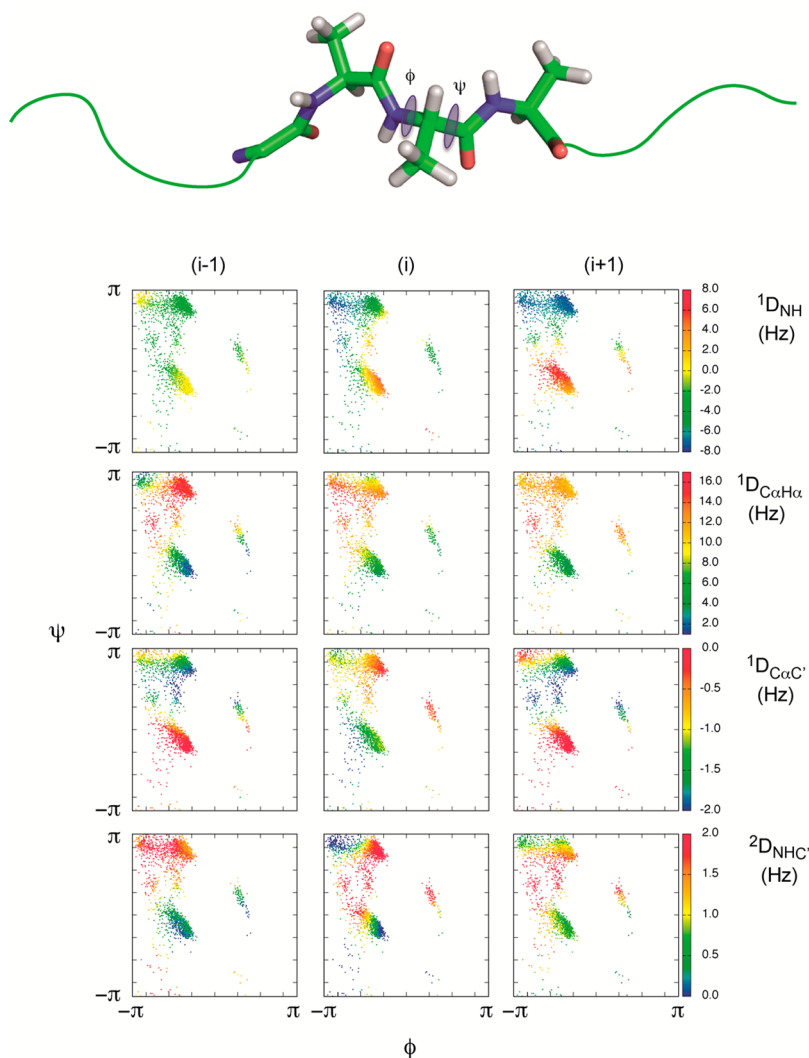
Figure 7. Dependence of different types of RDCs of an -Ala-Ala$_{i-1}$-Ala$_i$-Ala$_{i+1}$-Ala- peptide on the backbone dihedral angles of residue i. An ensemble comprising 1 000 000 conformers of the peptide was constructed using *flexible-meccano*, and the RDCs were predicted for each conformation. The conformations were clustered into bins with a radius of 1° according to the dihedral angles of residue i, and the RDCs within each bin were subsequently averaged and plotted against ϕ and ψ of residue i. Reprinted in part with permission from ref 82. Copyright 2012 American Chemical Society.

CSs. A step forward in this direction was recently taken by Brüschweiler and co-workers who proposed the predictor PPM that explicitly accounts for protein dynamics on CSs by parametrizing the predictor against long MD simulations of a set of proteins with known experimental CSs.[89] The importance of taking into account the effect of dynamics on CSs has been amply demonstrated as predicted CSs averaged over long MD simulations significantly better reproduce experimental CSs as compared to their prediction from a single static structure.[90−92]

CS prediction using these parametrized hypersurfaces can in turn provide insight into the ability of CSs to distinguish between different conformational propensities. Statistical coil simulations are shown in Figure 3, where backbone CSs have been predicted for trialanines over a large ensemble of conformers using an empirical CS predictor.[82,86] It is clear that ^{13}C CSs of the βP region of Ramachandran space are, to a large extent, degenerate with random coil shifts that would be measured by taking a population weighted average over the entire free-energy surface. On the other hand, ^{1}H^N and ^{15}N^H

CSs depend more strongly on the presence of βP conformations; however, they have a stronger dependence on experimental conditions such as pH, temperature, and ionic strength, making their prediction more error-prone.

2.3. Scalar Couplings

Information about secondary structural propensities can also be extracted from scalar coupling constants measured between nuclei of the protein backbone. Scalar couplings depend on backbone dihedral angles[93−95] and, in the same way as CSs, represent a population-weighted average over all conformations sampled in solution up to the millisecond time scale. The dependence of scalar couplings on the main-chain torsion angles can be described using a Karplus relationship[93] that is normally parametrized against experimental scalar coupling constants measured in several proteins of known structure.[65,96] For example, the three-bond coupling constant $^3J_{HN\alpha}$ depends on the backbone dihedral angle ϕ allowing a clear distinction between α-helical ($^3J_{HN\alpha} < 5$ Hz) and β-sheet conformations ($^3J_{HN\alpha} > 8$ Hz).[96,97] Scalar couplings have also been measured from short peptides providing estimates of random coil
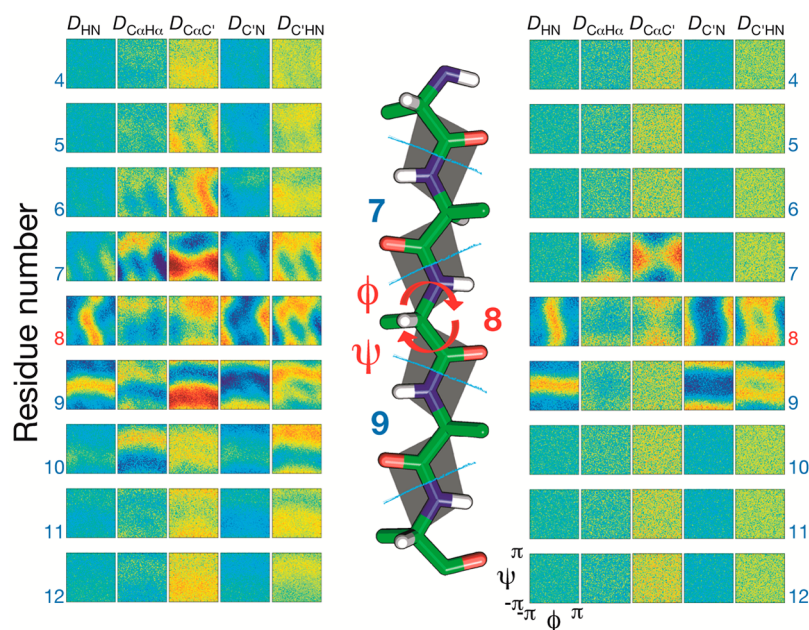
Figure 8. Propagation of the influence of conformational sampling along the peptide chain and its effect on predicted RDCs. RDCs predicted from a *flexible-meccano* ensemble of one million conformers of pentadeca-alanine are plotted against the backbone dihedral angles of the central residue averaged over bins with a radius of 1°. (Right) All residues sample {ϕ,ψ} space uniformly. (Left) The central residue samples conformational space uniformly, while the others sample {ϕ,ψ} angles according to the statistical coil distribution of alanine. Red to blue colors correspond to higher to lower RDC values. Reprinted with permission from ref 225. Copyright 2013 Wiley-VCH.

couplings that upon comparison with experimental couplings allow a direct identification of secondary structures.[98] In common with CSs, parametrization of Karplus curves is often calculated against known angles measured in folded proteins, so that some a priori unknown proportion of the ensemble-averaged property is again incorporated into the fitted parameters. Karplus parameters calculated using density functional theory and applied to each snapshot of an MD trajectory have been shown to reproduce experimental couplings more accurately in folded proteins,[99,100] and such an approach will likely improve J-coupling analysis in IDPs.

Early comparison of measured $^3J_{HNH\alpha}$ couplings with the distribution of ϕ angles of amino acids in loop regions of solved protein structures provided experimental evidence that amino acids possess specific intrinsic ϕ propensities.[65,101] This observation has consequences for IDPs, where the intrinsic conformational propensity of each amino acid is expected to dictate the conformational and dynamic behavior of the protein and eventually its function. Measurement of backbone scalar couplings in IDPs has been used to locate regions of transiently populated secondary structure as well as to report on the general distribution of dihedral angles in Ramachandran space at amino acid resolution.[65,101–110] They have also been used to cross-validate ensembles developed on the basis of other experimental data.[110–113] Other scalar couplings reporting on backbone ψ angle distributions, such as $^3J_{C\alpha C\alpha}$, $^3J_{NH\alpha}$, $^3J_{NC\beta}$, and $^3J_{NN}$,[114–118] although smaller in magnitude than couplings reporting on ϕ, can be very useful in IDPs to more accurately probe the populations of βP conformations.[116,119]

Information about side-chain conformations can also be obtained through scalar coupling constants. For example, the $^3J_{H\alpha H\beta}$, $^3J_{NC\gamma}$, and $^3J_{C'C\gamma}$ couplings report on the χ_1 torsion angles and have been used to study side-chain conformations in both folded and disordered proteins.[120–122] The measurement of these couplings in denatured or intrinsically disordered

proteins is particularly challenging due to the poor dispersion of side-chain resonances due to motional averaging. Schwalbe and co-workers studied side-chain conformations in denatured lysozyme through $^3J_{NC\gamma}$, $^3J_{C'C\gamma}$, $^3J_{C'H\beta}$, and $^3J_{H\alpha H\beta}$ couplings and obtained estimates of the populations of different staggered χ_1 rotamers.[123,124] Grzesiek and co-workers exploited multiple scalar couplings involving Hβ nuclei ($^3J_{NH\beta2,3}$, $^3J_{C'H\beta2,3}$, and $^3J_{H\alpha H\beta2,3}$) in urea-denatured ubiquitin and protein G to obtain χ_1 angle distributions with unprecedented precision together with stereospecific assignment of the H$^\beta$ protons.[125] The results showed that the derived χ_1 rotamer populations are in general similar to those derived from loop regions of folded protein structures; however, some variations occur locally indicating contributions from sequence- and residue-specific interactions. $^3J_{NC\gamma}$ couplings measured in phosphorylated arginine-serine repeats were also in good agreement with values calculated from ensemble descriptions driven by NMR backbone restraints, indicating a coupling between backbone and side-chain sampling.[110]

Analysis of the local conformational distributions of tripeptides using diverse scalar couplings combined with vibrational spectroscopy (infrared, Raman, and vibrational CD) has been used to determine how sampling propensity in the unfolded state can affect folding of peptides and proteins. These studies indicate that observed enhanced propensities for Thr, Ser, Cys, Asn, and Asp to stabilize different turn motifs in folded proteins reflect the intrinsic sampling of these amino acids in short peptides.[33,119,126,127] Related work compared MD simulation to measured couplings in alanine peptides of different lengths,[128] and scalar couplings were recently used to measure the accuracy of state-of-the-art force fields.[71,129] Finally, a recent study of the effects of pressure on the conformational sampling of α-synuclein used $^3J_{HNH\alpha}$ couplings to invoke a general increase in the population of βP at 2500 bar as compared to standard pressure conditions.[130]

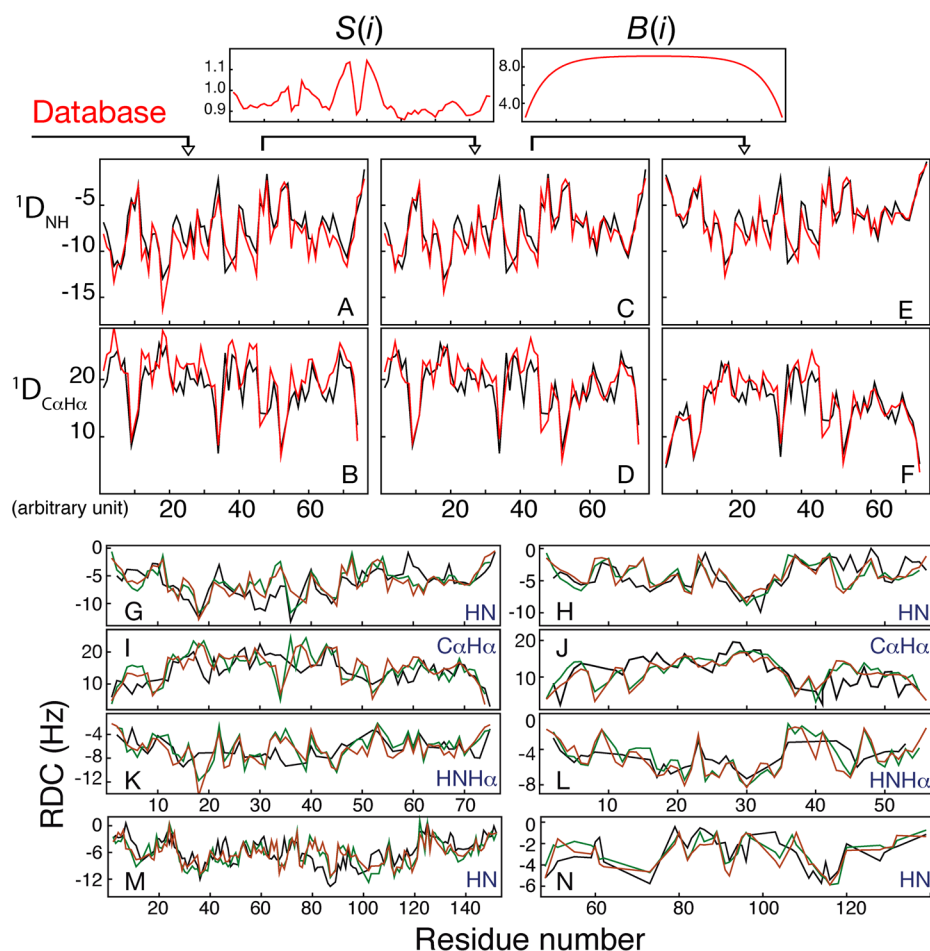6639

dx.doi.org/10.1021/cr400688u | Chem. Rev. 2014, 114, 6632–6660

Figure 9. (Top) Flow-chart of the *seq2rdc* approach for fast and reliable prediction of RDCs from unfolded systems. Simulated $^1D_{NH}$ (A) and $^1D_{C\alpha H\alpha}$ (B) RDCs for ubiquitin using local alignment windows (black lines) and using the *seq2rdc* approach (red lines) that employs a RDC look-up table composed of all possible triplets, thereby taking into account the conformational sampling of the residue of interest and that of nearest neighbors. The RDCs from *seq2rdc* were subsequently corrected with a scaling factor $S(i)$ taking into account the composition of residues $i-7$ to $i-2$ and $i+2$ to $i+7$ (C,D) as well as a length-dependent hyperbolic baseline (E,F). (Bottom) Prediction of RDCs using *seq2rdc* (red lines) as compared to experimentally measured RDCs (black lines) and RDCs predicted for *flexible-meccano* ensembles using a global alignment tensor for each conformation (green lines). Data are shown for the following proteins: denatured ubiquitin (G,I,K), denatured GB1 (H,J,L), denatured apo-myoglobin (M), and denatured Δ131Δ staphylococcal nuclease (N). Reprinted with permission from ref 225. Copyright 2013 Wiley-VCH.

2.4. Paramagnetic Relaxation Enhancement

By their nature IDPs are not expected to exhibit stable tertiary structure. Transient interactions between distant regions of the primary chain can however play important roles in terms of physiological function,[131,132] for example, through early folding or misfolding events, to avoid aggregation or proteolysis, to control access to binding sites, or to allow colocalization of partners via fly casting interactions. Standard $^1H-^1H$ NOESY measurements[133] are relatively insensitive to such transient contacts if they are very rare or weakly populated. To enhance the sensitivity of detection of such long-range information, a paramagnetic moiety can be introduced in the protein, so that the far stronger dipolar relaxation between nuclear spins and the paramagnetic spin can be detected at much lower populations, or much further distances. Such paramagnetic spins are generally introduced into the protein by attaching a nitroxide group to the reactive side chain of a cysteine that replaces a native amino acid in a mutated form of the protein.[134−136] The gyromagnetic ratio of the electron spin is 658 times higher than the proton spin, so that observed relaxation rates are enhanced massively. As an example, the transverse relaxation rates of an observed proton spin are enhanced according to:

$$\Gamma_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_H^2 g_e^2 \mu_B^2 s_e(s_e + 1)\left[4J(0) + 3J(\omega_H)\right] \quad (2)$$

where g_e is the electron g-factor, γ_H is the gyromagnetic ratio of the observed nucleus (proton), s_e is the electron spin quantum number, ω_H is the proton Larmor frequency, μ_B is the Bohr magneton, and μ_0 is the permittivity of free space. Similar expressions are valid for other nuclei or for longitudinal relaxation rates. The spectral density function, $J(\omega)$, comprises the orientational and distance-dependent components:

$$J(\omega) = \langle r_{H-e}^{-6}\rangle\left\{\frac{\tau_c}{1 + \omega^2\tau_c^2}\right\} \quad (3)$$

where $\tau_c = \tau_r\tau_s/(\tau_r + \tau_s)$ is defined in terms of τ_s and τ_r, the electron spin and rotational correlation times, respectively. r_{H-e} is the instantaneous distance between the nuclear and the electron spins.

Such paramagnetic relaxation enhancements (PRE) are clearly dominated by the distance distribution functions
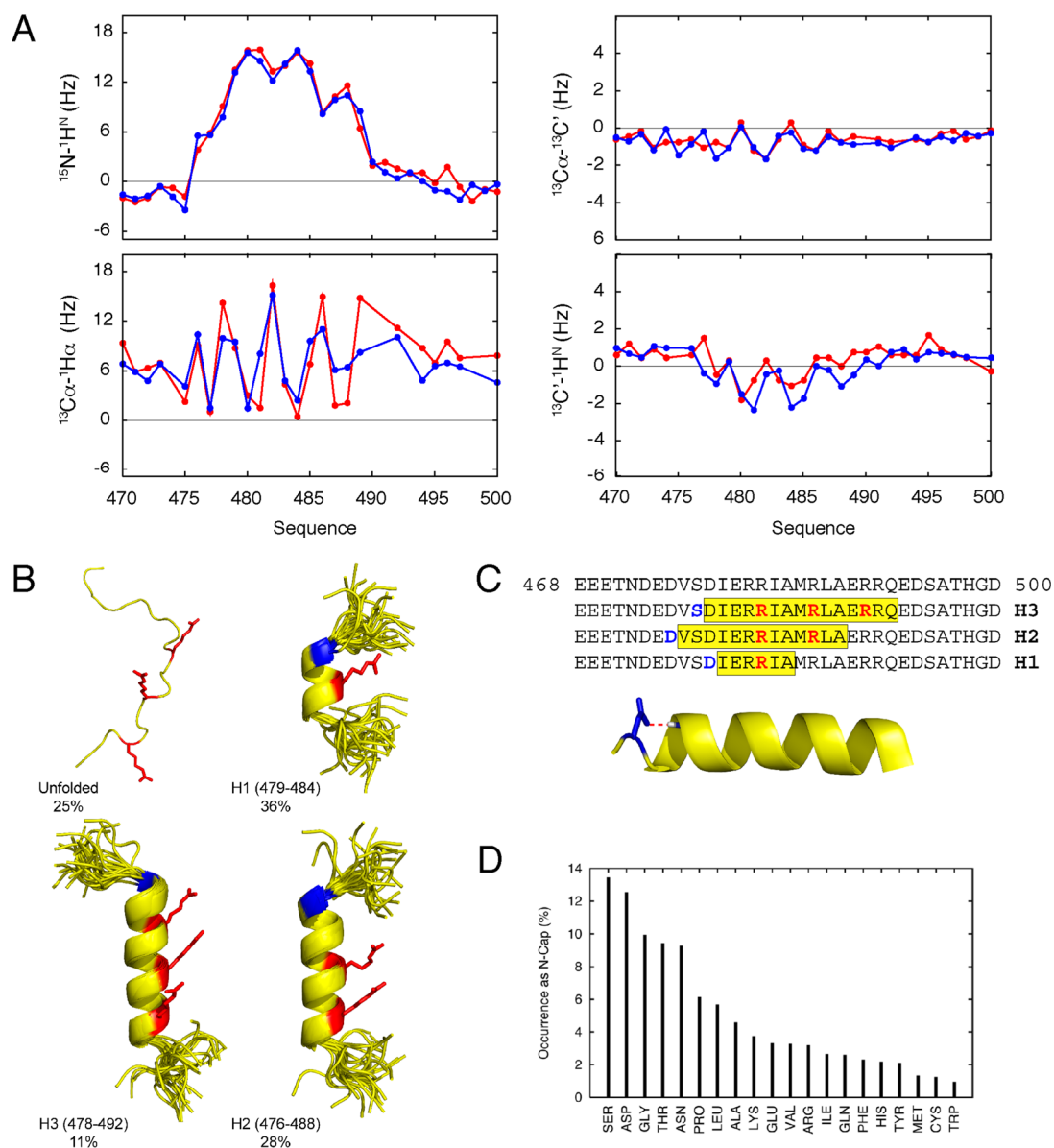
Figure 10. (A) Reproduction of the four types of RDCs in the molecular recognition element of Sendai virus N_{TAIL} for the best-fitting model containing three helical ensembles. Experimental RDCs are shown in red, while back-calculated RDCs from the model are shown in blue. (B) Molecular representation of the proposed conformational equilibrium of the molecular recognition element of N_{TAIL} in solution. The four conformations are presented as a single structure for the completely disordered form and as 20 randomly selected conformers for the three helical conformations. The molecular recognition arginines are displayed in red, while N-capping residues are shown in blue. (C) The amino acid sequence of the molecular recognition element of N_{TAIL} showing the positions of the selected helices. The cartoon figure illustrates an N-capping aspartic acid side chain–backbone interaction. (D) The occurrence of different types of amino acids as N-capping residues in helices of folded proteins. Reprinted with permission from ref 226. Copyright 2013 American Chemical Society.

between the observed nuclear spin and the attached probe,[137−143] but are also dependent on the dynamic properties of the relaxation active interactions. This can be affected both by the amplitude and by the time scale of the motion of the MTSL-bearing side chain, which are not generally known, although models have been derived from EPR in folded proteins.[144,145] We note that a number of spin probes with reduced mobility have been proposed that greatly alleviate the problem of local dynamics,[146,147] but that these probes often require the presence of specific and relatively rigid geometries on the backbone that make them less appropriate for IDPs.

Mobility of the side chain can be taken into consideration by introducing an order parameter relating to the motion of the spin label relative to the molecular frame of each conformer:[148,149]

$$J(\omega) = \langle r_{H-e}^{-6} \rangle \left\{ \frac{S_{H-e}^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S_{H-e}^2)\tau_e}{1 + \omega^2 \tau_e^2} \right\} \quad (4)$$

where the order parameter S_{H-e}^2 describes the motion of the dipolar interaction vector, and $\tau_e = 1/(\tau_i^{-1} + \tau_r^{-1} + \tau_s^{-1})$ where τ_i represents the effective internal correlation time of the spin label (Figure 4). Using a formalism directly comparable to that developed for the incorporation of motional effects into the interpretation of $^1H-^1H$ NOESY interactions,[149] the order parameter is then decomposed into radial and angular components:

$$S_{H-e}^2 = \frac{4\pi}{5} \sum_{m=-2}^{2} |\langle Y_2^m(\Omega^{mol})\rangle|^2 \langle r_{H-e}^{-6}\rangle^{-1} \langle r_{H-e}^{-3}\rangle^2 \tag{5}$$

where Ω^{mol} is the orientation of the relaxation active vector in the frame of the conformer. Assuming that the interconversion between different side-chain conformations is independent of, and faster than, the interconversion between different discrete backbone conformers, and that the interconversion between the different substates over the entire ensemble is also fast as compared to both chemical shifts and the relaxation rates, the effective rates can then be calculated for each conformation, and averaged over an ensemble of backbone conformers:

$$\Gamma_2^{total} = \frac{1}{N} \sum_{c=1}^{N} \Gamma_{2,c} \tag{6}$$

Whether internal motion of the spin label is taken into consideration or not, PREs are often interpreted in terms of distance restraints between the unpaired electron and the nuclear spin and incorporated in simulated annealing type calculations using single-copy or more commonly ensemble restrained molecular dynamics simulation.[135,141–143,150–154] Measured relaxation rates can also be incorporated as constraints,[155] or in terms of probability distributions.[152,156,157] PREs can also be used to select representative ensembles (vide infra) from a large pool of possible conformers using the same kind of formalism.[106,112,158,159]

As mentioned above, it is important to consider the time scale of the intrinsic dynamics of the chain with respect to the effective relaxation rates of the different exchanging substates. The above formalism is only valid if the interconversion rates are faster than the differences between the effective relaxation rates. If this is not the case, the effective relaxation rates would be quenched by the interconversion between substates, as predicted by the McConnell equations.[160] These intermediate exchange effects can in principle be used to probe the time scales of dynamics in highly disordered systems. We also note that a recent study proposed a promising formalism to combine paramagnetic relaxation measurements from multiple magnetic fields to estimate the distribution radius of a disordered group relative to a rigid portion of the macromolecule independent of any prior structural knowledge.[161] In vacuo high temperature MD simulation, to enhance the conformational sampling radius, was also used to study the dynamic regimes that would be necessary to reproduce experimentally observed PRE profiles measured in urea-denatured ubiquitin.[162] In this case, the interaction between the nuclear and electron spins was expressed rather in terms of translational diffusion of two tethered spheres in a harmonic potential, and the results were compared to those derived from FRET measurements of denatured proteins.

2.5. Nuclear Spin Relaxation

Nuclear spin relaxation is sensitive to the reorientational dynamics of relaxation active interactions; in the case of ^{15}N spin relaxation this is normally dominated by the dipolar interaction between the amide proton and nitrogen, providing sensitive probes of the motional properties of bond vectors positioned throughout the unfolded protein.[163] The information that can be derived from such measurements is limited to time scales shorter than the time taken for complete decorrelation of the angular correlation function, which in an IDP is normally in the range of 1–5 ns (at 298 K).

Nevertheless, ^{15}N spin relaxation has been shown to correlate with local order, or bulkiness of side chains along the peptide chain,[164] and has been used to study backbone and side-chain motions in intrinsically disordered and partially folded states.[165–173] Increasing computational power and the development of MD-dedicated processors have made longer time scale MD simulations available for comparison with experimental NMR relaxation rates as was shown recently for partially folded acyl coenzyme A binding protein (ACBP).[174] To address the interpretation of different spin relaxation rates in unfolded and partially folded proteins, where the concept of an overall motion common to all sites that is appropriate for folded proteins has no real relevance, Bruschweiler et al developed an MD-based framework for assessing relaxation active correlation times.[175] A recent MD study of the partially disordered DNA binding basic leucine zipper domain of the yeast transcription factor GCN4 used chemical shifts and ^{15}N spin relaxation rates to probe populations of residual helical elements and helix capping. The authors performed numerous long MD simulations, and compared relaxation rates calculated from autocorrelation functions to experimental data.[176] Recent developments have also addressed the important effects of long disordered chains on the rotational diffusion properties of multidomain proteins,[177–179] which have clear relevance for the diffusional properties of IDPs in crowded physiological environments.

Finally, relaxation dispersion CPMG[180,181] or rotating frame relaxation[182] experiments provide remarkable insight into the structural and dynamic origins of enhanced effective transverse relaxation measurements due to microsecond to millisecond time-scale exchange between sites experiencing distinct chemical shifts. Even when the minor species represents a few percent of the total population, these experiments can be used to determine the conformational behavior of the otherwise barely visible state, and have been successfully used to characterize partially folded intermediates and to characterize IDP folding-upon-binding events.[183–185] Although the physical basis of interactions between IDPs and their partner proteins is the subject of a growing number of theoretical[186–188] and numerical[189–191] studies, there are still relatively few experimental examples of atomic resolution characterizations of IDP interactions.[10] In this respect, NMR relaxation dispersion studies hold great promise for our understanding of the molecular mechanisms underlying the function of IDPs, and in particular their relationship to intrinsic conformational sampling, the subject of this Review.

2.6. Residual Dipolar Couplings

Dipolar couplings between pairs of nuclei are expected to average to zero in free solution due to the orientational averaging of the dipolar interaction, for example, for two covalently bound nuclei i and j:

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2 r^3} \left\langle \frac{3\cos^2 \Omega - 1}{2} \right\rangle \tag{7}$$

where Ω is the orientation of the internuclear vector with respect to the static magnetic field, and r is the vibrationally averaged distance between nuclei i and j. The angular brackets report on the average over all orientations sampled by the different substates present in solution, and exchanging with rates faster than the millisecond range. Although the static value of dipolar couplings measured between a single pair of covalently bound spins can be very large (11 kHz for

^{15}N–^{1}H^N spin pair), in the absence of a preferential order in the sample, all orientations are equivalently sampled and zero coupling is measured. If a protein is dissolved in a dilute liquid crystalline medium, the angular average is no longer uniform, some orientations have higher probability, so that the measured dipolar coupling in eq 7 has a nonzero, or residual, value.[192−194] Such residual dipolar couplings (RDCs) represent extremely powerful probes of angular order, and this long-range order can be used to complement short-range distance information, for example, to determine the orientation of different structural elements in folded proteins. RDCs have also been successfully used in folded proteins to probe time and ensemble-averaged conformational equilibria at atomic resolution, on time scales up to the millisecond.[194−198] Most importantly for this Review, RDCs provide precise probes of the conformational behavior of IDPs in solution.[62,199−204]

Alignment of the protein results from a restriction of the conformational sampling due to the presence of the liquid crystalline medium such as lipid bicelles,[192] filamentous phages,[205−207] lyotropic ethylene glycol/alcohol phases,[208] and polyacrylamide gels that have been strained either laterally or longitudinally to produce anisotropic cavities.[209,210] RDCs are measured as the difference between the experimental measurement in the presence and absence of the medium; in the latter case, only the appropriate scalar coupling is measured. RDCs are commonly measured in folded proteins in alignment media that are either neutral or electrostatically charged, resulting in either steric repulsion or repulsion between charged groups on the protein surface and the alignment medium. RDCs measured in IDPs, on the other hand, are almost universally measured in sterically aligning media, because of the relative simplicity of prediction of the alignment tensor under these conditions (based on the three-dimensional shape of the conformation).[211] As an exception to this rule, we note that electrostatic alignment of IDPs with strong charge polarity along the chain has been successfully shown to behave as predicted on the basis of electric fields calculated between solute surface charges and the electric field of the phage.[212]

^{1}D$_{HN}$ RDCs were first measured in denatured forms of folded proteins, with predominantly negatively signed couplings exhibiting a characteristic bell-shaped dependence with respect to the primary sequence.[199−201,213] These observations were explained on the basis of the nearly perpendicular orientation of the N–H^N bond vector relative to the principal alignment axis and therefore the static magnetic field direction (concerning the sign), and the chain like nature of the protein, increasing the efficiency of the angular average in eq 7 toward the termini of the chain.[200,214,215] Two important findings resulted from these experiments; the first concerned the high sensitivity of RDCs to elements of secondary structure in partially folded IDPs, while the second concerned the range of experimental RDCs measured at different sites along the primary chain, even in the presence of high concentrations of denaturant. It was evident that RDCs were highly sensitive to differential order; RDCs measured in the immediate vicinity of prolines (more rigid) gave rise to larger RDCs than those measured in the vicinity of glycines (more flexible).[62] It was also noted that the volume of the side chain, or bulkiness, correlated with the magnitude of RDCs measured in IDPs.[216] It was also noted that the presence of local helical elements would align the N–H^N bond vector along the direction of the magnetic field, and therefore change the sign of the RDC as compared to the otherwise unfolded chain.[200,214] These observations pointed the

way toward the development of qualitative, and later more quantitative, procedures for interpreting experimentally measured RDCs in terms of conformational behavior.

While CSs can be predicted using only local conformational sampling information, RDCs depend on both local and long-range structure, complicating their analysis. In folded proteins, the angular expression shown in eq 7 can be resolved into two independent terms, corresponding to an alignment tensor (or order matrix) describing the orientational properties of the molecule with respect to the magnetic field, and the orientation of each internuclear vector relative to the principal axis system of this alignment tensor:

$$D = D_{max}A_{zz}[P_2(\cos\vartheta) + \eta/2\sin^2\vartheta\cos 2\varphi] \qquad (8)$$

A_{zz} and η define the axial and rhombic components of the alignment tensor, which report on the level and nature of the alignment of the global conformation, and $\{\vartheta\varphi\}$ the orientation of the internuclear vector. In folded proteins, RDCs then report on the orientation of internuclear bond vectors in distant parts of the molecule that can provide extremely powerful conformational constraints.[217] Such an approach assumes that the alignment tensor can be considered to be invariant for all copies of folded proteins in the statistical ensemble and that the tensor is independent of the conformational behavior of each internuclear vector; these assumptions are clearly not justifiable for unfolded proteins.

Analytical approaches have been developed that model the unfolded protein as a random flight chain or a semistiff polymer combined with a mean-field diffusion equation.[215,218−220] Such studies resulted in a closed-form analytical expression whereby the expected RDC can be expressed in terms of the first few terms of a Taylor expansion over $N^{-0.5}$, where N is the number of independent segments comprising the chain. The bell-shaped dependence and efficiency of RDC averaging as a function of the length of the chain were both correctly predicted using these models. We note that while affording insight, such approaches remain essentially homopolymeric, and therefore do not yet provide a framework for describing differential, sequence-dependent RDCs along the peptide chain, which remains the principal advantage of ensemble-based approaches.

To account for the steric alignment of each conformation in a statistical coil ensemble, it is necessary to numerically predict RDCs on the basis of the shape of each conformer, essentially by calculating A_{zz}, η, ϑ, and φ in eq 8 for every structure.[221] The use of large conformational ensembles to predict RDCs in this way generally reproduced both the local and the global distribution of experimental N–H^N RDCs measured along the length of the unfolded chain.[62,202,215] As described below, refinement of backbone sampling potentials is necessary to simultaneously reproduce RDCs measured between different nuclei in the peptide unit, especially when local conformational propensities deviate from the statistical coil used for the simulation. Simulations of homopolymers exhibit featureless bell-shaped curves, while those carried out using heteropolymeric sequences exhibited the heterogeneous distribution of RDCs observed in IDPs or denatured proteins (Figure 5). This observation confirms that the observed distributions depend strongly on the differential flexibility and conformational heterogeneity of the primary sequence.

2.7. Insight into RDC Averaging Properties from Numerical Simulation

The ability to predict average RDC values from a statistical coil ensemble thus establishes references for comparison and possible identification of deviations from so-called random coil behavior. Simulation can also be used to probe the conformational origin of experimentally observed fluctuations in terms of specific or persistent local or long-range structure in hypothesis-driven or trial-and-error type approaches. As an example of the application of such approaches, $N-H^N$ RDCs measured from the monomeric, intrinsically disordered form of α-synuclein, although in reasonable agreement with simulation using the statistical coil, supporting the model that the protein was lacking in secondary structure, were shown to be less well-reproduced in the N- and C-terminal regions of the protein.[222] A transient contact had been observed from experimental PREs between these regions that carry complementary net charges, and experimental RDCs were shown to be significantly better reproduced when the unfolded ensemble also contained this weak contact. This example serves as a reminder of the dual dependence of RDCs on both local structural propensity and long-range order. Figure 5 illustrates the modulation of the underlying bell-shaped RDC baseline in the presence of transient contacts between different regions of the chain, predicted from numerical simulation. RDCs in the region between the two disordered segments experiencing long-range contacts are almost entirely quenched due to this modulation.[112] Note that the simulations show predictions for omnipresent contacts between different regions of the chain. When exchange is fast on the chemical shift time scale, these effects would be expected to be linearly population-weighted with a bell-shaped baseline when transiently populated.

Importantly it was noted that the dependence on local and long-range conformational degrees of freedom throughout the molecule is responsible for the highly inefficient convergence properties of ensemble-averaged RDCs calculated from statistical coil ensembles, which require many tens of thousands of conformers before reaching a stable average.[62,223] This observation has clear implications not only for portability of ensemble representations, but also for subsequently designed restrained or selected ensembles (vide infra). The problem can be alleviated by using a smaller number of conformers if the protein were divided into short, uncoupled segments (local alignment windows) and calculating RDCs using alignment tensors for each overlapping segment.[224] However, if only local alignment is considered, all long-range order is lost, removing one of the richest sources of conformational information available from RDCs, and potentially leading to misinterpretation of long-range effects in terms of local sampling. Importantly, therefore, the combination of parametrized analytical forms of the underlying baseline modulations shown in Figure 5 and the efficiently averaged RDCs using local alignment windows was shown to successfully account for both local and long-range averaging (Figure 6).[112,223]

Ensemble-based simulation also demonstrated that RDCs between different pairs of spins in the peptide unit provide complementary information about the underlying conformational sampling. Combinations of multiple RDCs significantly raise degeneracy between different sampling regimes that can give rise to the measured data.[105] Figure 7 shows the dependences of different RDCs predicted for residues $i-1$, i, and $i+1$ on the $\{\phi\psi\}$ sampling of amino acid i.[82] $^1D_{NH}$ exhibits the expected sensitivity to αR sampling, but also degeneracy

between βS and βP, either for the amino acid of interest or for immediate neighbors. In general, dependence on sampling of immediate neighbors as well as the amino acid of interest complicates interpretation in terms of local conformational basins, and clearly demonstrates why no single RDC can represent a "read-out" value for conformational propensity. This also underlines the importance of using ensemble approaches to interpret experimental RDCs, where neighbor effects are explicitly taken into account.

2.8. Deconvoluting Conformational Factors Contributing to RDCs Measured in IDPs

Numerical simulations can provide remarkable insight into the complex origin of RDCs from unfolded systems. In a recent study, extensive *flexible-meccano* simulations were performed to identify the importance of local and global sequence determinants for the predicted RDCs.[225] On the left-hand side of Figure 8, the sampling of the central amino acid is uniform, from $-180°$ to $180°$ for both ϕ and ψ, but the remainder of the sequence samples the conformational potential of an alanine residue. On the right-hand side of the figure, the whole sequence samples Ramachandran space uniformly. It is clear that constraining the flanking residues to the more physical model significantly modulates the predicted RDCs, showing how the dependence on the $\{\phi,\psi\}$ sampling of the amino acid of interest propagates along the chain. Removal of explicit steric interactions does not change this result. The angular sampling available to the peptide of interest therefore depends on the backbone dihedral distribution of the neighbors, which projects the peptide of interest into a restricted cone of angular space. This study demonstrated that RDC prediction can be deconvoluted to only four contributing factors: the sampling of the amino acid of interest, the conformational sampling of the nearest-neighbors that defines the orientational space of the amino acid of interest, sequence-dependent scaling factors that correct the local alignment tensor (a polyglycine will be more flexible and therefore exhibit lower order than a polyproline), and a length-dependent baseline to incorporate the polymeric nature of the unfolded protein, which in the case of transient tertiary contacts will be modulated by these in the manner shown in Figure 5. A database of combinations of triplets of amino acids, combined with rigidity corrections dependent on the composition of the six neighbors on either side, defines to a very good approximation (Figure 9) expected values of RDCs in unfolded states (*seq2rdc* algorithm). While ensemble averages are still generally used, this demonstration shows that it is possible to develop a look-up table for expected RDCs from an unfolded polymer that provides both important physical insight and a significant improvement in efficiency of calculating random coil RDC values.

2.9. Comparison with Data Sets Comprising Multiple Experimental RDCs and CSs

In an early example of the use of RDCs for the study of unfolded states, a combination of multiple RDC types was used to investigate the conformational behavior of urea-denatured ubiquitin. $N-H^N$, $C'-H^N$, $C'-C^\alpha$, $C^\alpha-H^\alpha$, H^N-H^N, and H^N-H^α RDCs clearly contradicted statistical coil models, but were shown to be in agreement with potential energy basins with higher propensities to sample more extended regions of Ramachandran space (βS and βP) and lower propensities to sample helical regions (αR).[105] This example is particularly revealing, because $N-H^N$ RDCs alone were in general
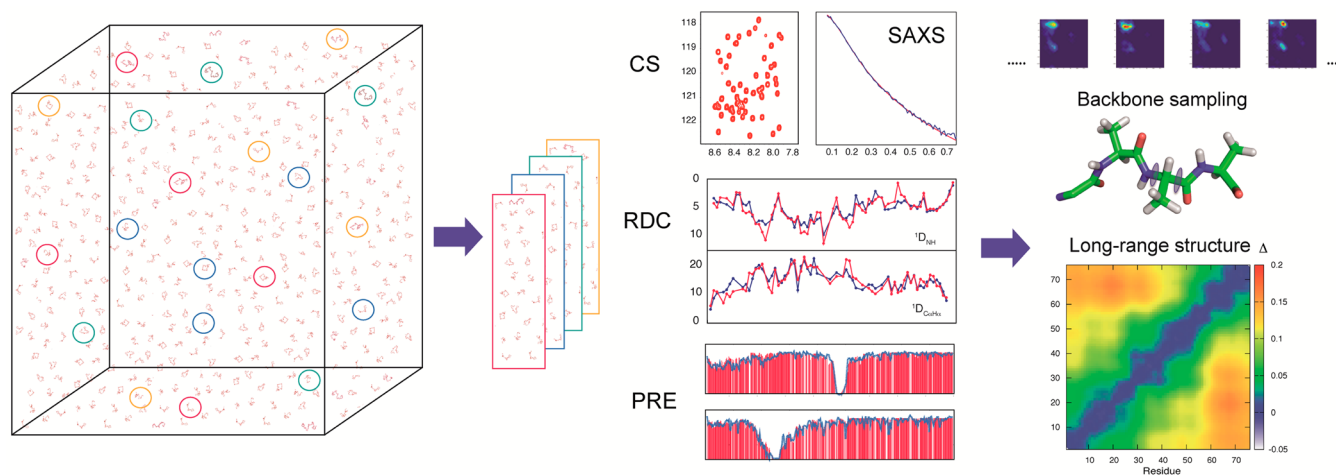
Figure 11. Overview of the *flexible-meccano*/ASTEROIDS approach for deriving representative ensembles of IDPs. An initial pool of structures is created using *flexible-meccano* that is made sufficiently large to maximize sampling of conformational space. The genetic algorithm ASTEROIDS is used to select equivalent, representative subensembles in agreement with experimental data such as CSs, RDCs, PREs, and SAXS data. The selected subensembles are subsequently analyzed in terms of backbone conformational sampling and long-range structure.

agreement with the statistical coil model, underlining the importance of measuring the other RDC types to extract quantitative information.

Multiple RDC types (N−H^N, C′−H^N, C′−C$^\alpha$, and C$^\alpha$−H$^\alpha$) were also measured from the C-terminal domain of Sendai virus nucleoprotein.[226] ^{13}C CSs showed a high (but not complete) α-helical propensity in the phosphoprotein binding site. To examine the details of the helical sampling in this protein, a systematic approach was developed, postulating the presence of all possible helical elements. *Flexible-meccano* ensembles were calculated for each of 154 helical lengths (with a minimum of four residues) between the beginning and end of the helical region defined by the CSs. Inspection of the predicted RDCs from each of these ensembles already revealed that "dipolar waves" of RDC values along the sequence varied significantly depending on the length and termination of the helix.[227] These oscillations, due to the orientation of the alignment axis relative to the helix, allowed for a systematic combination of helical ensembles and their populations, in increasingly complex equilibria, until the data were adequately reproduced, and no significant improvement was achieved with additional helices.[226] This approach demonstrated that all experimental data could be reproduced from only three significantly populated helical elements, all stabilized by N-capping interactions, in exchange with an unfolded population (Figure 10). Importantly, this analysis demonstrates that CSs and RDCs are complementary in their ability to distinguish cooperative and noncooperative secondary structure formation. While CSs report essentially on the conformational properties of an amino acid and its neighbors, so that a peptide in a long helix can have the same CSs as the central amino acid of a helical tripeptide, RDCs report on the length of the helices present in the conformational equilibrium, as well as their capping characteristics, and can therefore distinguish these scenarios. CSs, on the other hand, provide a more robust measure of the absolute level of secondary structure, as opposed to RDCs whose absolute values depend on the level of alignment.

Application of the approach to a homologous domain from Measles virus, this time combining CSs and RDCs, resulted in a similar helical ensemble in the interaction site, comprising four helical elements, all stabilized by amino acids favoring N-

capping interactions.[228] Similarly, N−H^N, C′−H^N, C′−C$^\alpha$, and C$^\alpha$−H$^\alpha$ RDCs measured from the intrinsically disordered transactivation domain of p53 were shown to agree with the presence of a single turn α-helical element in the MDM2 binding site, also stabilized by an N-capping interaction, and populated to approximately 30%.[229] N-capping appears to be a general mode adopted by IDP sequences to place-mark nascent helical elements in otherwise unfolded chains.

In all of the above cases, the analysis of RDCs was guided by the observation from ^{13}C CSs that a helical element was present. More generally, however, there is a need for the development of representative ensembles of conformers directly from experimental data, a challenge that has motivated considerable interest over recent years.[15,18,20,142,230]

3. ANALYTICAL APPROACHES TO DESCRIBING THE FREE-ENERGY LANDSCAPE

The development of ensemble representations of IDPs requires the resolution of a number of important problems. These include the assurance of a broad enough sampling of the available conformational space, the optimization of the number of structures present in the ensemble average, the combination of different experimental restraints, and the integration of appropriate dynamic averaging time scales of each experimental parameter within the conformational ensemble. More generally, as mentioned above, overfitting, under-determination, and validation of ensembles remain critical challenges.

3.1. Molecular Dynamics-Based Approaches

One approach to the development of such ensembles is to use restrained molecular dynamics (MD) or restrained replica-exchange MD (REMD) calculations. In restrained MD, additional terms are incorporated into a physical potential energy function to bias the conformational sampling, to drive a single or, more appropriately in the case of IDPs, multiple copies of the protein into conformations in overall agreement with experimental data.[113,152,153,231−233] The ability to predict CSs on the basis of the local geometry and atomic proximity as is the case for the program CamShift allows for CSs to be calculated rapidly at each step of the trajectory, making such approaches feasible for CSs. Similarly, the steric molecular
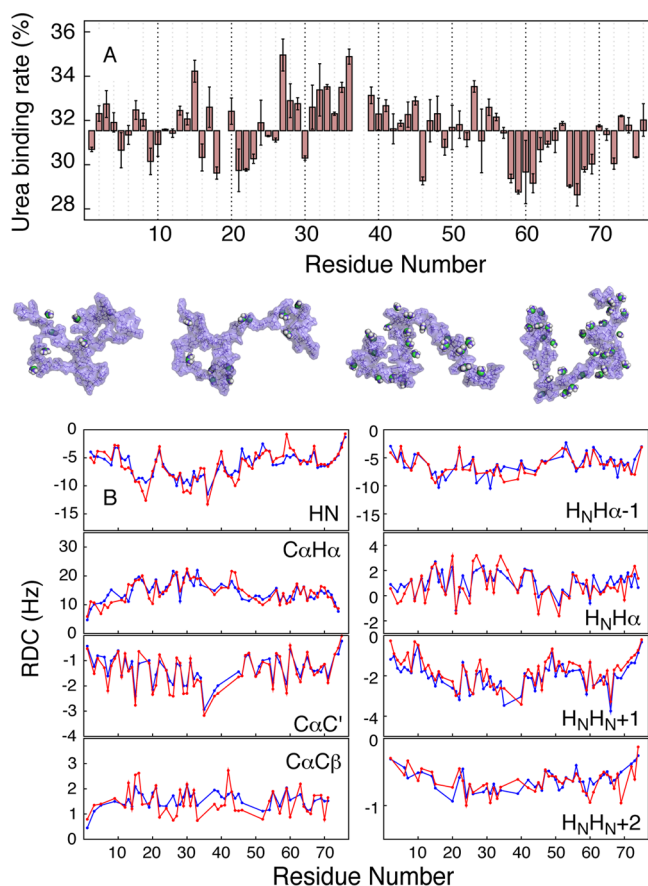
Figure 12. Modification of the *flexible-meccano* algorithm to include explicit sampling of urea molecules on the protein backbone. (A) Application of the explicit urea sampling to urea-denatured ubiquitin showing the average urea binding rate along the sequence from 50 ensembles selected by ASTEROIDS on the basis of multiple RDCs. Conformers constructed using *flexible-meccano* with explicit urea sampling with different degrees of saturation (10−40%) are shown in sticks representations. (B) Comparison of experimental RDCs (red) and back-calculated RDCs from selected ASTEROIDS ensembles (blue). The selections were carried out from a pool of ubiquitin conformers carrying explicit urea sampling with all possible degrees of saturation. Reprinted with permission from ref 287. Copyright 2012 American Chemical Society.

alignment tensor of each copy in the ensemble can be calculated at each step of the simulation,[155] thereby allowing for the effect of the instantaneous dipolar coupling on the change in shape of the protein. Assuming that correlation times for local and long-range motions have been estimated, PREs then depend only on the geometric properties of the molecule, so that calculation of the appropriate restraint energy term is again trivial.[234]

MD-based approaches most closely resemble the commonly used simulated annealling protocols used for structure determination of folded proteins, and are therefore the most accessible to NMR spectroscopists and structural biologists in general. However, they may not be the most appropriate for IDPs, as it remains unclear how the introduction of nonphysical restraints into the potential energy force field affects the completeness of phase space sampling, which is of course critical for such under-determined systems. An accurate representation of the Boltzmann-weighted distribution of conformers can only be faithfully reproduced if sampling is as

complete as possible. Using REMD, sampling is enhanced by exchanging between copies of the biomolecular system that evolve under more or less constrained conditions (for example, at different temperatures) using Monte Carlo criteria.[235,236] Many groups have applied such techniques to the study of IDPs by directly restraining against experimental data or comparing the resulting ensembles to experimental NMR data.[154,191,237−242] A number of recent articles have addressed the statistical distribution produced by restrained replica-exchange ensemble simulations[243−245] and identified the conditions under which restrained MD is formally consistent with the statistical distribution produced using more rigorous maximum entropy principles.[246]

Restraint-free MD does not suffer from the combination of empirical and physical force field terms, and can be directly compared to experimental data. A great deal of valuable insight has been gained from MD simulations of protein folding and unfolding or denaturation, and although this paradigm is strongly related to the simulation of IDPs, this work will not be covered in more detail here.[247−252] MD simulations of Abeta40 and Abeta42, peptides associated with Alzheimer's disease, were also compared to cross relaxation, CS, and scalar coupling data.[237,253] More recently, a long (200 μs) fully solvated MD simulation was used to study the acid-unfolded state of the protein ACBP, comparing NMR spin relaxation rates and long-range contacts derived from experimental PREs.[174] Even for such long time-scale simulations, sampling remains a problem, because the conformational space sampled by a single MD trajectory is likely to be restricted as compared to the highly heterogeneous statistical ensemble, so that it still remains challenging to ensure efficient sampling of the phase space accessible to the unfolded state. The behavior of unfolded chains depends strongly on protein−solvent interactions, which means that the inaccuracies of currently accessible water models to reproduce solvent behavior, including fluctuating weak interactions such as solvent−solute hydrogen bonding, may compromise the accuracy of MD simulations of IDPs. Nevertheless, there is a great deal to learn from such simulations, not least because this is one of the few approaches that can provide insight into the characteristic time scales of IDP dynamics, and the molecular mechanisms controlling protein folding upon binding.[241,254−256] Comparison with experimental data will no doubt contribute to the already active development of more accurate force fields that are applicable for simulations of IDPs.[257−259]

The statistical sampling problem can also be addressed using enhanced sampling approaches that directly modify the applied force field, such as metadynamics,[260] where the potential energy surface is modulated as a function of the memory of the previous conformational search, forcing the system to explore conformations that have not yet been visited. Such methods, which can be usefully combined with replica-exchange,[261] have been applied to highly dynamic peptides and proteins.[262] The resulting free-energy surfaces, derived from the inverse of the applied potentials, can be used to predict experimental NMR data to gauge the accuracy of the resultant sampling.[263,264] Similarly, a recent study incorporated CSs as collective variables to guide the metadynamics calculation.[265] Another approach to enhanced sampling uses accelerated molecular dynamics (AMD),[266−268] which modulates the potential energy surface in a very simple way, scaling the potential energy so as to reduce the barriers between substates, but crucially retaining the rugosity of the underlying surface. This increases the
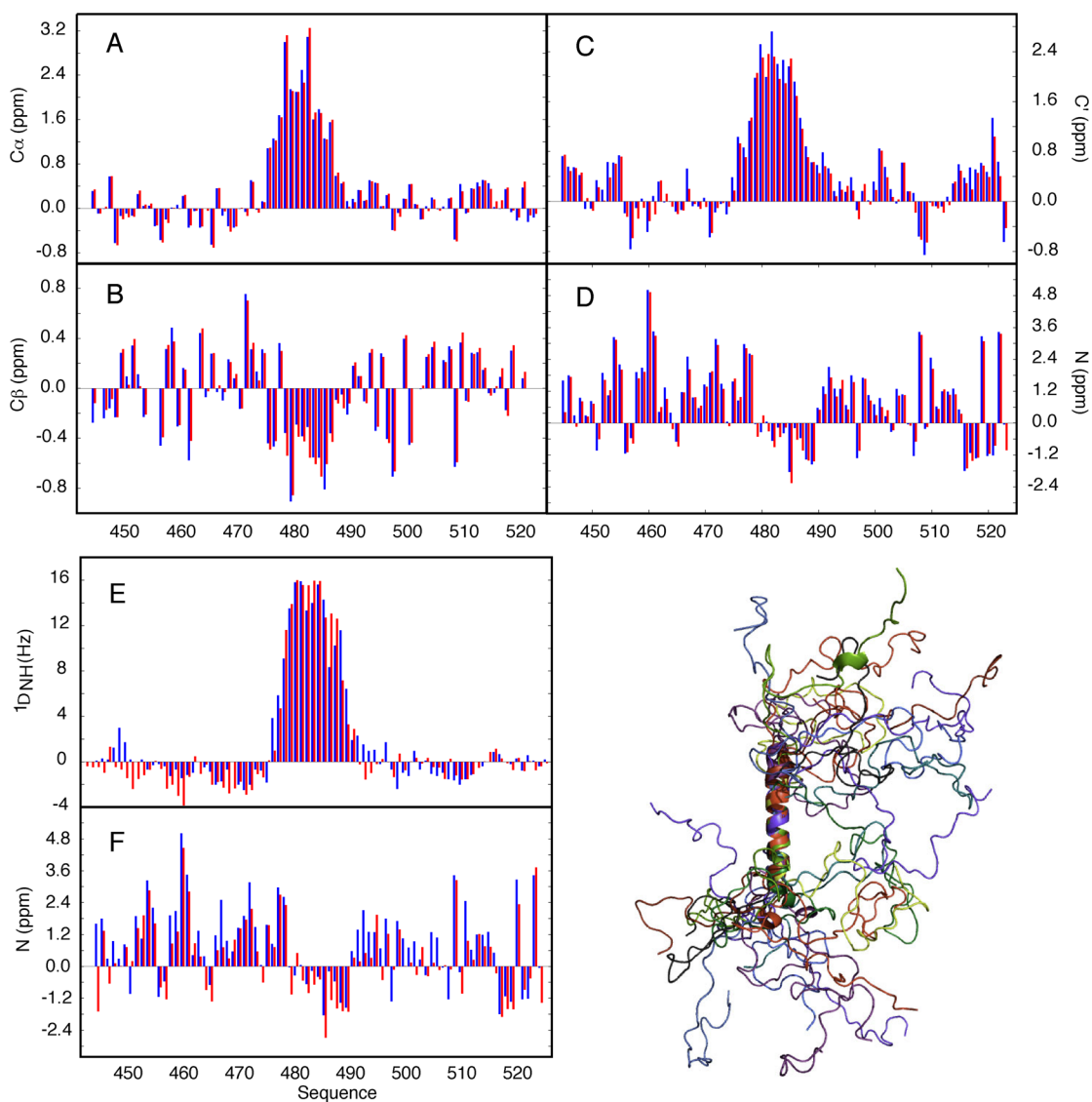
Figure 13. Selection of representative ensembles using ASTEROIDS on the basis of experimental CSs. (A–D) Agreement between experimental secondary chemical shifts (blue) in the C-terminal domain, N_{TAIL}, of Sendai virus nucleoprotein and back-calculated secondary chemical shifts (red) from an ASTEROIDS-selected ensemble comprising 200 conformers. (E) Reproduction of independent data not used in the ASTEROIDS selection. Comparison of experimental $^1D_{NH}$ RDCs (blue) with back-calculated RDCs from a *flexible-meccano* ensemble of 50 000 conformers constructed using the backbone dihedral angles obtained on the basis of experimental $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, and ^{15}N chemical shifts (red). (F) Comparison of experimental ^{15}N secondary chemical shifts (blue) and secondary chemical shifts back-calculated from the ASTEROIDS ensemble selected on the basis of $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ (red). Representative conformers of N_{TAIL} from the ASTEROIDS ensembles are shown in cartoons representations. Reprinted with permission from ref 296. Copyright 2012 American Chemical Society.

probability of transitions between existing energy minima and allows for highly efficient sampling of conformational space. AMD has been combined with NMR, predominantly for the study of long time scale motions in folded proteins,[269] but also for the study of β-turn regions in the K18 domain of Tau protein,[111] and the proline-rich domain of p53 transactivation domain.[229] In both cases, deviation from random coil RDC values was observed experimentally, and the statistical coil sampling for the regions of interest was replaced by AMD-derived {ϕ,ψ} distributions. Average RDCs using these distributions reproduced measured values within experimental error. This again indicates that enhanced sampling of the available potential energy surface can provide further insight when statistical coil sampling fails, although in general classical MD force fields seem to describe amino-acid specific

differences less well than statistical coil models, most notably overestimating α-helical populations.[128,257,259]

3.2. Ensemble Selection Approaches To Mapping the Potential Energy Surface

An alternative approach has been developed over recent years that oversamples conformational space in as unbiased a way as possible, and then selects representative subensembles from this sampling that are in agreement with experimental data. The advantages of such approaches are that the underlying conformational sampling available to the chain is established independently of the experimental data, the extent of the sampling can be tested for completeness at this stage, and the delineation of regions of conformational space that can give rise to the experimental data occurs in a second step. The first of such approaches to be developed for unfolded proteins is
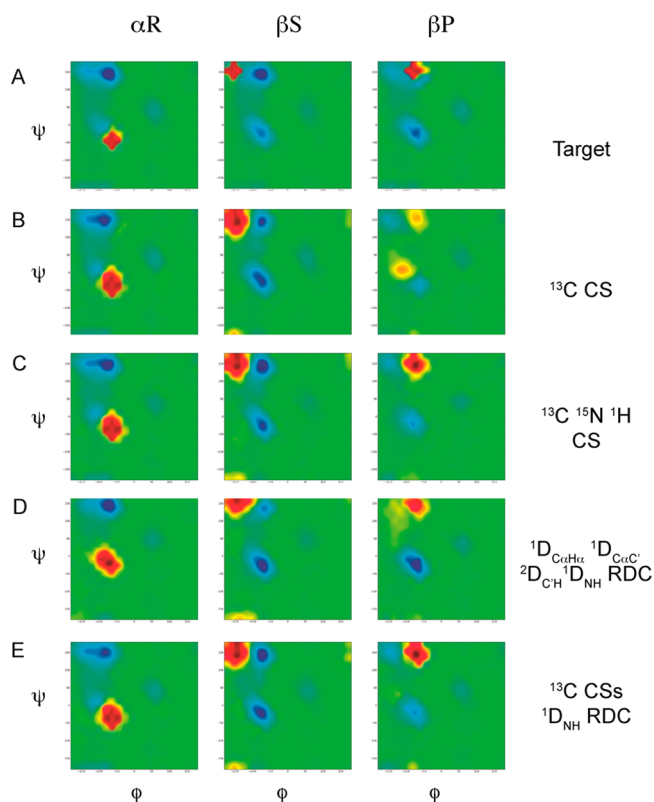
Figure 14. Systematic assessment of the accuracy to which experimental CSs and RDCs can map site-specific conformational sampling in IDPs. (A) An ensemble of 20 000 conformers of a 60 amino acid protein of arbitrary sequence was generated using *flexible-meccano* with three distinct five-residue regions oversampling α-helix (αR), β-strand (βS), and the poly proline II region (βP). Predicted CSs and RDCs from this ensemble were used as target for ASTEROIDS ensemble selections. The Ramachandran plots show difference maps between the statistical coil sampling and the sampling obtained by ASTEROIDS in the three regions. (B) If only ^{13}C chemical shifts are used, the populations of αR and βS can be accurately determined, while degeneracy is observed between the βP region and the upper αR region. (C) The inclusion of the experimentally more volatile ^{1}H$_N$ and ^{15}N CSs in the ensemble selection raises this degeneracy. (D) Conformational sampling determined on the basis of RDCs alone shows degeneracy between the β-extended regions. (E) By combining CSs with RDCs, the ensemble selection accurately defines all (αR, βS, and βP) populations. Reprinted in part with permission from ref 82. Copyright 2012 American Chemical Society.

ENSEMBLE,[270,271] which uses the statistical coil generator TraDES,[272] and diverse MD-based protocols to sample available conformational space. The selection protocol, which assigns weights to the initial pool of conformers in a way that best reproduces the experimental data, has been applied to the study of the unfolded form of drkN SH3 domain,[158] the cell cycle IDP Sic1,[273] and protein phosphatase 1 regulators.[274,275] Recently, unbiased MD simulations were combined with NMR restraints, to derive an ensemble of conformations for phosphorylated arginine-serine repeats, which were in best agreement with experimental N−H^N, C^α−H^α, C^α−C′ RDCs and ^{3}J$_{HN−Hα}$ scalar couplings.[110] The selection was achieved using a combination of a Monte-Carlo search and exhaustive scanning. Thirty conformations were estimated to be optimal to avoid under- or overfitting of the NMR data. The ensemble was analyzed by calculation of the Jensen−Shannon divergence and

partial least-squares analysis. The ensembles were compared to ^{1}J$_{Cα−Hα}$ and ^{1}J$_{Cα−Cβ}$ couplings together with chemical shifts for the backbone and ^{3}J$_{NCγ}$ and ^{3}J$_{C′Cγ}$ couplings for the side-chain orientations.

A similar approach was developed by Stultz et al. where the prior pool of structures is assembled using REMD simulations of extended, or fragmented, pieces of the protein that are then reassembled to enhance sampling diversity.[276] This prior pool is then clustered on the basis of energetic and topological considerations into a smaller library of a few hundred conformers that is used to represent the available sampling, and weights are again assigned to these conformers as a function of agreement with experimental data. The approach was applied to the study of a domain of Tau protein[277] and to monomeric α-synuclein.[278] In a more recent study of α-synuclein, in addition to monomeric conformers, the initial pool was also populated with trimeric and tetrameric conformations obtained by threading the primary sequence of α-synuclein through crystallographic structures of myosin and ferritin.[279] These conformers represented over one-third of all populations in the final selection against experimental CSs, despite the fact that they would be effectively spectroscopically invisible under the experimental conditions. Although it is not possible to disprove the presence of such species (in the absence of SAXS data, for example), there is no experimental evidence for their presence. This illustrates one of the risks involved in ensemble selection, where the nature of the selected ensemble depends entirely on the reliability of the prior sampling. We note that similar sample and select type approaches have been developed for the ensemble representation of RNA molecules and multidomain proteins on the basis of both NMR and SAXS data.[280−285]

The ASTEROIDS (A Selection Tool for Ensemble Representation Of Intrinsically Disordered States) approach uses the *flexible-meccano* statistical coil approach to establish the prior sampling of phase space available to an unfolded model of the protein. The experimental data are not used at this stage, the aim being to sample the full energy landscape accessible to an ultrasimple statistical coil model for the specific primary sequence in the absence of bias either from experimental data or from molecular dynamics force fields. Experimental restraints are only used in the second, selection step, performed by the ASTEROIDS algorithm,[112,223,286] which delimits ensembles sampling conformational space that can give rise to the experimental data (Figure 11). ASTEROIDS uses a genetic algorithm, where one conformer is represented by a gene, the size of the chromosome (n) is constant, and generations are obtained by random selection, mutations, and crossings of conformers. Populations of different conformers are not optimized, so that $p_i = 1/n$ for all substates i (the population of such a state is augmented by selection of similar substates with the same characteristics relative to the experimental data).

Every statistical coil definition carries bias depending on the specific protocol used to create the model, so that care must be taken to test the effect of this bias. Dependence of subsequent selections on the initial *flexible-meccano* sampling can been tested by modifying the initial sampling and comparing the characteristics of the final selection. Because of the low probability of randomly producing significant amounts of secondary structure directly from a statistical pool, iterative cycling of biased *flexible-meccano* sampling and ASTEROIDS selection can be used to modify the amino-acid specific potentials using previous selections until agreement is achieved
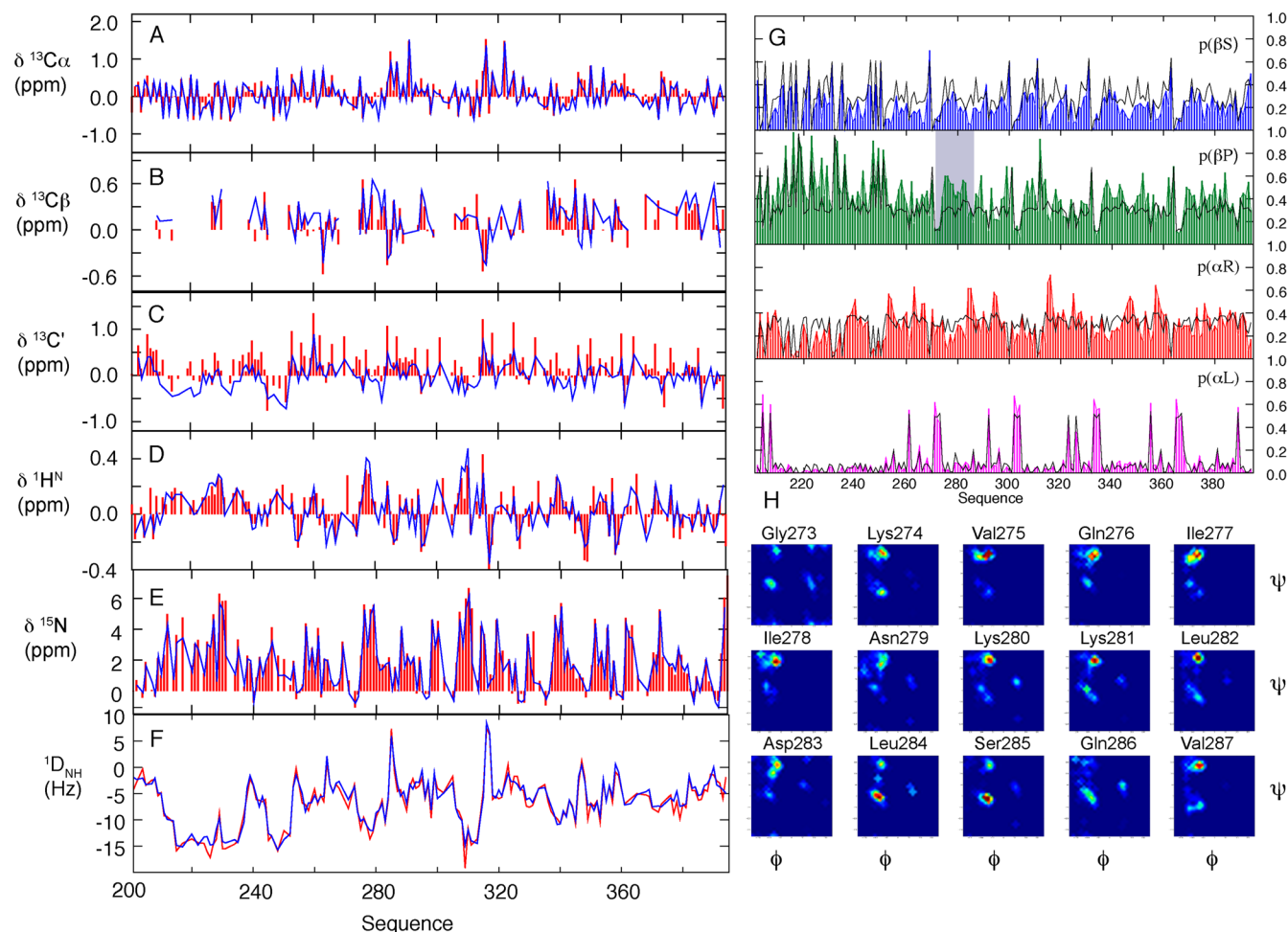
Figure 15. Site-specific conformational sampling in the K32 construct of Tau protein derived from ASTEROIDS selections against experimental CSs and RDCs. (A−F) Agreement between experimental CSs and RDCs (red) and those back-calculated from selected ASTEROIDS ensembles (blue). (G) Populations in the four regions of Ramachandran space defined in Figure 2. The populations were derived from the selected ASTEROIDS ensembles on the basis of CSs and RDCs (blue, green, red, and magenta) and compared to the populations in the statistical coil library (black). (H) Ramachandran plots of the region G273−V287 of K32 showing enhanced βP sampling. Reprinted in part with permission from ref 298. Copyright 2014 Cell Press.

between experimental and back-calculated data. The *flexible-meccano*/ASTEROIDS approach is used as a means to map the potential energy landscape sampled by the protein backbone, and as such repetition of the procedure determines solutions that are not unique, containing different structures for each selection. The backbone sampling characteristics, however, do not vary between selections, which are converged and therefore unique in terms of conformational substates and their populations. This would not be the case if sampling were insufficiently complete. It is also important to note that the conformational ensembles are used to map complex probability distributions in terms of a distribution of discrete states. Although ensemble approaches are optimized to reproduce the true conformational equilibrium, the individual structures comprising the ensembles depend on the presence of all other members of the ensemble, which can only be meaningfully interpreted as a whole.

3.3. Application of Ensemble Descriptions to Protein Denaturation in the Presence of Urea

The behavior of IDPs is defined to a large extent by their interaction with solvent molecules, so that while urea-denatured proteins are not intrinsically disordered, the description of their

conformational behavior in solution is of direct relevance for understanding the stability of IDPs and unfolded proteins in general. In this context, the ASTEROIDS approach was initially applied to urea-denatured ubiquitin for which a large set of RDCs had been measured from throughout the peptide chain.[105] The potential energy basins for each amino acid are derived from the simultaneous analysis of multiple complementary experimental RDCs measured on each peptide unit. This identified local conformational sampling properties of the protein, highlighting modifications of backbone sampling of charged or polar amino acids, in particular threonine, glutamic acid, and arginine.

This approach highlighted two important points: the first being that for alignment windows of 15 amino acids 200 distinct conformations are required to reach RDC convergence. It was shown that the use of 20 conformers still reproduces the experimental data, but the conformational sampling can be incorrect because the average is not yet stable. Consequently, the attractive idea of reducing the number of conformers to a minimum that reproduces the experimental data can be dangerous as it may result in a poorly converged solution. Within this range, the optimal number of structures is determined by repeating the selection analysis and comparing
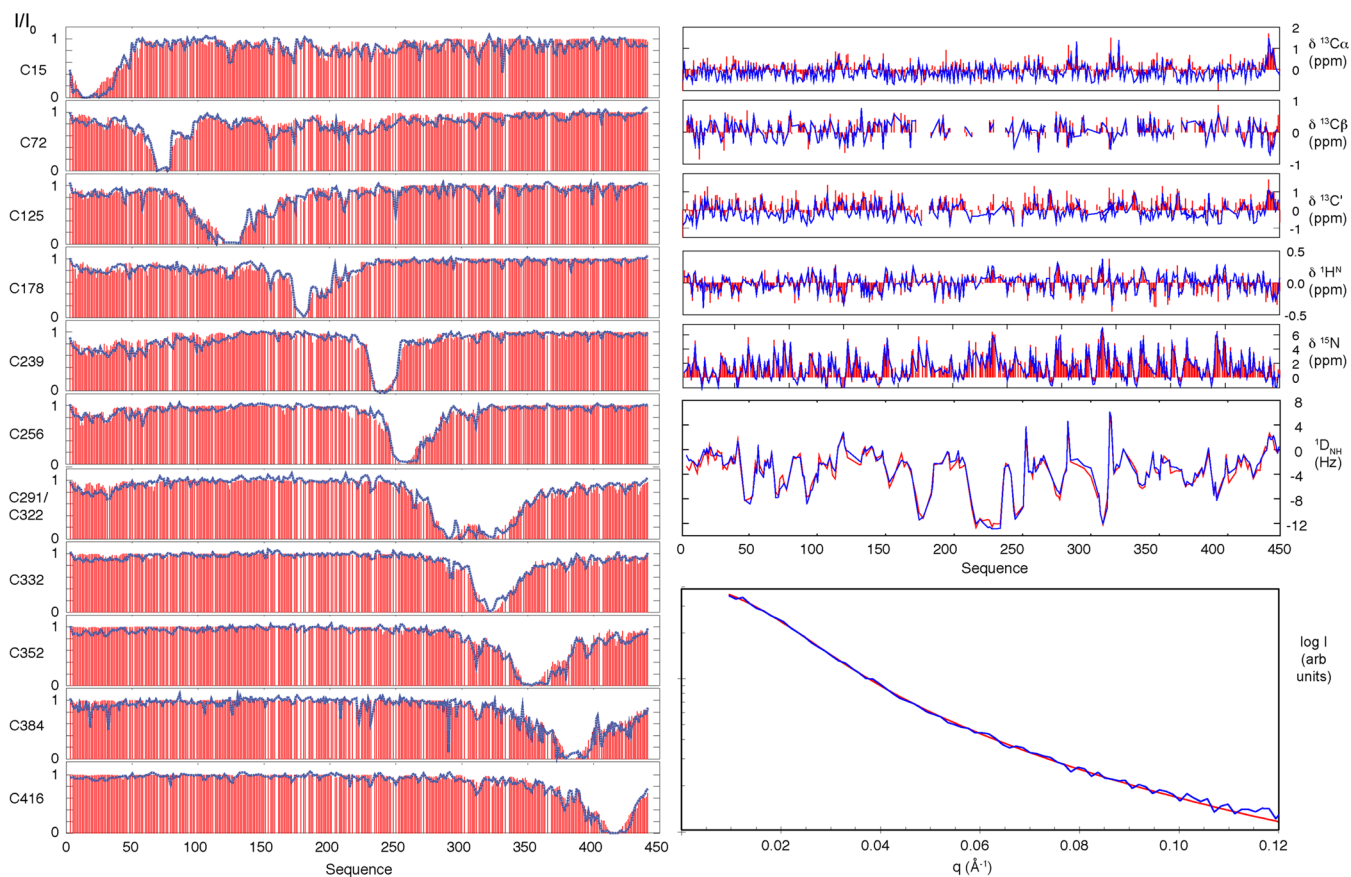
Figure 16. Structural ensemble description of Tau from experimental PREs (left-hand panels), CSs (top right-hand panels), RDCs (middle right-hand panel), and SAXS data (bottom panel). In all panels, the experimental data (red) are compared to those back-calculated from selected ASTEROIDS ensembles (blue). Reprinted in part with permission from ref 298. Copyright 2014 Cell Press.

to independent experimental data that have not been included in the target function. This emphasis on the predictive capacity of ensemble descriptions is crucial for testing the accuracy and robustness of the solutions. Second, the demonstration that the modulation of the RDC profile due to long-range information can be combined with the efficient local alignment approach (Figure 6), to simultaneously account for both local and long-range order contributions within the same ensemble average, paves the way to combine RDCs with, for example, PREs and SAXS restraints (vide infra).

The application of ASTEROIDS to the study of denatured ubiquitin was taken one step further by modifying the *flexible-meccano* approach to construct full-atom conformational ensembles, including side-chain conformations derived from rotamer libraries, and random binding of explicit urea molecules to the protein backbone that modifies the potential energy basins of each amino acid. The ASTEROIDS selection then determines the number of bound urea molecules over the ensemble, and identifies the sites that have the highest propensity for binding (Figure 12).[287] The direct-binding model of urea to the protein backbone is thus shown to be compatible with all available experimental data. Between 30% and 40% of the backbone peptide groups bind urea, in agreement with independent results from a model-free analysis of small-angle neutron and X-ray scattering data.[288] The trend of urea-binding reveals a higher level in the central part of the protein, resembling independent results derived from chemical shift mapping.

3.4. Combining Local and Long-Range Effects for Ensemble Interpretation of RDCs

The combination of RDCs with PREs presents a more difficult problem, as a significant modulation of RDCs is expected due to the existence of long-range intrachain contacts (Figures 5, 6). It was subsequently shown, again using the ASTEROIDS approach, that RDCs from the protein α-synuclein were more accurately predicted using a combination of local alignment and RDC modulation due to long-range contact information derived from PREs, as compared to a prediction based on local structure alone.[112] The same study also demonstrated that the inclusion of an explicit term describing the dynamics of the MTSL-carrying side chain better reproduced passive PREs not included in the selection. ASTEROIDS ensembles were recently derived from PREs and RDCs to study the modulation of intrinsic helical propensity in the protein ACTR (activator for thyroid hormone and retinoid receptors).[289] Eight amino acid mutations were used to vary helical populations and to study the coupling between secondary and tertiary structure formation. ASTEROIDS ensembles derived from PRE data were also used to describe changes in long-range order in acid-unfolded ACBP induced by single-point mutations previously identified to be important for the folding of the protein.[290]

3.5. Ensemble Interpretation of CSs for the Study of Local Conformational Propensity

CSs are extremely sensitive to backbone conformation, and the possibility of exploiting this sensitivity to map the potential energy surface of IDPs is particularly attractive considering that
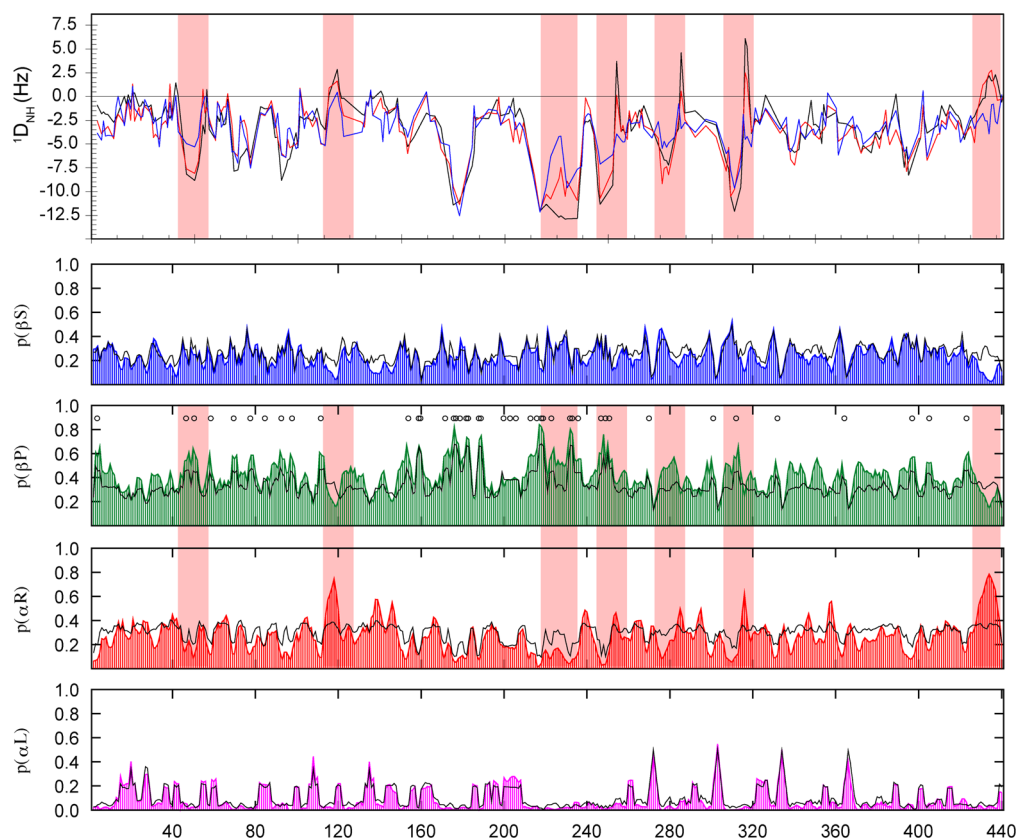
Figure 17. Analysis of selected ASTEROIDS ensembles of Tau protein. (Top) Comparison of experimental $^1D_{NH}$ (black) with RDCs obtained from a statistical coil ensemble (blue) and with RDCs back-calculated from an ASTEROIDS ensemble of Tau selected on the basis of CSs only (red). (Bottom panels) Populations in the four regions of Ramachandran space defined in Figure 2. The populations were derived from the selected ASTEROIDS ensembles on the basis of CSs, RDCs, PREs, and SAXS data (blue, green, red, and magenta) and compared to the populations in the statistical coil library (black). The black circles indicate positions of proline residues. The shaded bars indicate regions for which the derived conformational sampling is significantly different from statistical coil, and which coincide with improved agreement between ASTEROIDS predicted $^1D_{NH}$ RDCs as compared to statistical coil. Reprinted in part with permission from ref 298. Copyright 2014 Cell Press.

the determination of (backbone) CSs is a prerequisite for any NMR study. This possibility becomes even more pertinent with the recent developments within the field of in-cell NMR[291−294] that allow the observation of proteins within their physiological environment, but where additional structural parameters, besides CSs, cannot be easily measured.[295] The advantage of CSs is that they depend in a different, complementary manner on the backbone conformational sampling (Figure 3) and therefore, in principle, allow a mapping of the potential energy surface in a site-specific manner.

ASTEROIDS was also applied to experimental CSs, to study the conformational sampling of the intrinsically disordered tail of Sendai virus nucleoprotein at atomic resolution.[296] Starting from a pool of statistical coil conformers containing negligible secondary structure, an iterative approach was developed where ensemble selections were carried out followed by regeneration of the pool using the site-specific potentials derived from the selected ensemble in the previous iteration. Five iterations of this kind were necessary to achieve convergence with respect to the experimental data in the case of Sendai virus nucleoprotein that contains a molecular recognition element populating up to 75% α-helical conformations. Excellent agreement was obtained between the experimental CSs and those back-calculated from selected ensembles each comprising 200 conformers (Figure 13). The site-specific conformational sampling of the tail of Sendai virus nucleoprotein derived from the chemical shifts

revealed a slightly enhanced propensity to populate PPII conformations as compared to the statistical coil populations employed within *flexible-meccano*.

It is important to consider the relative weights of each CS type in the ensemble selections to avoid overfitting of one data type as compared to the other. One way of avoiding overfitting, and of testing the consistency of ensembles derived from CSs, is to back-calculate independent parameters that were not used in the ensemble selection. In the case of the tail of Sendai virus nucleoprotein, the structural ensembles derived from CSs alone were capable of reproducing experimentally measured $^1H-^{15}N$ RDCs both within the α-helical molecular recognition element as well as in the flanking regions testifying to the predictive nature of the CS-derived ensembles (Figure 13).

3.6. Calibration of Ensemble Mapping of Potential Energy Surfaces − How Well Can We Do?

Although the conformational dependences of individual CSs and RDCs could be predicted (see Figures 3 and 7), no analytical framework for the determination of the potential energy landscape of unfolded proteins at amino acid specific resolution was yet available. It therefore remained unclear how accurately CSs and RDCs could be used to define backbone conformational sampling in IDPs. Experimental data are often combined to best restrain the ensemble, but without calibration it is unclear what resolution can be achieved, or which regions of Ramachandran space can be differentiated on the basis of the
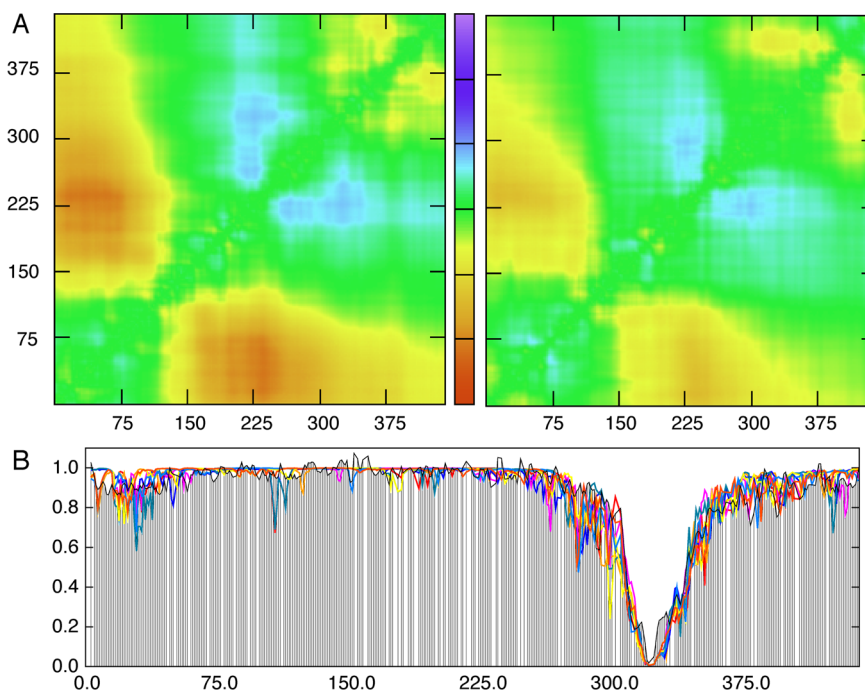
Figure 18. Analysis of long-range interactions in selected ASTEROIDS ensembles of Tau protein. (A, left) Contact map derived from selected ASTEROIDS ensembles of wild-type Tau on the basis of experimental PREs (see Figure 16, left panel) showing a transient long-range contact between the N-terminus of Tau and the proline-rich region. (A, right) Contact map derived from selected ASTEROIDS ensemble of pseudophosphorylated Tau (S199E, S202E, T205E, T231E, S396E, S404E) on the basis of PRE data showing a release of long-range interactions as compared to wild-type Tau. (B) Cross-validation of a single PRE data set (C322) for wild-type Tau. Gray bars indicate the experimental data, while the different colors represent the calculated data from eight independent runs of the ASTEROIDS algorithm for ensembles comprising 200 structures. The selected ensembles reproduce the "passive" data as well as when these data are used actively in the ensemble selection, indicating that an ensemble size of 200 conformers is suitable and that the approach is not significantly prone to overfitting. Reprinted in part with permission from ref 301. Copyright 2011 American Chemical Society.

available data. Having access to the algorithms described above, a recent study performed a systematic analysis of the ability of NMR data to map the conformational landscape of disordered proteins, identifying combinations of RDCs and CSs that can be used to raise conformational degeneracies inherent to the different data types.[286]

Conformationally biased disordered ensembles of a model protein were generated using the *flexible-meccano* algorithm so that three different regions of the protein obeyed specific sampling properties (enhanced αR, βS, and βP as compared to statistical coil). Synthetic, noise modulated, RDC, and CS data were simulated from these ensembles and used as target data for the ASTEROIDS algorithm, with subensembles selected from a prior distribution of 20 000 structures calculated using the statistical coil model. The results (Figure 14) identify conformational propensities that can be distinguished on the basis of experimentally available data. As was already predicted in Figure 3, although available backbone ^{13}C chemical shifts can be used to accurately measure the populations of βS and αR regions of Ramachandran space, the βP region is degenerate with average values predicted for random statistical coil sampling as well as a peripheral region to αR. Although prediction accuracy is lower ^{15}N and ^{1}H^{N} shifts raise this degeneracy. Similarly, even combinations of multiple commonly measured RDCs (N−H^{N}, C′−H^{N}, C′−C$^{\alpha}$, C$^{\alpha}$−H$^{\alpha}$) are unable to distinguish between βS and βP regions, while the extended regions can be distinguished from helical bias. Importantly, however, the combination of backbone ^{13}C (and ^{15}N and ^{1}H^{N}) CSs and ^{1}D$_{HN}$ RDCs raises both of these inherent degeneracies, and represents a tractable solution that is accessible for many experimental studies, while remaining robust with respect to experimental noise, spectral calibration, and prediction error. This study of the ability of NMR data to map conformational potentials provides an essential calibration for the *flexible-meccano*/ASTEROIDS approach. Although alternative algorithms will likely differ slightly in the detail of the discrimination of conformational space from experimental data, these differences are not expected to be major.

3.7. Application to IDPs Involved in Neurodegenerative Disease − Tau and α-Synuclein

This calibrated approach was recently applied to a number of experimental systems, including to the study of three constructs of increasing length of the protein Tau, an IDP that is implicated in the development of Alzheimer's disease. Initially, local conformational sampling of the 130 amino-acid K18 domain, which contains four highly homologous repeat sequences, was analyzed using ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$, ^{13}C′, ^{15}N, and ^{1}H^{N} CSs and ^{1}D$_{HN}$ RDCs,[111,286,297] and the longer form (K32),[298] containing 198 amino acids stretching from 200 to 397, using additional ^{2}D$_{CHN}$ and ^{4}D$_{HNiH\alpha i-1}$ RDCs. K32 encompasses K18, as well as the second proline-rich region (PRR; residues 198−244),[299] whose structural propensities are particularly interesting as this region is targeted by kinases and involved in assembly of tubulin into microtubuli.[300] Local conformational sampling can be represented in terms of free-energy maps of each Ramachandran plot (Figure 15), or more accessibly in terms of populations of the four main regions of Ramachandran space presented in Figure 2. Comparison with the populations of these regions taken from the initial pool reveals whether the
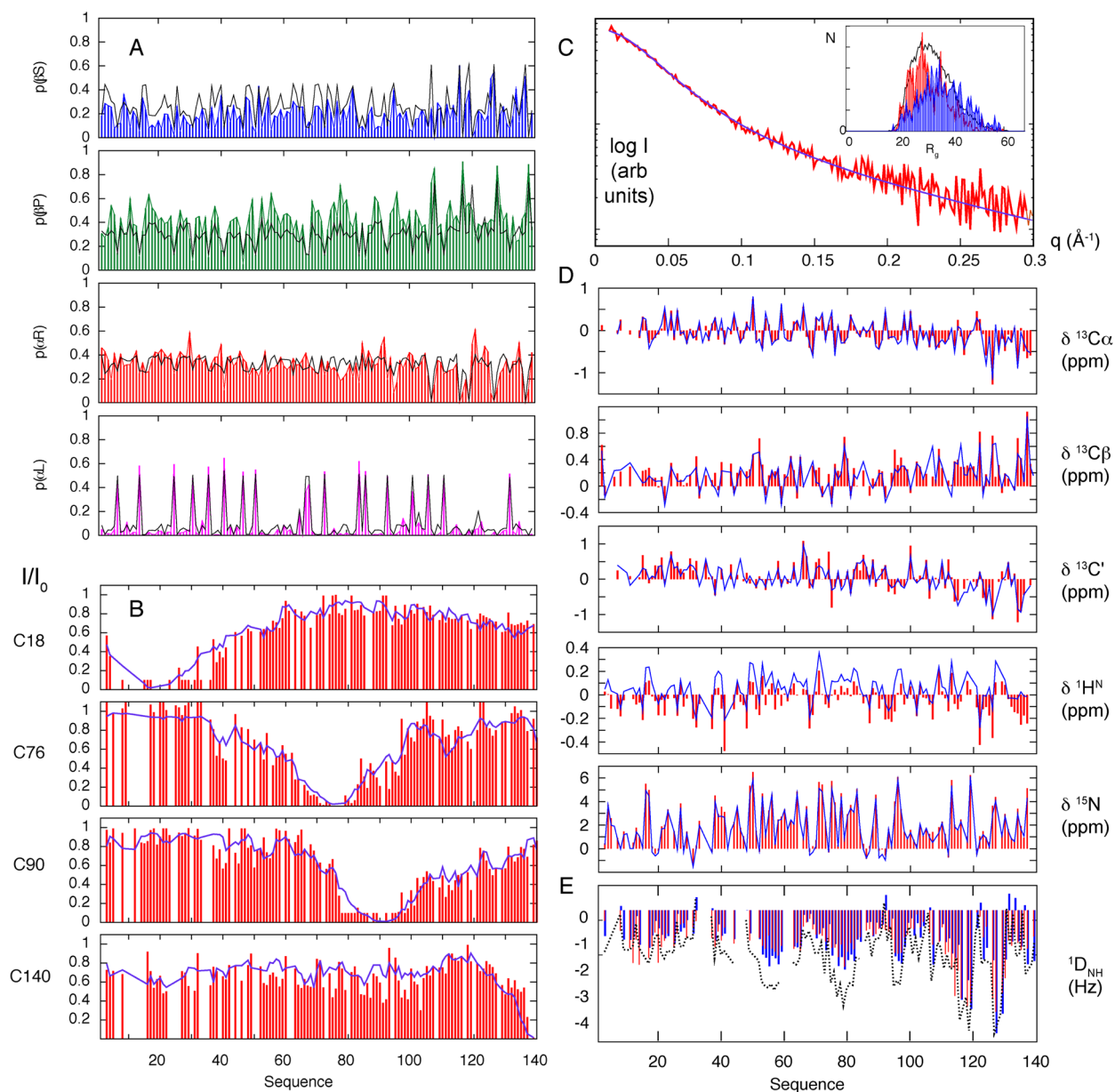
Figure 19. Structural ensemble description of α-synuclein from experimental CSs, RDCs, PREs, and SAXS data. (A) Populations in the four regions of Ramachandran space defined in Figure 2. The populations were derived from selected ASTEROIDS ensembles (blue, green, red and magenta) and compared to the populations in the statistical coil library (black). (B) Comparison of experimental PREs (red) with those back-calculated from selected ASTEROIDS ensemble (blue). (C) Comparison of experimental SAXS data (red) with the scattering curve calculated from the selected ASTEROIDS ensemble (blue). The inset shows the distribution of the radius of gyration, R_g, in the selected ASTEROIDS ensembles employing both PRE and SAXS data (blue) as compared to selected ensembles on the basis of PRE data only (red), and the statistical coil (black). (D) Reproduction of experimental CSs (red) by the ASTEROIDS selected ensembles (blue). (E) Cross-validation of experimental $^1D_{NH}$ RDCs. Comparison of experimental $^1D_{NH}$ RDCs (red) with RDCs back-calculated from selected ASTEROIDS ensemble using SAXS, PREs, and CSs only (blue) and RDCs calculated for a statistical coil ensemble (dotted line). Reprinted in part with permission from ref 298. Copyright 2014 Cell Press.

experimental selections provide evidence for systematic and significant deviation from the statistical coil model. The presence of type I β-turns (populated between 15% and 25% over the four repeat domains) is clearly identified. This conformational sampling is very similar to that predicted by AMD simulation in a previous study.[111] Continuous stretches of enhanced population of βP are also observed in the strands comprising residues (256−261), (275−282), (307−313), and (338−346), the central two of which mediate binding to microtubules and have been identified as aggregation nucleation sites important for the formation of Tau oligomers

and fibrils. Perhaps not surprisingly, the levels of βP sampling in the PRR were up to 20% higher than expected from the statistical coil.

The study of the protein Tau was extended further by combining NMR CSs (^{15}N, ^{13}C and $^1H^N$), RDCs ($^1D_{HN}$), and PREs from 12 cysteine mutants, and SAXS data, to develop an ensemble description of the full length Tau protein (441 amino acids in length).[298] Ensemble selection of such a large protein is challenging. The volume of experimental data available for such a large molecule comprises more than 2500 PRE-derived peak intensity ratios, more than 2000 chemical shifts, SAXS data, as
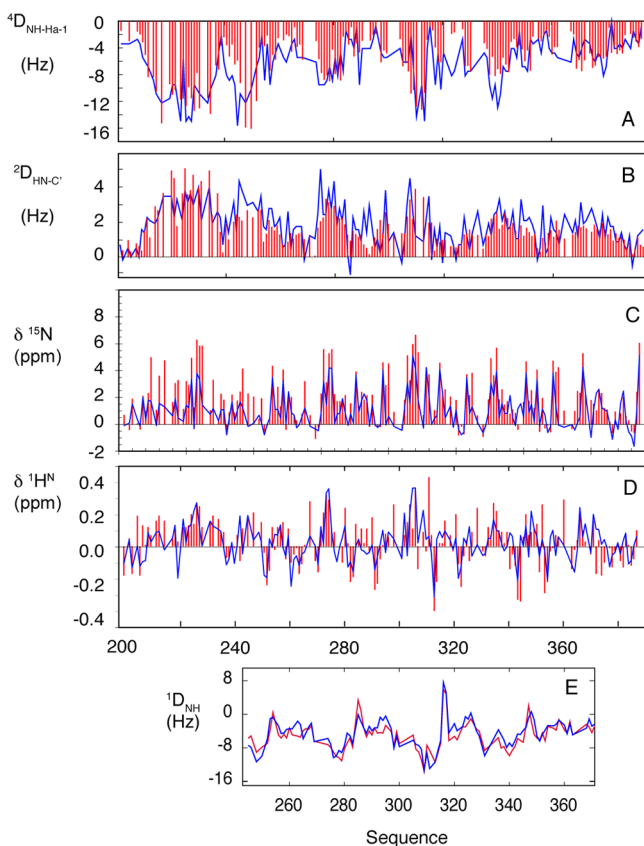
Figure 20. Cross-validation of the ASTEROIDS ensemble of the K32 construct of Tau. Reproduction of experimental data when not included as active data in the ASTEROIDS target selection function. (A) Experimental (red) and back-calculated (blue) $^4D_{NH-H\alpha-1}$ RDCs from selection using $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, $^1H^N$, and ^{15}N CSs and $^1D_{HN}$ RDCs. (B) Experimental (red) and back-calculated (blue) $^2D_{HN-C'}$ RDCs from selection using $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, $^1H^N$, and ^{15}N CSs and $^1D_{HN}$ RDCs. (C) Experimental (red) and back-calculated (blue) ^{15}N CSs from selection using $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$ CSs and $^1D_{NH}$ and $^2D_{HN-C'}$ RDCs. (D) Experimental (red) and back-calculated (blue) $^1H^N$ CSs from selection using $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$ CSs and $^1D_{NH}$ and $^2D_{HNC'}$ RDCs. (E) Experimental (red) and back-calculated (blue) $^1D_{NH}$ RDCs in K18 from selection using ^{13}C Appro$^\alpha$, $^{13}C^\beta$, $^{13}C'$, $^1H^N$, and ^{15}N CSs. Reprinted in part with permission from ref 298. Copyright 2014 Cell Press.

well as local and long-range RDC predictions, all of which are calculated for each of 36 000 conformers present in the prior sampling pool. The optimal number of structures was estimated to be approximately 400, requiring averaging of around 10^7 simulated observables for each of 100 000 iterations of the ASTEROIDS genetic algorithm.

The resulting data reproduction and local conformational sampling are shown in Figures 16 and 17. Within the K32 region (198−394), sampling is essentially identical to that observed independently from this domain, while helical regions are observed in the C-terminal region and around residue 118. Elevated populations of βP are again seen at specific regions in the protein even in the absence of proline residues. The ensemble also characterizes transient long-range order in Tau protein from both PREs and SAXS between the N-terminus and the proline-rich region[22,301] that has also been identified using EPR and FRET,[302,303] and that has been shown, using ASTEROIDS, to be abrogated in a pseudophosphorylated mutated form of the protein where additional charges mimic

phosphorylation of S199, S202, T205, T212, S214, S396, and S404 (Figure 18).[301] The long-range contact plot also reveals more extended sampling in the repeat domain of Tau as has been seen previously from SAXS.[304] This study also underlines the complementary nature of SAXS and PRE data, which are both required to correctly define the overall dimensions of the ensemble, notably because only the SAXS data are unambiguously sensitive to the more extended conformations.

ASTEROIDS was again applied to the protein α-synuclein (Figure 19), this time combining $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, ^{15}N, and $^1H^N$ CSs, $^1D_{HN}$ RDCs, SAXS, and four PRE data sets in a single ensemble selection.[298] According to the calibration study described above, this combination is sufficient to unambiguously distinguish populations of βP, βS, αR, and αL regions in the resulting ensemble. The population of the βP region is continuously elevated over residues 76−81, which forms the central region of the nonamyloid-β component of the amyloid plaque region, suggesting, in combination with the observations from Tau, that the βP region may have some importance as an intermediate for aggregation propensity.[305] We note that the resulting ensemble, which is in agreement with available NMR and SAXS data, exhibited no evidence for large molecular weight assemblies that have previously been evoked from ensemble analysis.[279]

3.8. Cross-Validation: Testing the Predictive Capacity of Ensemble Descriptions

Access to the large volumes of experimental data described above provides the opportunity for rigorous testing of the predictive nature of ASTEROIDS ensembles. In the case of K32 $^2D_{CHN}$ and $^4D_{HNiH\alpha i-1}$, RDCs and ^{15}N and $^1H^N$ CSs were independently removed from the target selection and their values calculated from the selected ensembles (Figure 20).[298] A similar procedure was applied to N−H^N RDCs from K18 (Figure 20), α-synuclein (Figure 19), Sendai virus N$_{TAIL}$ (Figure 13), and full-length Tau protein (Figure 17). In total, nine independent sets, five different types of RDC and CS ($^1D_{HN}$, $^2D_{CHN}$, and $^4D_{HNiH\alpha i-1}$ RDCs and $^1H^N$ and ^{15}N CSs from K18, K32, htau40, and αSyn) were predicted from ASTEROIDS selections on the basis of the remaining NMR data. The results demonstrate significant improvement in the reproduction of independent data sets as compared to prediction from statistical coil. In most cases, the improvement increases when only regions exhibiting significantly different conformational sampling are compared, providing quantifiable evidence that the improvement does not result from fitting to noise or poor conformational sampling. A recent study proposed a different approach, whereby data were removed from ensemble analysis and the effects on the physical properties of the resultant ensembles were monitored.[306] In general, we believe that the use of such robust comparison procedures will be essential for the development of credible ensemble descriptions of highly disordered proteins.[307]

4. CONCLUSIONS

In this Review, we have attempted to describe recent advances in the investigation of the free-energy surface explored by intrinsically disordered proteins using NMR in combination with other solution-state experimental techniques. In particular, we have presented different approaches to the difficult transformation of experimental data into valid conformational distributions. In general, restrained MD-based approaches combine experimental data with potential energy functions to

determine representative ensembles, while ensemble selection approaches sample an unsupervised conformational space and then attempt to delineate free-energy contours on the basis of experimental data. The advantages and disadvantages of the different approaches are not yet clear, so that both approaches can gain insight from the other; for example, one could hope that force fields appropriate for IDPs can be refined from information derived from ensemble descriptions. In all cases, it is clear that the highly under-determined conformational space fundamental to IDPs requires the development of robust validation procedures that can test the predictive nature of the descriptions and to provide estimates of uncertainty and statistical confidence in the ensemble descriptions.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +33 457 428 554. E-mail: martin.blackledge@ibs.fr.

Notes

The authors declare no competing financial interest.

Biographies

Malene Ringkjøbing Jensen studied chemistry and mathematics at the University of Copenhagen, and she received her Ph.D. degree in chemistry in 2006 with Prof. Jens J. Led from the Department of Chemistry, University of Copenhagen. From 2007 to 2009 she worked as a postdoctoral fellow in the group of Dr. Martin Blackledge at the Institut de Biologie Structurale in Grenoble focusing on the development of structural ensemble representations of intrinsically disordered proteins from experimental NMR data. Since 2010 she holds a position as an associate scientist at the Centre National de la Recherche Scientifique (CNRS) at the Institut de Biologie Structurale. Her research focuses on the role of protein intrinsic disorder in biological systems such as cell signaling pathways and viral transcription and replication.

Markus Zweckstetter studied physics at the Ludwig Maximilians University in Munich. He wrote his Ph.D. thesis under the supervision of Tad Holak at the Max Planck Institute for Biochemistry in Martinsried and received his Ph.D. from the Technical University of Munich. Subsequently, he worked from 1999 to 2001 as a postdoctoral fellow in Dr. Adrian Bax's group at the National Institutes of Health in Bethesda, MD. He has headed the research group "Structure Determination of Proteins Using NMR" at the Max-Planck-Institute for Biophysical Chemistry since 2001. Since June 2012 he is heading the senior research group "Structural Biology in Dementia" at the German Center for Neurodegenerative Diseases (DZNE) Göttingen. Since December 2012 he is also professor at the University Medical School, Göttingen. His laboratory tackles the structural biology challenges faced today in neurodegeneration that are intrinsically disordered proteins, protein aggregates, big proteins, nucleosome complexes, and proteins in membranes.

Jie-rong Huang received his bachelor degree in physics from National Taiwan University, Taiwan (1997–2001), and finished his Ph.D. thesis under the supervision of Dr. Sophie Jackson at the University of Cambridge, UK (2003–2007). Subsequently, he moved to the University of Basel, Switzerland for his postdoctoral training with Professor Stephan Grzesiek (2007–2010) and then worked at the Institute of Structural Biology of CNRS, France, with Dr. Martin Blackledge (2010–2013). He is currently an assistant professor at the Institute of Biochemistry and Molecular Biology at the National Yang-Ming University, Taiwan. His long-standing research interests focus on using NMR and computational modeling methods to understand the protein folding problem and the structural propensities of intrinsically disordered proteins.

Martin Blackledge studied physics at the University of Manchester, and received his doctoral degree with Professor Sir George Radda from the University of Oxford in 1987. He received a Royal Society postdoctoral fellowship to work in the group of Professor Richard Ernst in the Physical Chemistry department at the ETH Zürich, working on the study of conformational dynamics of biomolecules using NMR, in particular rotating frame relaxation. Since 1992 he has been working for the French Commissariat à l'energie atomique at the Institute de Biologie Structurale in Grenoble. He directs the Protein Dynamics and Flexibility by NMR group, which focuses on the use of NMR spectroscopy to study the role of protein dynamics in biomolecular function. His recent research efforts have concentrated on the study of weak protein−protein interactions, and the development of NMR-based approaches for the study of intrinsically disordered proteins; in particular, his group is interested in proteins involved in viral replication, neurodegenerative disease, and signaling.

ACKNOWLEDGMENTS

REFERENCES

(1) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739.
(2) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(3) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
(4) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* **2010**, *1804*, 1231.
(5) Frauenfelder, H.; Sligar, S.; Wolynes, P. *Science* **1991**, *254*, 1598.
(6) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539.
(7) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35.
(8) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756.
(9) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.
(10) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31.
(11) Van Roey, K.; Gibson, T. J.; Davey, N. E. *Curr. Opin. Struct. Biol.* **2012**, *22*, 378.
(12) Forman-Kay, J. D.; Mittag, T. *Structure* **2013**, *21*, 1492.
(13) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. *Molecules* **2013**, *18*, 10802.
(14) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607.
(15) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
(16) Eliezer, D. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23.
(17) Tompa, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419.
(18) Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426.
(19) Ito, Y.; Selenko, P. *Curr. Opin. Struct. Biol.* **2010**, *20*, 640.
(20) Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426.
(21) Rezaei-Ghaleh, N.; Blackledge, M.; Zweckstetter, M. *ChemBioChem* **2012**, *13*, 930.
(22) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *PLoS Biol.* **2009**, *7*, e34.
(23) Pervushin, K.; Vögeli, B.; Eletsky, A. *J. Am. Chem. Soc.* **2002**, *124*, 12898.
(24) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128*, 9042.
(25) Deschamps, M.; Campbell, I. D. *J. Magn. Reson.* **2006**, *178*, 206.
(26) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187*, 163.
(27) Narayanan, R. L.; Dürr, U. H. N.; Bibow, S.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. *J. Am. Chem. Soc.* **2010**, *132*, 11906.
(28) Wen, J.; Wu, J.; Zhou, P. *J. Magn. Reson.* **2011**, *209*, 94.
(29) Zawadzka-Kazimierczuk, A.; Koźmiński, W.; Sanderová, H.; Krásný, L. *J. Biomol. NMR* **2012**, *52*, 329.
(30) Nováček, J.; Haba, N. Y.; Chill, J. H.; Žídek, L.; Sklenář, V. *J. Biomol. NMR* **2012**, *53*, 139.
(31) Krimm, S.; Tiffany, M. *Isr. J. Chem.* **1974**, *12*, 189.
(32) Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2003**, *125*, 8178.
(33) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2013**, *117*, 6927.
(34) McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D. *J. Am. Chem. Soc.* **2004**, *126*, 5076.
(35) Syme, C. D.; Blanch, E. W.; Holt, C.; Jakes, R.; Goedert, M.; Hecht, L.; Barron, L. D. *Eur. J. Biochem.* **2002**, *269*, 148.
(36) Mästle, W.; Dukor, R. K.; Yoder, G.; Keiderling, T. A. *Biopolymers* **1995**, *36*, 623.
(37) Nettels, D.; Mueller-Spaeth, S.; Kuester, F.; Hofmann, H.; Haenni, D.; Rueegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20740.
(38) Ferreon, A. C. M.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A. *Nature* **2013**, *498*, 390.
(39) Schuler, B.; Hofmann, H. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36.
(40) Lipfert, J.; Doniach, S. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 307.
(41) Sibille, N.; Bernadó, P. *Biochem. Soc. Trans.* **2012**, *40*, 955.
(42) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363.
(43) Osapay, K.; Case, D. A. *J. Biomol. NMR* **1994**, *4*, 215.
(44) Oldfield, E. *J. Biomol. NMR* **1995**, *5*, 217.
(45) Mulder, F. A. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578.
(46) Wishart, D. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 62.
(47) Case, D. A. *Curr. Opin. Struct. Biol.* **2013**, *23*, 172.
(48) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.
(49) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.
(50) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171.
(51) Kjaergaard, M.; Poulsen, F. M. *J. Biomol. NMR* **2011**, *50*, 157.
(52) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67.
(53) Kjaergaard, M.; Brander, S.; Poulsen, F. M. *J. Biomol. NMR* **2011**, *49*, 139.
(54) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Wright, P. E.; Dyson, H. J. *J. Biomol. NMR* **2000**, *18*, 43.
(55) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970.
(56) Prestegard, J. H.; Sahu, S. C.; Nkari, W. K.; Morris, L. C.; Live, D.; Gruta, C. *J. Biomol. NMR* **2013**, *55*, 201.
(57) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402.
(58) Wang, Y.; Jardetzky, O. *J. Am. Chem. Soc.* **2002**, *124*, 14075.
(59) Wang, Y.; Jardetzky, O. *Protein Sci.* **2002**, *11*, 852.

(60) Zhang, H.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173.

(61) De Simone, A.; Cavalli, A.; Hsu, S.-T. D.; Vranken, W.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 16332.

(62) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002.

(63) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-R.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463.

(64) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322.

(65) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494.

(66) Swindells, M.; Macarthur, M.; Thornton, J. *Nat. Struct. Biol.* **1995**, *2*, 596.

(67) Griffiths-Jones, S. R.; Sharman, G. J.; Maynard, A. J.; Searle, M. S. *J. Mol. Biol.* **1998**, *284*, 1597.

(68) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691.

(69) Fitzkee, N. C.; Fleming, P. J.; Rose, G. D. *Proteins* **2005**, *58*, 852.

(70) Seki, Y.; Shimbo, Y.; Nonaka, T.; Soda, K. *J. Chem. Theory Comput.* **2011**, *7*, 2126.

(71) Jiang, F.; Han, W.; Wu, Y.-D. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3413.

(72) Kjaergaard, M.; Poulsen, F. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *60*, 42.

(73) Kragelj, J.; Ozenne, V.; Blackledge, M.; Jensen, M. R. *ChemPhysChem* **2013**, *14*, 3034.

(74) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795.

(75) Tamiola, K.; Acar, B.; Mulder, F. A. A. *J. Am. Chem. Soc.* **2010**, *132*, 18000.

(76) Tamiola, K.; Mulder, F. A. A. *Biochem. Soc. Trans.* **2012**, *40*, 1014.

(77) Maiti, N. C.; Apetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. *J. Am. Chem. Soc.* **2004**, *126*, 2399.

(78) Shi, Z.; Chen, K.; Liu, Z.; Kallenbach, N. R. *Chem. Rev.* **2006**, *106*, 1877.

(79) Woody, R. W. *J. Am. Chem. Soc.* **2009**, *131*, 8234.

(80) Cao, W.; Bracken, C.; Kallenbach, N. R.; Lu, M. *Protein Sci.* **2004**, *13*, 177.

(81) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. *Protein Sci.* **2010**, *19*, 1555.

(82) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2012**, *134*, 15138.

(83) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* **2012**, *51*, 2224.

(84) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215.

(85) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43.

(86) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289.

(87) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13.

(88) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894.

(89) Li, D.-W.; Brüschweiler, R. *J. Biomol. NMR* **2012**, *54*, 257.

(90) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 1220.

(91) Li, D.-W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246.

(92) Robustelli, P.; Stafford, K. A.; Palmer, A. G., III. *J. Am. Chem. Soc.* **2012**, *134*, 6365.

(93) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.

(94) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.

(95) Bystrov, V.; Gavrilov, Y.; Solkan, V. *J. Magn. Reson.* **1975**, *19*, 123.

(96) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741.

(97) Vuister, G. W.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772.

(98) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285.

(99) Case, D. A.; Scheurer, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390.

(100) Markwick, P. R. L.; Showalter, S. A.; Bouvignies, G.; Brüschweiler, R.; Blackledge, M. *J. Biomol. NMR* **2009**, *45*, 17.

(101) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322.

(102) Penkett, C. J.; Redfield, C.; Jones, J. A.; Dodd, I.; Hubbard, J.; Smith, R. a. G.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1998**, *37*, 17054.

(103) Danielsson, J.; Jarvet, J.; Damberg, P.; Graslund, A. *FEBS J.* **2005**, *272*, 3938.

(104) Tang, Y. F.; Goger, M. J.; Raleigh, D. P. *Biochemistry* **2006**, *45*, 6940.

(105) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799.

(106) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359.

(107) Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2010**, *49*, 9964.

(108) Oh, K.-I.; Jung, Y.-S.; Hwang, G.-S.; Cho, M. *J. Biomol. NMR* **2012**, *53*, 25.

(109) Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2013**, *425*, 3338.

(110) Xiang, S.; Gapsys, V.; Kim, H.-Y.; Bessonov, S.; Hsiao, H.-H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Lührmann, R.; de Groot, B.; Zweckstetter, M. *Structure* **2013**, *21*, 2162.

(111) Mukrasch, M. D.; Markwick, P.; Biernat, J.; Bergen, M.; von Bernadó, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235.

(112) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.

(113) Esteban-Martín, S.; Fenwick, R. B.; Salvatella, X. *J. Am. Chem. Soc.* **2010**, *132*, 4626.

(114) Seip, S.; Balbach, J.; Kessler, H. *J. Magn. Reson., Ser. B* **1994**, *104*, 172.

(115) Wang, A. C.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 1810.

(116) Peti, W.; Hennig, M.; Smith, L. J.; Schwalbe, H. *J. Am. Chem. Soc.* **2000**, *122*, 12017.

(117) Permi, P.; Kilpeläinen, I.; Annila, A. *J. Magn. Reson.* **2000**, *146*, 255.

(118) Löhr, F.; Schmidt, J. M.; Maurer, S.; Rüterjans, H. *J. Magn. Reson.* **2001**, *153*, 75.

(119) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2010**, *132*, 540.

(120) Demarco, A.; Llinás, M.; Wüthrich, K. *Biopolymers* **1978**, *17*, 617.

(121) Karimi-Nejad, Y.; Schmidt, J. M.; Rüterjans, H.; Schwalbe, H.; Greisinger, C. *Biochemistry* **1994**, *33*, 5481.

(122) West, N. J.; Smith, L. J. *J. Mol. Biol.* **1998**, *280*, 867.

(123) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705.

(124) Sziegat, F.; Silvers, R.; Hähnke, M.; Jensen, M. R.; Blackledge, M.; Wirmer-Bartoschek, J.; Schwalbe, H. *Biochemistry* **2012**, *51*, 3361.

(125) Vajpai, N.; Gentner, M.; Huang, J.-R.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 3196.

(126) Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. *Chem.—Eur. J.* **2011**, *17*, 6789.

(127) Rybka, K.; Toal, S. E.; Verbaro, D. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. *Proteins* **2013**, *81*, 968.

(128) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179.

(129) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257.

(130) Roche, J.; Ying, J.; Maltsev, A. S.; Bax, A. *ChemBioChem* **2013**, *14*, 1754.

(131) Uversky, V. N. *Biochim. Biophys. Acta* **2011**, *1814*, 693.

(132) Baker, J. M. R.; Hudson, R. P.; Kanelis, V.; Choy, W.-Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738.

(133) Macura, S.; Ernst, R. *Mol. Phys.* **1980**, *41*, 95.

(134) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355.

(135) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170.

(136) Clore, G. M.; Tang, C.; Iwahara, J. *Curr. Opin. Struct. Biol.* **2007**, *17*, 603.

(137) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945.

(138) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383.

(139) Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108.

(140) Iwahara, J.; Clore, G. M. *Nature* **2006**, *440*, 1227.

(141) Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J. Mol. Biol.* **2005**, *347*, 1053.

(142) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C.; Poulsen, F.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291.

(143) Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430.

(144) Columbus, L.; Hubbell, W. L. *Trends Biochem. Sci.* **2002**, *27*, 288.

(145) Sezer, D.; Freed, J. H.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 5755.

(146) Keizers, P. H. J.; Saragliadis, A.; Hiruma, Y.; Overhand, M.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 14802.

(147) Fawzi, N. L.; Fleissner, M. R.; Anthis, N. J.; Kálai, T.; Hideg, K.; Hubbell, W. L.; Clore, G. M. *J. Biomol. NMR* **2011**, *51*, 105.

(148) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.

(149) Bruschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. *J. Am. Chem. Soc.* **1992**, *114*, 2289.

(150) Dedmon, M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. *J. Am. Chem. Soc.* **2005**, *127*, 476.

(151) Song, J.; Guo, L.-W.; Muradov, H.; Artemyev, N. O.; Ruoho, A. E.; Markley, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1505.

(152) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314.

(153) Ganguly, D.; Chen, J. *J. Mol. Biol.* **2009**, *390*, 467.

(154) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784.

(155) Huang, J.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694.

(156) Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6278.

(157) Xue, Y.; Podkorytov, I. S.; Rao, D. K.; Benjamin, N.; Sun, H.; Skrynnikov, N. R. *Protein Sci.* **2009**, *18*, 1401.

(158) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W.-Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2007**, *367*, 1494.

(159) Cho, M.-K.; Nodet, G.; Kim, H.-Y.; Jensen, M. R.; Bernado, P.; Fernandez, C. O.; Becker, S.; Blackledge, M.; Zweckstetter, M. *Protein Sci.* **2009**, *18*, 1840.

(160) McConnell, H. M. *J. Chem. Phys.* **1958**, *28*, 430.

(161) Iwahara, J.; Clore, G. M. *J. Am. Chem. Soc.* **2010**, *132*, 13346.

(162) Xue, Y.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2011**, *133*, 14614.

(163) Palmer, A. *Chem. Rev.* **2004**, *104*, 3623.

(164) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.

(165) Alexandrescu, A.; Shortle, D. *J. Mol. Biol.* **1994**, *242*, 527.

(166) Farrow, N.; Zhang, O.; Forman-Kay, J.; Kay, L. *Biochemistry* **1995**, *34*, 868.

(167) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121*, 8671.

(168) Yang, D. W.; Mok, Y. K.; Muhandiram, D. R.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 3555.

(169) Tollinger, M.; Skrynnikov, N. R.; Mulder, F. a. A.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 11341.

(170) Ochsenbein, F.; Neumann, J. M.; Guittet, E.; Van Heijenoort, C. *Protein Sci.* **2002**, *11*, 957.

(171) Choy, W. Y.; Shortle, D.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 1748.

(172) Wirmer, J.; Peti, W.; Schwalbe, H. *J. Biomol. NMR* **2006**, *35*, 175.

(173) Houben, K.; Blanchard, L.; Blackledge, M.; Marion, D. *Biophys. J.* **2007**, *93*, 2830.

(174) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787.

(175) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.

(176) Robustelli, P.; Trbovic, N.; Friesner, R. A.; Palmer, A. G. *J. Chem. Theory Comput.* **2013**, *9*, 5190.

(177) Bae, S.-H.; Dyson, H. J.; Wright, P. E. *J. Am. Chem. Soc.* **2009**, *131*, 6814.

(178) Amoros, D.; Ortega, A.; Garcia de la Torre, J. *J. Chem. Theory Comput.* **2013**, *9*, 1678.

(179) Rezaei-Ghaleh, N.; Klama, F.; Munari, F.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 11410.

(180) Carver, J.; Richards, R. *J. Magn. Reson.* **1972**, *6*, 89.

(181) Hansen, D. F.; Vallurupalli, P.; Lundström, P.; Neudecker, P.; Kay, L. E. *J. Am. Chem. Soc.* **2008**, *130*, 2667.

(182) Palmer, A. G.; Massi, F. *Chem. Rev.* **2006**, *106*, 1700.

(183) Korzhnev, D.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A.; Davidson, A.; Dobson, C.; Kay, L. *Nature* **2004**, *430*, 586.

(184) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021.

(185) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312.

(186) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868.

(187) Motlagh, H. N.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4134.

(188) Liu, J.; Faeder, J. R.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19819.

(189) Long, D.; Brueschweiler, R. *PLoS Comput. Biol.* **2011**, *7*, e1002035.

(190) Knott, M.; Best, R. B. *PLoS Comput. Biol* **2012**, *8*, e1002605.

(191) Zhang, W.; Ganguly, D.; Chen, J. *PLoS Comput. Biol.* **2012**, *8*, e1002353.

(192) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111.

(193) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23.

(194) Tolman, J.; Ruan, K. *Chem. Rev.* **2006**, *106*, 1720.

(195) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098.

(196) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.

(197) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968.

(198) Guerry, P.; Salmon, L.; Mollica, L.; Ortega Roldan, J.-L.; Markwick, P.; van Nuland, N. A. J.; McCammon, J. A.; Blackledge, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 3181.

(199) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487.

(200) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131.

(201) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191.

(202) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099.

(203) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.

(204) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M. *Structure* **2009**, *17*, 1169.

(205) Torbet, J.; Maret, G. *J. Mol. Biol.* **1979**, *134*, 843.

(206) Hansen, M.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065.

(207) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571.

(208) Ruckert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793.

(209) Sass, H.; Musco, G.; Stahl, S.; Wingfield, P.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303.

(210) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340.

(211) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679.

(212) Skora, L.; Cho, M.-K.; Kim, H.-Y.; Becker, S.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 7012.

(213) Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci.* **2003**, *12*, 2132.

(214) Meier, S.; Guthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051.

(215) Louhivuori, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647.

(216) Cho, M.-K.; Kim, H.-Y.; Bernado, P.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *J. Am. Chem. Soc.* **2007**, *129*, 3032.

(217) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563.

(218) Fredriksson, K.; Louhivuori, M.; Permi, P.; Annila, A. *J. Am. Chem. Soc.* **2004**, *126*, 12646.

(219) Obolensky, O. I.; Schlepckow, K.; Schwalbe, H.; Solov'yov, A. V. *J. Biomol. NMR* **2007**, *39*, 1.

(220) Cubrovic, M.; Obolensky, O. I.; Solov'yov, A. V. *Eur. Phys. J. D* **2009**, *51*, 41.

(221) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791.

(222) Bernadó, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968.

(223) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908.

(224) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7804.

(225) Huang, J.-R.; Ozenne, V.; Jensen, M. R.; Blackledge, M. *Angew. Chem., Int. Ed.* **2012**.

(226) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W. H.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055.

(227) Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 11266.

(228) Jensen, M. R.; Communie, G.; Ribeiro, E. A., Jr.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 9839.

(229) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762.

(230) Tompa, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 419.

(231) Bonvin, A. M.; Rullmann, J. A.; Lamerichs, R. M.; Boelens, R.; Kaptein, R. *Proteins* **1993**, *15*, 385.

(232) Gsponer, J.; Hopearuoho, H.; Whittaker, S. B.-M.; Spence, G. R.; Moore, G. R.; Paci, E.; Radford, S. E.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 99.

(233) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784.

(234) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 5879.

(235) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(236) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.

(237) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2007**, *368*, 1448.

(238) Terakawa, T.; Takada, S. *Biophys. J.* **2011**, *101*, 1450.

(239) Knott, M.; Best, R. B. *PLoS Comput. Biol.* **2012**, *8*, e1002605.

(240) Narayanan, C.; Weinstock, D. S.; Wu, K.-P.; Baum, J.; Levy, R. M. *J. Chem. Theory Comput.* **2012**, *8*, 3929.

(241) Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3743.

(242) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. *J. Phys. Chem. B* **2013**, *117*, 118.

(243) Roux, B.; Weare, J. *J. Chem. Phys.* **2013**, *138*, 084107.

(244) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(245) Im, W.; Jo, S.; Kim, T. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 252.

(246) Jaynes, E. *Phys. Rev.* **1957**, *106*, 620.

(247) Caflisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746.

(248) Daggett, V.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5142.

(249) Duan, Y.; Kollman, P. *Science* **1998**, *282*, 740.

(250) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925.

(251) Boczko, E.; Brooks, C. *Science* **1995**, *269*, 393.

(252) Das, A.; Sin, B. K.; Mohazab, A. R.; Plotkin, S. S. *J. Chem. Phys.* **2013**, *139*, 121925.

(253) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* **2008**, *130*, 6145.

(254) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comput. Biol.* **2008**, *4*, e1000060.

(255) Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 5214.

(256) Higo, J.; Nishimura, Y.; Nakamura, H. *J. Am. Chem. Soc.* **2011**, *133*, 10448.

(257) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07.

(258) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673.

(259) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004.

(260) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148.

(261) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553.

(262) Palazzesi, F.; Barducci, A.; Tollinger, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14237.

(263) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. *PLoS Comput. Biol.* **2009**, *5*, e1000452.

(264) Camilloni, C.; Schaal, D.; Schweimer, K.; Schwarzinger, S.; De Simone, A. *Biophys. J.* **2012**, *102*, 158.

(265) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6817.

(266) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(267) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.

(268) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 2997.

(269) Salmon, L.; Pierce, L.; Grimm, A.; Roldan, J.-L. O.; Mollica, L.; Jensen, M. R.; van Nuland, N.; Markwick, P. R. L.; McCammon, J. A.; Blackledge, M. *Angew. Chem., Int. Ed.* **2012**, *51*, 6103.

(270) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011.

(271) Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D. *Bioinformatics* **2013**, *29*, 398.

(272) Feldman, H. J.; Hogue, C. W. V. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 112.

(273) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. *Structure* **2010**, *18*, 494.

(274) Marsh, J. A.; Dancheck, B.; Ragusa, M. J.; Allaire, M.; Forman-Kay, J. D.; Peti, W. *Structure* **2010**, *18*, 1094.

(275) Pinheiro, A. S.; Marsh, J. A.; Forman-Kay, J. D.; Peti, W. *J. Am. Chem. Soc.* **2011**, *133*, 73.

(276) Fisher, C. K.; Huang, A.; Stultz, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 14919.

(277) Huang, A.; Stultz, C. M. *PLoS Comput. Biol.* **2008**, *4*, e1000155.

(278) Ullman, O.; Fisher, C. K.; Stultz, C. M. *J. Am. Chem. Soc.* **2011**, *133*, 19536.

(279) Gurry, T.; Ullman, O.; Fisher, C. K.; Perovic, I.; Pochapsky, T.; Stultz, C. M. *J. Am. Chem. Soc.* **2013**, *135*, 3865.

(280) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656.

(281) Bertini, I.; Giachetti, A.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Pierattelli, R.; Ravera, E.; Svergun, D. I. *J. Am. Chem. Soc.* **2010**, *132*, 13553.

(282) Bernado, P.; Modig, K.; Grela, P.; Svergun, D.; Tchorzewski, M.; Pons, M.; Akke, M. *Biophys. J.* **2010**, *98*, 2374.

(283) Yang, S.; Blachowicz, L.; Makowski, L.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15757.

(284) Francis, D. M.; Rozycki, B.; Koveal, D.; Hummer, G.; Page, R.; Peti, W. *Nat. Chem. Biol.* **2011**, *7*, 916.

(285) Salmon, L.; Bascom, G.; Andricioaei, I.; Al-Hashimi, H. M. *J. Am. Chem. Soc.* **2013**, *135*, 5457.

(286) Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.-R.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2012**, *134*, 15138.

(287) Huang, J.; Gabel, F.; Jensen, M. R.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2012**, *134*, 4429.

(288) Gabel, F.; Jensen, M. R.; Zaccaï, G.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 8769.

(289) Iešmantavičius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. *J. Am. Chem. Soc.* **2013**, *135*, 10155.

(290) Ozenne, V.; Noel, J.; Heidarsson, P. O.; Brander, S.; Poulsen, F. M.; Jensen, M. R.; Kragelund, B. B.; Blackledge, M.; Danielsson, J. *J. Mol. Biol.* **2013**, *426*, 722.

(291) Serber, Z.; Dötsch, V. *Biochemistry* **2001**, *40*, 14317.

(292) Selenko, P.; Frueh, D. P.; Elsaesser, S. J.; Haas, W.; Gygi, S. P.; Wagner, G. *Nat. Struct. Mol. Biol.* **2008**, *15*, 321.

(293) Binolfi, A.; Theillet, F.-X.; Selenko, P. *Biochem. Soc. Trans.* **2012**, *40*, 950.

(294) Theillet, F.-X.; Rose, H. M.; Liokatis, S.; Binolfi, A.; Thongwichian, R.; Stuiver, M.; Selenko, P. *Nat. Protoc.* **2013**, *8*, 1416.

(295) Waudby, C. A.; Camilloni, C.; Fitzpatrick, A. W. P.; Cabrita, L. D.; Dobson, C. M.; Vendruscolo, M.; Christodoulou, J. *PLoS One* **2013**, *8*, e72286.

(296) Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270.

(297) Mukrasch, M. D.; Biernat, J.; von Bergen, M.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *J. Biol. Chem.* **2005**, *280*, 24978.

(298) Schwalbe, M.; Ozenne, V.; Bibow, S.; Jaremko, M.; Jaremko, L.; Gajda, M.; Jensen, M. R.; Biernat, J.; Becker, S.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *Structure* **2014**, *22*, 238.

(299) Mukrasch, M. D.; von Bergen, M.; Biernat, J.; Fischer, D.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *J. Biol. Chem.* **2007**, *282*, 12230.

(300) Preuss, U.; Biernat, J.; Mandelkow, E. M.; Mandelkow, E. *J. Cell Sci.* **1997**, *110*, 789.

(301) Bibow, S.; Ozenne, V.; Biernat, J.; Blackledge, M.; Mandelkow, E.; Zweckstetter, M. *J. Am. Chem. Soc.* **2011**, *133*, 15842.

(302) Jeganathan, S.; von Bergen, M.; Brutlach, H.; Steinhoff, H.-J.; Mandelkow, E. *Biochemistry* **2006**, *45*, 2283.

(303) Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. *Biophys. J.* **2012**, *103*, 1940.

(304) Mylonas, E.; Hascher, A.; Bernadó, P.; Blackledge, M.; Mandelkow, E.; Svergun, D. I. *Biochemistry* **2008**, *47*, 10345.

(305) Kim, H.-Y.; Cho, M.-K.; Kumar, A.; Maier, E.; Siebenhaar, C.; Becker, S.; Fernandez, C. O.; Lashuel, H. A.; Benz, R.; Lange, A.; Zweckstetter, M. *J. Am. Chem. Soc.* **2009**, *131*, 17482.

(306) Marsh, J. A.; Forman-Kay, J. D. *Proteins* **2011**, *80*, 556.

(307) Jensen, M. R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, in press.