PAPER

Check for updates

Conformational sampling of the intrinsically disordered dsRBD-1 domain from *Arabidopsis thaliana* DCL1†

Irina P. Suárez,†[a] Diego F. Gauto,‡[a] Guillermo Hails,[a] Florencia C. Mascali,[a] Roberta Crespo,[a] Lingzi Zhao, [b] Jin Wang[c] and Rodolfo M. Rasia ★[ad]

DCL1 is the ribonuclease that carries out miRNA biogenesis in plants. Substrate pri-miRNA recognition by DCL1 requires two double stranded RNA binding domains located at the C-terminus of the protein. We have previously shown that the first of these domains, DCL1-A, is intrinsically disordered and folds upon binding pri-miRNA. Integrating NMR and SAXS data, we study here the conformational landscape of free DCL1-A through an ensemble description. Our results reveal that secondary structure elements, corresponding to the folded form of the protein, are transiently populated in the unbound state. The conformation of one of the dsRNA binding regions in the free protein shows that, at a local level, RNA recognition proceeds through a conformational selection mechanism. We further explored the stability of the preformed structural elements *via* temperature and urea destabilization. The C-terminal helix is halfway on the folding pathway in free DCL1-A, constituting a potential nucleation site for the final folding of the protein. In contrast, the N-terminal helix adopts stable non-native structures that could hinder the correct folding of the protein in the absence of RNA. This description of the unfolded form allows us to understand details of the mechanism of binding-induced folding of the protein.

Introduction

The classical paradigm stating that protein function depends on a fixed three-dimensional conformation has been subverted since the beginning of this century, as the functions of disordered regions in proteins or even fully disordered proteins were brought to light. Since then, the study of a range of intrinsically disordered proteins (IDPs) has revealed a wide landscape of previously underestimated functionality.[1]

Proteins that are fully disordered or show local disordered regions in their structures appear in a variety of contexts. Disordered fragments usually act as linkers between protein domains, conferring motional flexibility. Relatively short regions (10–15 residues) can act as molecular recognition elements, where the function is dictated mostly by the sequence itself and is not dependent on the conformational sampling.

Long disordered regions are sometimes also present at either or both the C- and N-termini of proteins. Finally, some proteins are fully disordered, their function depending on their high degree of flexibility.[2]

In terms of flexibility, the degree of disorder can be variable as well. Evidence from biophysical characterization indicate that many IDPs are well-described by simple random or statistical coil models. In other cases, proteins sample a more restricted conformational space, due to the presence of local (short range) and/or global (long range) partial order that imposes spatial restrictions. These conformational constraints were usually found to be essential for their function. A global evaluation of several IDPs show that the level of conformational restriction can be understood as a continuum ranging from stably folded proteins, going through forms that retain long range inter-actions described as molten globules, proteins only with local folding propensities and finally proteins with mostly random conformations.[3]

Experimental information on conformational sampling by IDPs is usually obtained *via* spectroscopic methods. Among the available techniques, solution NMR is the most revealing, as it is the only experimental method that provides structural information on IDPs with atomic resolution.[4] Despite the ever-increasing number of IDPs identified and described, precise information on conformational sampling obtained by spectroscopic methods is

[a] *Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET-UNR), Rosario, Santa Fe, Argentina. E-mail: rasia@ibr-conicet.gov.ar*

[b] *College of Physics, Jilin University, Changchun, Jilin, China*

[c] *State University of New York at Stony Brook, USA*

[d] *Área Biofísica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Santa Fe, Argentina*

This journal is © the Owner Societies 2018

Phys. Chem. Chem. Phys., 2018, **20**, 11237–11246 | **11237**

relatively scarce. In this context, the measurement and interpretation of data on new IDPs is undeniably relevant.[5]

DCL1 is the ribonuclease that carries out miRNA biogenesis in plants. The protein harbours two double stranded RNA binding domains (dsRBDs) located in tandem at the C-terminus that were shown to be essential for the function of the protein in pri-miRNA processing. The importance of these domains was demonstrated through the analysis of truncated forms of DCL1. Truncation of the sequence at the level of the second domain (dcl1-9 mutant) strongly diminishes miRNA processing, while truncation at the first domain (dcl1-6 mutant) is embryonic lethal, presumably because it abolishes DCL1 activity.[6] It is therefore evident that the first dsRBD (DCL1-A hereafter) is crucial for the activity of DCL1 in the context of the miRNA-processing complex. More recently, a construct containing the two tandem dsRBDs of DCL1 was shown to complement the pri-miRNA processing defective hyl1-2 mutant plants, highlighting the importance of this region in substrate recognition and in the formation of the miRNA processing complex in plants.[7]

We have previously shown that DCL1-A is intrinsically disordered but it binds its substrate pri-miRNA acquiring a canonical dsRBD fold.[8] We have described in depth the mechanism of binding-induced folding of this domain. Examination of the spectroscopic data shows that the free disordered form shows tendency to adopt folded conformations, and goes through an unfolded bound state prior to the folding event. Based on these observations we proposed a conformational selection mechanism for the encounter complex formation and an induced fit process for the folding event on the surface of the substrate. Here, we present a detailed structural analysis of the free form of DCL1-A. Using chemical shifts, RDCs, and SAXS data we generated a structural ensemble that represents the conformational space sampled by the protein. The ensemble reveals preformed structural elements with restrained flexibility that could be essential to drive binding and subsequent folding of the protein on its target RNA. By means of urea destabilization and temperature stabilization of the partially folded structural elements we could further analyse their relative stability and obtain an independent estimate for their population. This allowed us to show that the protein explores locally native-like conformations and to characterize its energy landscape. The stability patterns of different residual structure elements suggest possible causes for the disordered nature of the free protein.

Experimental

Protein expression and purification

DCL1-A was purified as described in Suarez *et al.*, 2015.[8] Briefly, the expression plasmid for DCL1-A was transformed in *E. coli* BL21(DE3) cells, which were then grown at 37 °C in M9 minimal medium supplemented with either 1 g L^{-1} ^{15}NH$_4$Cl or 1 g L^{-1} ^{15}NH$_4$Cl and 2 g L^{-1} [U–^{13}C]glucose (Cambridge Isotope Laboratories). Protein expression was induced at OD600 $\approx$ 0.7 and cells were grown overnight at 25 °C. The cells were harvested by centrifugation. Cell pellets were resuspended in 50 mM

Tris (pH 8.0), 500 mM NaCl, 5 mM imidazole, and 1 mM beta-mercaptoethanol and lysed by sonication. The clarified supernatant was purified using a Ni(ɪɪ)-affinity column and the protein was eluted with the same buffer supplemented with 350 mM imidazole. Fractions containing the protein were concentrated and digested with His-tagged TEV protease. The protease was removed with a Ni(ɪɪ)-affinity column, and the protein was further purified by a G75 size exclusion chromatography column equilibrated with 100 mM Phosphate, 50 mM NaCl, 10 mM beta-mercaptoethanol, pH 7.0. Protein concentration was measured by UV absorption at 280 nm with $\varepsilon_{280} = 1500$ M^{-1} cm^{-1}.

NMR spectroscopy

NMR spectra were acquired on a Bruker Avance II 600 MHz spectrometer or on a Bruker Avance III 700 MHz spectrometer, both equipped with TXI triple resonance probes. We employed pulse sequences that use the BEST selective excitation scheme[9] as implemented the standard Bruker library.

Urea titration data was obtained by stepwise addition of solid urea to a sample of ^{13}C^{15}N labelled DCL1-A. Urea concentrations were estimated from the mass of urea added, correcting the ensuing increment in volume. The concentrations reported were estimated by integration of the urea signal on the sample compared with that of individually externally prepared urea solutions. At each titration step ^{1}H–^{15}N-HSQC and HNCO spectra were acquired at 298 K. A full set of backbone assignment spectra was acquired at the highest urea concentration to confirm signal assignments.

RDCs were measured at 293 K in a C8E5-octanol alignment medium.[10] This medium was shown to induce a steric alignment being mostly inert and not interfering with the protein conformation. HN–N RDCs were obtained on a ^{15}N labelled sample using the standard IPAP sequence, as implemented in the Bruker library.

NOESY-^{15}N-HSQC spectra were acquired at 278 K and 298 K with 200 ms mixing time. TOCSY-^{15}N-HSQC spectra were acquired at 278 K and 298 K with 60 ms mixing time. Pulse sequences from the standard Bruker library were used in both cases.

NMR spectra acquired with full sampling schemes were transformed with NMRPipe,[11] those acquired with non-uniform sampling schemes were transformed with qMDD.[12] All NMR data were analysed using CcpNmr Analysis.[13] Secondary shifts were obtained using the software SSP.[14]

Circular dichroism (CD) spectroscopy

CD spectra were acquired on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controlling unit, using a 0.1 cm path cuvette to minimize the buffer contribution to absorption. Spectra were acquired from 190 to 250 nm, averaging four scans to improve the signal to noise ratio. Buffer condition was 10 mM phosphate (pH 6.5) and protein concentration was 10 μM. For the urea titration, an individual sample was prepared for each urea concentration. The complete sets of spectra as a function of temperature or urea were analysed by singular value decomposition (SVD) in order to obtain the spectra of the components that develop in the sample.

11238 | *Phys. Chem. Chem. Phys.*, 2018, **20**, 11237–11246

This journal is © the Owner Societies 2018

Small-angle X-ray scattering

Samples for SAXS were freshly collected after size exclusion chromatography and diluted to concentrations from 0.75 to 3 mg mL^{-1} in the reference buffer, which was the same as used for the chromatography column (50 mM phosphate buffer at pH 6, 5, 400 mM NaCl, 2 mM beta-mercaptoethanol). SAXS measurements were performed at 293 K on a Rigaku BioSAXS-1000 home source system attached to a Rigaku MicroMaxTM-007HF rotating anode generator with a Pilatus 100 K detector (Dectris). Data were collected in capillary sample holders, at a fixed sample-to-detector distance using a silver behenate calibration standard. Data processing was performed according to standard procedures using the package ATSAS.[15] Briefly, buffer scattering was subtracted form sample signal with PRIMUS[16] and the R_g was estimated by Guinier's approximation and the pair distance distribution analysis with GNOM.[17]

Ensemble calculations

The representative structural ensemble for DCL1-A was calculated following the protocol outlined by Jensen *et al.*[18] Briefly, we first generated 10 000 random possible protein conformations using Flexible-Meccano (FM).[19] We then calculated the theoretical chemical shifts for the backbone atoms using SPARTA+.[20] We then proceeded with five cycles of residue specific dihedral angle selection and new ensemble generation using algorithms written in Python and PyRosetta.[21] With the new residue-specific dihedral angle database, we generated a new 10 000 strong ensemble.

The final 10 000 strong ensemble reproduces accurately the experimental chemical shift data. We then calculated RDC profiles for each conformer using PALES,[22] and used those profiles to select a 200-strong subensemble of conformers that fits the experimental HN–N RDC data as well as the chemical shift data. Local structural conformation analysis was performed on this final subensemble.

Urea denaturation

Urea denaturation curves were fitted to a simple two-state unfolding mechanism. The unfolding equilibrium constant K at each denaturant concentration is usually calculated using eqn (1).[23]

$$K = \frac{U}{N} = \frac{y - y_N}{y_U - y} = e^{\left(-\frac{\Delta G_U}{RT}\right)} \tag{1}$$

where y is an observable that is assumed to vary linearly with the degree of folding of the protein between y_N for the native form to y_U for the unfolded form. The free energy change of unfolding, ΔG_U, is known to vary linearly with denaturant concentration, following eqn (2).

$$\Delta G_U = \Delta G_{U,H_2O} - m \cdot [\text{denaturant}] \tag{2}$$

where $\Delta G_{U,H_2O}$ is the free energy change of unfolding in the absence of denaturant and the slope m is a constant. This slope, termed *m*-value, was shown to depend mostly on the solvent accessible surface area (SASA) exposed upon unfolding.

We used the ^{13}CO chemical shift as a proxy for the conformation of the protein y, defining y_N as the ^{13}CO shift of the folded bound form of DCL1-A and y_U as the ^{13}CO shift of the protein in 9.8 M urea.

Results

Modelling of the conformational landscape of DCL1

Backbone chemical shift based secondary structure propensities. We had previously assigned the backbone HN, N, CA, CB and CO atoms of free DCL1-A at 298 K.[8] The secondary chemical shifts of DCL1-A clearly indicate the propensity of the C-terminal region to populate helical conformations, but no other feature is evident at first glance from the data. In order to perform a global analysis of the secondary structure propensities considering all backbone chemical shifts simultaneously we resorted to the d2D software[24] (Fig. S1, ESI†). The results show clearly the presence of a partially populated helix between residues Q50 and E73 (residue numbering scheme corresponds to our construct). In addition, the d2D analysis reveals some strand propensity in the regions corresponding to beta sheets 2 and 3 of the folded bound form of the protein (Fig. S1, ESI†). Residues V7, L10 and C14 in the region corresponding to helix 1 and K24 and R27 in the region corresponding to strand 1 show secondary shifts corresponding to helical and strand propensities, respectively. However as they do not form continuous stretches, the corresponding regions are unlikely to adopt any preferential secondary structure conformation.

NOESY-^{15}N-HSQC spectra of the protein show few NOE crosspeaks, corresponding to intra and i-1 correlations, further supporting the absence of a stable structure. Only few weak HN–HN crosspeaks are observed in the region corresponding to helix 2 at 278 K (Fig. S2, ESI†).

Quite remarkably, the intensities of signals in the HNCO spectrum seem to correlate with expected beta strand regions. The signal intensity in IDPs depend on both the conformational exchange regime and on the exchange of amide protons with water. Measurement of the amide proton chemical shift temperature dependence (see below) shows no evidence of hydrogen bonding in the sheet regions, thus excluding the formation of stable beta sheets in this region in the free protein. Therefore, the pattern may arise from a lower conformational exchange rate in residues corresponding to turns, and is indicative of the presence of residual structure in those regions.

Disordered free DCL1-A ensemble calculations. In order to gain information on the unfolded form of DCL1-A we used the backbone chemical shifts to obtain a structural ensemble representing the conformational space explored by the protein following the protocol outlined by Jensen *et al.*[18] with modifications, as explained in the methods section. The full 10 000 strong conformer ensemble calculated using chemical shifts shows an excellent agreement with the experimental values, and no significant improvement can be obtained by selection of subensembles. Chemical shifts provide excellent probes for local conformation, but give no information on global structure. To refine our ensemble with global structural information we
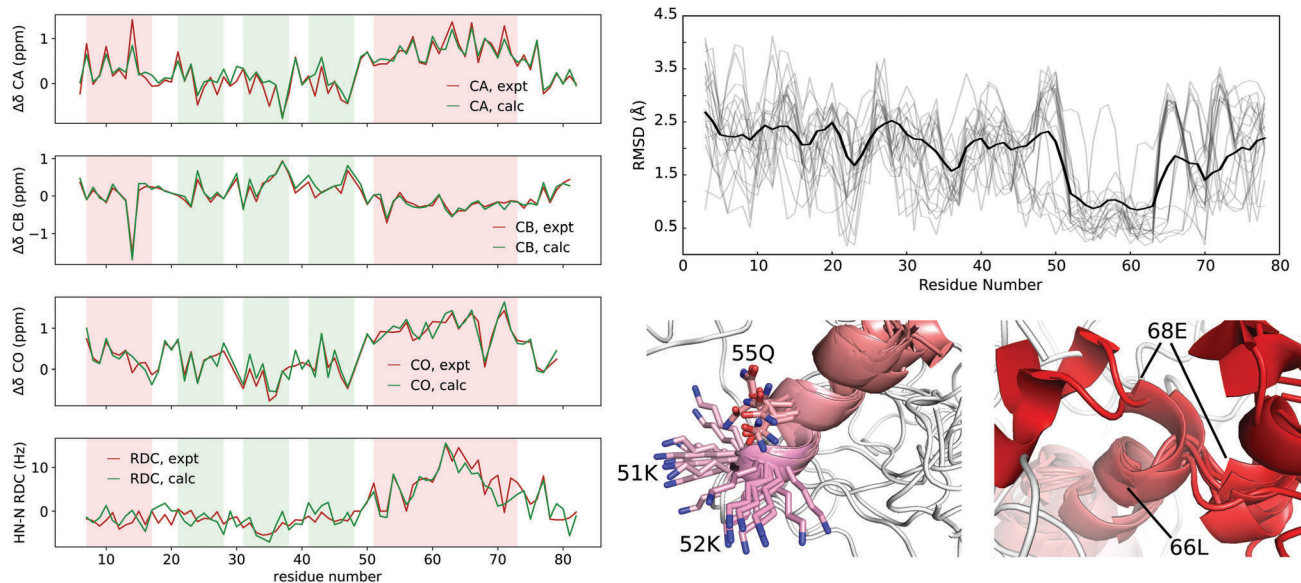
Fig. 1 Ensemble calculations on DCL1-A. Left, experimental (red) and calculated (green) secondary chemical shifts and HN–N RDCs for the selected structural ensemble. Regions corresponding to secondary structure elements on the folded form are shaded in red for helices and in green for sheets. Top right, local RMSD between the structures of the selected structural ensemble on a 7-residue sliding window. Bottom, details on the local structure. Left, RNA binding region at the N-terminus of helix 2. Right, split trajectories of helix 2 at residue 67.

acquired HN–N RDCs. The RDC pattern is very poorly reproduced by the final ensemble. We therefore used the HN–N RDC data to select a 200-strong subensemble from the full ensemble. The RDC baseline bears information on long range interactions[18] and the fine structure is particularly sensitive to the location, register an population of alpha helices[25] and beta turns.[26] The HN–N RDCs of DCL1-A show no apparent baseline, but a clear helical pattern is apparent in the helix 2 region. The selected subensemble fits very well all the experimental NMR data, and was used in the subsequent analysis (Fig. 1).

We employed SAXS data to obtain information on global shape of the ensemble. The Kratky plot shows the typical features of disordered proteins. Fit of the Guinier region reports a radius of gyration (R_g) of 22.1 Å (Fig. S3, ESI†). The R_g for the folded form of DCL1-A is predicted to be 16.4 Å while the R_g for the fully denatured protein, calculated as in Kohn *et al.*, 2004,[27] is 26.2 Å. The intermediate value of the experimental R_g of DCL1-A further confirms the partially folded nature of the free protein.

Structural alignment of the full proteins from the final selected ensemble does not show any common conformational feature. We therefore performed local alignments along the sequence on sliding windows. We found an optimal window of 7 residues, with shorter windows giving too noisy local RMSD values and longer windows not showing significant differences. The most prominent feature highlighted by low local RMSD values is the region corresponding to helix 2 (Fig. 1). Overlay of the structures from the ensemble in this region shows the N-terminal end of helix 2 adopts a conformation similar to that observed in the complex with RNA. Going further towards the C-terminus, residues around E68 seem to cluster in two groups with low RMSD within them and higher differences between
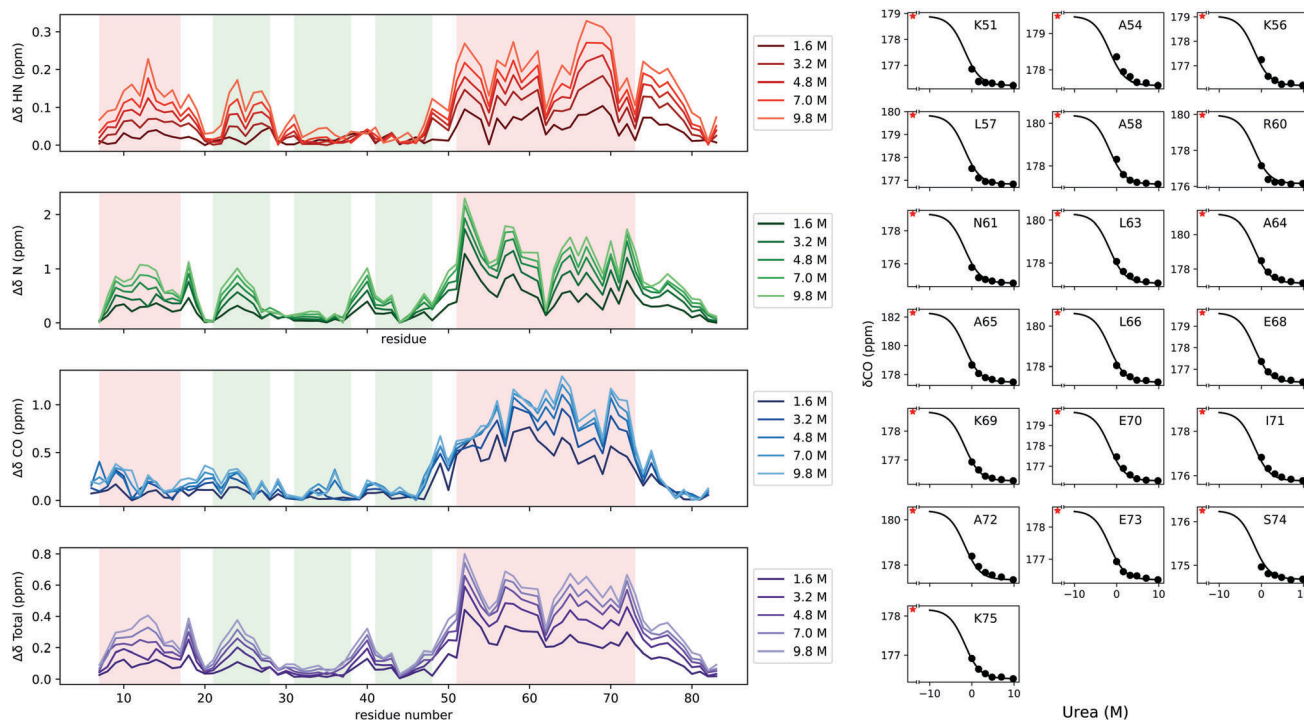
them. In this position the direction of the main chain splits in two, with a fraction of the conformers producing a longer helix and another showing a kink in the trajectory. The potential significance of this kink in the helix will be discussed below.

Stability of the disordered state of DCL1-A

Urea induced unfolding of residual structure. We have previously observed that the spectrum of DCL1-A changes significantly in the presence of 8 M urea, evidencing the presence of residual structure being denatured.[8] We decided to follow the complete unfolding of the structure by stepwise titration of a solution of DCL1-A with urea at 298 K. We acquired an HNCO spectrum at each point of the titration. This allowed to follow each signal's path more precisely while giving at the same time the ^{13}CO chemical shift as an additional probe of the unfolding process. At the final point (9.8 M urea) we acquired a full set of triple resonance assignment spectra. Further acquisition of ^{1}H–^{15}N HSQC spectra after dilution of the sample with buffer confirmed that urea induced unfolding of residual structure in DLC1-A is reversible.

The profile of the chemical shift evolution along the sequence show a similar picture for all three nuclei detected, ^{1}HN, ^{15}N and ^{13}CO (Fig. 2). The most affected regions are helix 2, helix 1 and beta sheet 1, in this order. Residues corresponding to beta sheets 2 and 3 in the folded protein are mostly unaffected, with the exception of the shifts observed for amide ^{15}N in the beta 2–beta 3 turn, suggesting that this turn is partially folded in the free protein. Quite strikingly, only residues corresponding to helix 2 show significant ^{13}CO shifts, with the rest of the protein being only slightly affected. Far UV CD spectra of the protein also show the loss of structure in the presence of urea as a decrease in ellipticity at 220 nm (Fig. S4, ESI†).

11240 | *Phys. Chem. Chem. Phys.*, 2018, **20**, 11237–11246

This journal is © the Owner Societies 2018

Fig. 2 Destabilization of DCL1A by urea at 298 K. Left, chemical shift changes as a function of sequence. From top to bottom, individual shifts for HN, N and CO nuclei, and combined distance. Regions corresponding to secondary structure elements on the folded form are shaded. Right, global fits of the chemical urea induced CO shift changes along helix 2. The graph is extended to unrealistic negative urea concentration values to show the shape of the extrapolated unfolding curve. The red star shows the chemical shift of the bound folded form, which is used as the limit value in the fit (y_U in eqn (1)).

The weighed residue specific chemical shift changes *vs.* urea concentration for most residues display a saturation behaviour at increasing urea concentrations (Fig. S5, ESI†). [1]HN shifts along beta sheets 2 and 3 are quite unusual, but their absolute values are rather small compared to other regions.

Chemical shifts of [1]HN and ^{15}N nuclei are affected by the backbone torsion angles bracketing the peptide bond, but also significantly by hydrogen bonding interactions. Their relationship to backbone conformation and secondary structure is therefore not direct. In contrast the [13]CO nuclei secondary shifts are mostly defined by the backbone dihedral angles and can be related directly to the conformation of the residue.[4,28,29] Using the known [13]CO chemical shift of the folded protein and the [13]CO chemical shift at 9.8 M urea as limiting value (y_U in eqn (1)) we calculated the fraction of folded conformation for each residue at each concentration of urea. We then applied the linear extrapolation method, as detailed in the Materials and methods section to obtain estimates for the stability of the protein in the absence of urea ($\Delta G_{U,H_2O}$) and of its dependence on urea concentration (*m*-value). Fits of the data for individual residues along the helix 2 region are very good, giving values for $\Delta G_{U,H_2O}$ and *m*-value of around −0.5 kcal mol[−1] and 0.4 kcal (mol M)[−1] respectively (Fig. S6, ESI†). The negative values of $\Delta G_{U,H_2O}$ are consistent with the protein being partially unfolded in the absence of denaturant.

Although residue specific information can be considered valuable, the model used is based on a global transition from the native to the unfolded state. Besides this, small errors in the

extrapolation to the native [13]CO shift can induce large variations in the parameters calculated for individual residues. We therefore obtained a global fit for residues K56-A72 along the helix 2 region. Fits for individual residues are still very good, and give values of −0.58 kcal mol[−1] and 0.32 kcal (mol M)[−1] for $\Delta G_{U,H_2O}$ and *m* respectively (Fig. 2). These values correspond to the global folding of the whole helical region.

Myers *et al.* demonstrated that *m*-values of unfolding for globular proteins depend linearly on the difference between the solvent accessible surface area (SASA) of the folded and unfolded protein.[30,31] We calculated SASAs for folded (4486 Å[2]) and unfolded (5840 Å[2]) forms of DCL1-A using the software freesasa.[32] To obtain the estimate for the unfolded protein we calculated the average SASAs on 1000 conformers of the initial unconstrained ensemble. The calculated *m*-value for DCL1-A using the parameters from Scholtz *et al.*[31] is 0.41 kcal (mol M)[−1], 28% higher than the *m*-value obtained from the fit of [13]CO data. This further confirms that the unfolding event monitored in the experiment does not correspond to the cooperative unfolding of the whole protein, but rather to the unfolding of helix 2.

Stabilization of folded conformations at low temperature. In our previous work we observed that lowering the temperature from 298 K to 278 K produces significant alterations in the ^{1}H–^{15}N HSQC spectrum of DCL1-A, and that the protein shows low but positive [1]H–[15]N NOEs at 278 K. While this information indicates that the protein becomes more rigid at this lower temperature, the protein does not acquire a fully folded conformation. We decided to acquire HNCO and HNCA spectra

between 278 K and 298 K with 5 K intervals to better understand the transition.

The temperature dependence of HN chemical shifts is a good indicator of the formation of hydrogen bonds in the backbone. All measured slopes fall well above the values corresponding to hydrogen bonded HN atoms (<5 ppb K^{-1} ref. 33). However, there is a clear reduction of the slopes in the helix 2 region, further supporting the presence of short lived hydrogen bonds as the helix is transiently populated (Fig. S7, ESI†).

We used the assigned backbone chemical shifts to estimate the population of residual secondary structure at each temperature by using the d2D algorithm.[24] Observation of the d2D results reveals that the protein starts populating conformations corresponding to the folded form as the temperature decreases (Fig. 3). Far UV CD spectra acquired as a function of temperature also show the loss of secondary structure in the protein as the temperature increases (Fig. S8, ESI†). Although the overall change in CD spectra is small, singular value decomposition of the whole set shows clearly a helical component that is increasingly populated at lower temperatures. At 278 K not only the region corresponding to helix 2, but also the regions corresponding to beta sheets 2 and 3 and, to a lesser extent, the region corresponding to beta sheet 1, show some secondary structure propensity. Whereas the coil conformation dominates the secondary structure population even at the lowest temperature tested, the central part of helix 2 populates more helical than coil conformations below 288 K, indicating the presence of a nascent helix and pinpointing the

region as a seed for DCL1-A folding. In contrast, the region corresponding to helix 1 remains fully disordered in the temperature range tested.

We then quantified the percentage of folding in a similar way as for the urea titrations. We obtained the d2D profile for the folded protein and used the values as the limit of fully folded protein. This allowed us to obtain a value for the folding equilibrium of each residue at each temperature. We used eqn (3) to describe the evolution of the free energy of unfolding *vs.* temperature:[34]

$$\Delta G(T) = \frac{T_m - T}{T_m}\Delta H(T_m) + \Delta C_p(T - T_m) + T\Delta C_p \ln\left(\frac{T_m}{T}\right) \tag{3}$$

The derivation of this equation assumes a constant value for ΔC_P in the temperature range studied, allowing to define $\Delta G(T)$ as a function of only three parameters: T_m, $\Delta H(T_m)$, and ΔC_P.

We obtained fits for residues along helix 2 (Fig. 3). As most points fall below the melting temperature, the curvature of the graph is ill defined and leaving ΔC_P as a fitting parameter gives unrealistic values for some residues. We decided therefore to calculate the value of ΔC_P from the $\Delta SASA$,[30] and fix the value obtained (0.15 kcal mol^{-1} K^{-1}) in the data fitting. In this manner we obtained estimations for T_m and $\Delta H(T_m)$ from fits of the d2D *vs.* temperature data for each residue along the helix.
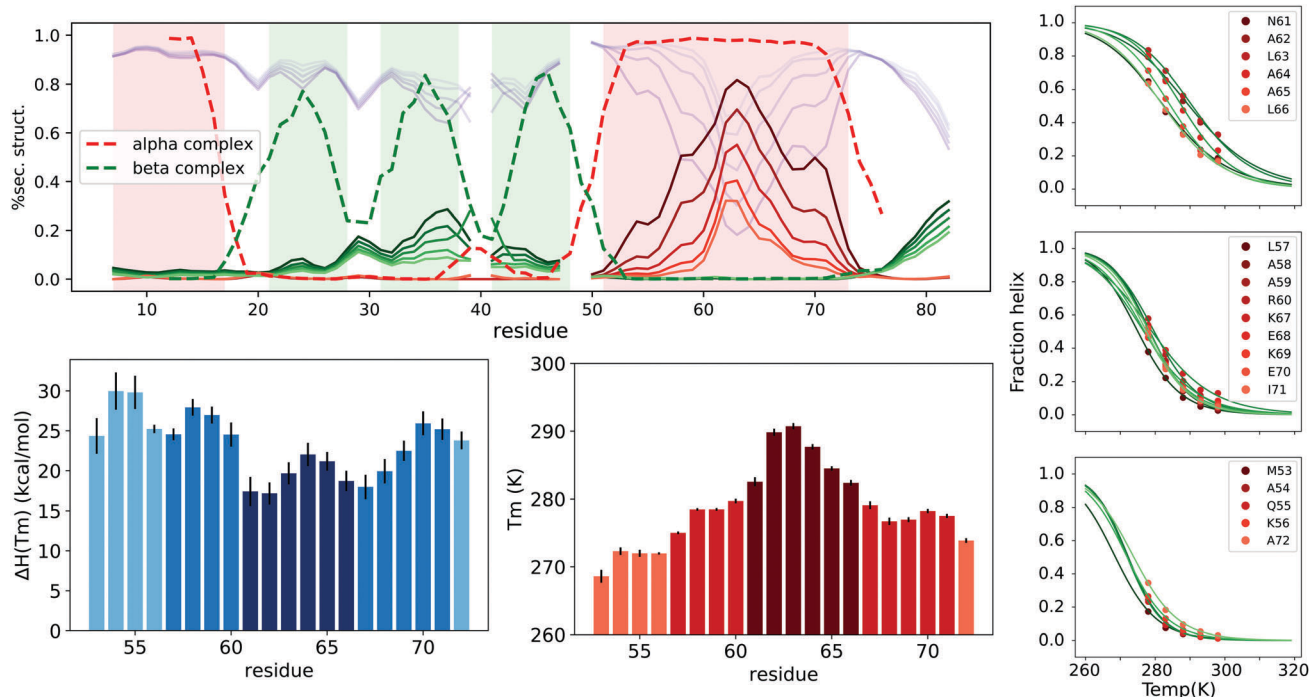

Fig. 3 Stabilization of the folded forms at lower temperature. Top, secondary structure populations along the sequence calculated using d2D. Temperatures go from 278 K to 298 K in 5 K increments, and are coloured in gradient from the lower (dark) to the higher (light) temperatures. Alpha helix content is shown in full lines, beta sheet in dashed lines and random coil full light grey gradient. The d2D profile of bound folded DCL1-A is shown with long dashed lines. Regions corresponding to secondary structure elements on the folded form are shaded in red for helices and in green for sheets. Middle, $\Delta H(T_m)$ and T_m values obtained from fitting the data for individual residues along helix 2. Three regions with comparable T_m values are highlighted. Below, data points and fit. Plots are separated according with the T_m values, as highlighted in the middle panel.

Fixing ΔC_P to zero, as sometimes assumed for single helices or small proteins,[35,36] gives similar results as those obtained with the ΔC_P value calculated from ΔSASA (Fig. S9, ESI†).

The stability of helix 2 varies along the sequence, as judged by the T_m values (Fig. 3). The most stable central part (residues 61–66), with T_m values above 280 K, is flanked by two stretches of 4–5 residues (57–60 and 67–72) with similar T_m (*ca.* 278 K). Four further residues at the N-terminus (53–56) present T_m values of *ca.* 271 K, and the stability decays rapidly before residue 52 and after residue 73.

Individual $\Delta H(T_m)$ values have no physical meaning, as the underlying model assumes a global unfolding event. The average $\Delta H(T_m)$ value obtained for the transition is 23 kcal mol^{-1}. This value is comparable to that expected for a simple helix–coil transition of the 20-residues helix 2 (considering per-residue enthalpy change ≈ 1 kcal per mol per residue[35]). Therefore, the process reported by the thermal stabilization of DCL1-A in the temperature range studied would correspond to the formation of helix 2, occurring independently form the whole protein folding.

Discussion

The system formed by DCL1-A and its substrate RNA constitutes an interesting model to further the knowledge of binding induced folding in nucleic acid:protein complexes. A thorough description of the conformational space of the free protein is necessary if one wishes to draw conclusions on the coupling of folding with binding.[37] It is believed that in general the unfolded state of proteins may dictate the conformations open for exploration during the folding process by restricting the conformational space available to the polypeptide chain. These restrictions channel the folding pathway thus accelerating protein folding. Therefore, the presence of residual structure in the unfolded state could play an important role in protein folding. DCL1-A, being mostly disordered in its unbound form, lends itself to the study of the unfolded state. In the present study we obtained an in-depth picture of free DCL1-A in terms of its conformational preferences and of its location in the folding pathway.

Different disorder [disopred,[38] IUPred,[39] PrDOS[40]] and secondary structure prediction [jpred;[41] s2d;[42] Agadir[43]] algorithms confirm that DCL1-A is expected to fold as a dsRBD based on its aminoacid sequence: the protein is predicted to have low disorder tendency and all secondary structure elements are correctly predicted (Fig. S10, ESI†). However CD, NMR and SAXS results show clearly that the protein is extended and mostly disordered in its free state, while retaining residual structure.

Using chemical shifts, RDCs and SAXS data as restraints we obtained an ensemble description of DCL1-A. The calculated ensemble shows that two of the three dsRNA binding sites are found in highly unstructured regions of the protein, allowing them to explore the surrounding space with a large degree of freedom. In contrast, the third binding site is in a partially folded region. In canonical dsRBDs, this site is precisely one of the determinants that allows specific binding of dsRNA over other nucleic acids.[44–46] The variation observed in orientations

of the side chains of the residues in this binding site in the calculated ensemble could be conferring versatility in the binding to the substrate. At the same time, the preformed global structure shows that a conformational selection mechanism is operating in RNA recognition, at least locally.

Helix 2 presumably constitutes a nucleation site for the folding event. The long helix contacts all the other secondary structure elements in the folded conformation. Analysis of the distance distribution in the conformational ensemble does not show any preferential long-range contacts. The change in the direction of the chain observed in L66 can be related to the known structures of two yeast dsRBDs, RNase III Rnt1p[47] and Dicer Dcr1.[48] Both dsRBDs present a third helix in the C-terminal region. In the case of Rnt1p, deletion of helix 3 decreases the affinity of the enzyme for the substrate. This helix establishes a series of contacts with the helix 1 of the domain, that are important for the stability and could be inducing the proper conformation in helix 1 to allow for substrate binding. Although the final folded conformation of DCL1-A does not show a break in the helix, the two conformation clusters could allow for two alternative paths of folding-induced binding, mediated by different interactions with helix 1.

Information on the relative energies of the fully unfolded and partially disordered state found in standard conditions of *in vitro* analysis is particularly important.[49] To describe the location of free DCL1-A in the folding pathway and obtain a quantitative understanding of its conformational landscape we analysed the destabilization of the residual structure with urea and the increase in population of structured segments at lower temperatures. The data obtained with both approaches could be fit to well established thermodynamic models of folding transitions. Whereas for most optical spectroscopy probes (CD, absorption, fluorescence infrared) the accuracy of thermodynamic fits relies on precise knowledge of baseline evolution and on the shape of the transition curve, the use of NMR chemical shifts allow a reliable estimation of the fraction of folded and unfolded protein with residue resolution under each condition.

Unfolding of the residual structure using urea showed the presence of residual structure all along the protein sequence. For the region corresponding to helix 2 in the folded conformation, we could obtain quantitative information by fitting the ^{13}CO chemical shift data. This revealed that the region is almost halfway in the folding pathway, with slightly negative $\Delta G_{U,H_2O}$ values corresponding to an equilibrium constant of *ca.* 3 for the unfolding reaction, that is there is three times as much unfolded protein as folded protein in standard buffer conditions. The *m*-value obtained is somewhat below the value calculated for DCL1-A folding, based on the ΔSASA. Reduction of the *m*-value could be due to a deviation from a two-state mechanism,[31] suggesting that the transition observed corresponds to the unfolding of an intermediate species and hence that the free form of DCL1-A is an actual intermediate in the folding pathway. The study of an extensive set of mutants of staphylococcal nuclease demonstrated that mutants exhibiting reduced *m*-values give rise to more compact forms of the protein domains in unfolding conditions.[31,50] Considering this, the partially folded
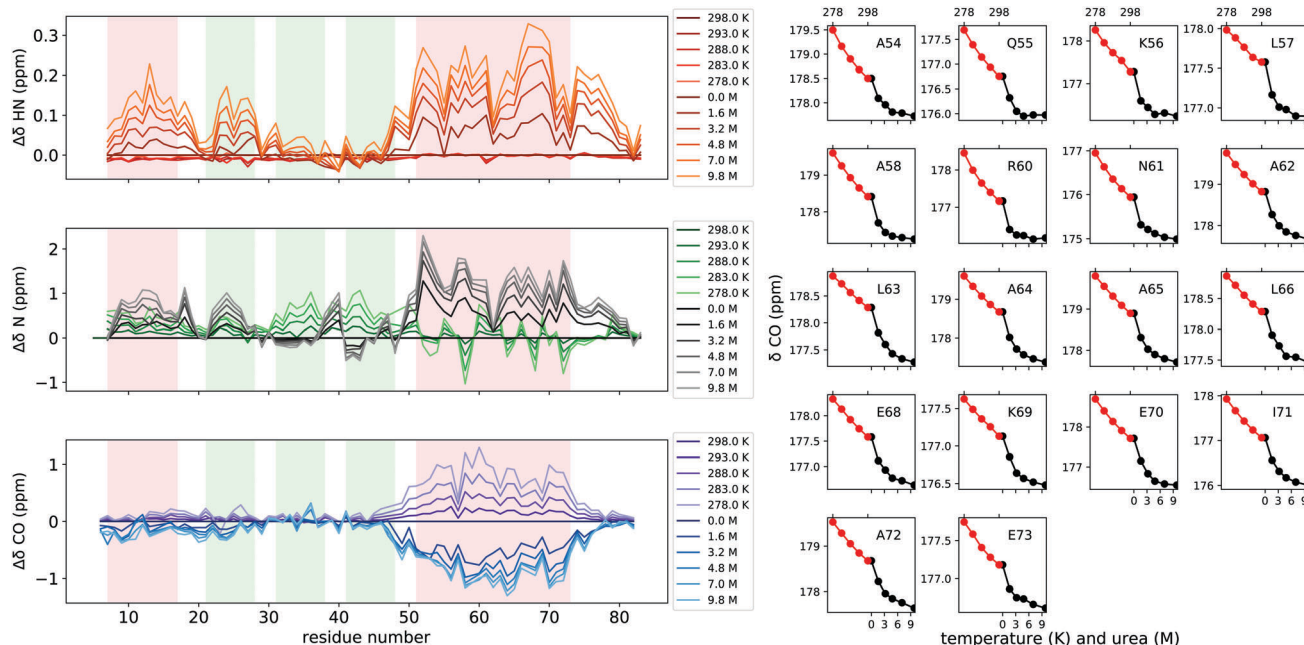
Fig. 4   Left, chemical shift evolution as a function of temperature or urea concentration for HN, N and CO nuclei. Each line corresponds to the difference in chemical shift between the condition reported in the legend and the chemical shift at 298 K in the absence of urea. HN chemical shifts *vs.* temperature are plotted subtracting the linear temperature dependence (shown in Fig. S7, ESI†). Regions corresponding to secondary structure elements on the folded form are shaded. Right, sequential loss of helical propensity with temperature increment and urea concentration reported by CO chemical shifts. Temperature data (gray) and urea titration data (black) are plotted on the same CO shift axis for residues along the helix 2 region.

nature of disordered free DCL1-A could result in an *m*-value lower than the one calculated on the basis of ΔSASA. Alternatively, compaction of the denatured state ensemble at low denaturing concentrations could also result in non-constant *m*-values.[51]

Lowering the temperature leads to a stabilization of the partial structures along the sequence, extending even to the beta 2–beta 3 region. We could also obtain good fits of the temperature dependence data to a simple unfolding model for the helix 2 region. The central region, where the helix is most stable, could behave as a seed for the subsequent folding of the remainder of the helix and eventually of the whole protein. The calculated $\Delta H(T_{\mathrm{m}})$ corresponds to that expected for the folding of the helical region alone, suggesting that we are monitoring the helix–coil transition in these experiments, rather than the folding transition of the whole protein. Dependence of folding parameters with temperature suggest that in the free form DCL1-A behaves almost as a downhill folder, where structural elements are formed sequentially and converge to a fully folded conformation with a low energy barrier.

It is informative to study the datasets of urea unfolding and temperature stabilization together. To observe the evolution of DCL1-A towards the folded form or the fully disordered state as a whole, we plotted the chemical shifts in different conditions with respect to those of the protein in standard buffer at 298 K (Fig. 4). The plot shows clearly that helix 2 is halfway folded in the free protein at 298 K. The simultaneous observation of urea unfolding and low temperature stabilization of DCL1-A gives a further hint on the underlying cause of its disordered nature, that contradicts prediction algorithms. The region corresponding to helix 1 does not show helical propensity, as revealed by d2D

analysis of the different chemical shift datasets. In fact, alternate residues along the region exhibit secondary chemical shifts that denote a helical preference, but not having contiguous helical conformational tendencies the helix cannot form. At the same time, urea denaturation results in changes of the HN and N chemical shifts almost comparable to those observed in the well-populated helix 2 region, without corresponding changes in CO chemical shifts (Fig. 4). Taken together the observations suggest that the helix 1 region does have a residual structure that is disrupted by urea, but does not correspond to the secondary structure of the folded protein. Having this N-terminal region locked in a non-native conformation would hinder the acquisition of the native fold in the free form of DCL1-A. Helix 1 is one of the regions that participate in substrate binding in dsRBDs.[46] These considerations allow us to propose a sequence for binding induced folding of DCL1-A. The formation of a first encounter complex would be facilitated by binding region 3, in the N-terminus of helix 2, that populates conformations corresponding to the folded form. Once the protein is anchored to substrate RNA, interaction of the sidechains of residues in region 1 with the RNA backbone would induce the correct conformation on helix 1.[46] When the N-terminal region adopts a helical structure, the rest of the protein would collapse to the folded conformation, forming a stable complex with its substrate RNA.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 V. N. Uversky, Intrinsic disorder here, there, and everywhere, and nowhere to escape from it, *Cell. Mol. Life Sci.*, 2017, **74**, 3065–3067.

2 R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Old, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, Classification of Intrinsically Disordered Regions and Proteins, *Chem. Rev.*, 2014, **114**, 6589–6631.

3 J. Habchi, P. Tompa, S. Longhi and V. N. Uversky, *Chem. Rev.*, 2014, **114**, 6561–6588.

4 M. R. Jensen, M. Zweckstetter, J. Huang and M. Blackledge, Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy, *Chem. Rev.*, 2014, **114**, 6632–6660.

5 M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. Wishart, P. E. Wright and P. Tompa, PE-DB: A database of structural ensembles of intrinsically disordered and of unfolded proteins, *Nucleic Acids Res.*, 2014, **42**, 326–335.

6 S. E. Schauer, S. E. Jacobsen, D. W. Meinke and A. Ray, DICER-LIKE1: Blind men and elephants in Arabidopsis development, *Trends Plant Sci.*, 2002, **7**, 487–491.

7 Q. Liu, Q. Yan, Y. Liu, F. Hong, Z. Sun, L. Shi, Y. Huang and Y. Fang, Complementation of Hyponastic Leaves1 by Double-Strand RNA-Binding Domains of Dicer-Like1 in Nuclear Dicing Bodies, *Plant Physiol.*, 2013, **163**, 108–117.

8 I. P. Suarez, P. Burdisso, M. P. M. H. Benoit, J. Boisbouvier and R. M. Rasia, Induced folding in RNA recognition by Arabidopsis thaliana DCL1, *Nucleic Acids Res.*, 2015, **43**, 6607–6619.

9 E. Lescop, P. Schanda and B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, *J. Magn. Reson.*, 2007, **187**, 163–169.

10 M. Rückert and G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc.*, 2000, **122**, 7793–7797.

11 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR*, 1995, **6**, 277–293.

12 V. Y. Orekhov and V. A. Jaravine, Analysis of non-uniformly sampled spectra with Multi-Dimensional Decomposition, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **59**, 271–292.

13 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline, *Proteins*, 2005, **59**, 687–696.

14 J. A. Marsh, V. K. Singh, Z. Jia and J. D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation, *Protein Sci.*, 2006, **15**, 2795–2804.

15 D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries and D. I. Svergun, ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *J. Appl. Crystallogr.*, 2017, **50**, 1212–1225.

16 P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J. Appl. Crystallogr.*, 2003, **36**, 1277–1282.

17 D. I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr.*, 1992, **25**, 495–503.

18 M. R. Jensen, L. Salmon, G. Nodet and M. Blackledge, Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts, *J. Am. Chem. Soc.*, 2010, **132**, 1270–1272.

19 V. Ozenne, R. Schneider, M. Yao, J. Huang, L. Salmon, M. Zweckstetter, M. R. Jensen and M. Blackledge, Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution, *J. Am. Chem. Soc.*, 2012, **134**, 15138–15148.

20 Y. Shen and A. Bax, SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, *J. Biomol. NMR*, 2010, **48**, 13–22.

21 S. Chaudhury, S. Lyskov and J. J. Gray, PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta, *Bioinformatics*, 2010, **26**, 689–691.

22 M. Zweckstetter and A. Bax, Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR, *J. Am. Chem. Soc.*, 2000, **122**, 3791–3792.

23 T. O. Street, N. Courtemanche and D. Barrick, Protein Folding and Stability Using Denaturants, *Methods Cell Biol.*, 2008, **84**, 295–325.

24 C. Camilloni, A. De Simone, W. F. Vranken and M. Vendruscolo, Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts, *Biochemistry*, 2012, **51**, 2224–2231.

25 M. R. Jensen and M. Blackledge, On the Origin of NMR Dipolar Waves in Helical Elements of Partially Folded Proteins, *J. Am. Chem. Soc.*, 2008, **130**, 11266–11267.

26 M. D. Mukrasch, P. Markwick, J. Biernat, M. Von Bergen, P. Bernadó, C. Griesinger, E. Mandelkow, M. Zweckstetter

and M. Blackledge, Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation, *J. Am. Chem. Soc.*, 2007, **129**, 5235–5243.

27 J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnik, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach and K. W. Plaxco, Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12491–12496.

28 M. V. Berjanskii and D. S. Wishart, Unraveling the meaning of chemical shifts in protein NMR, *Biochim. Biophys. Acta, Proteins Proteomics*, 2017, **1865**, 1564–1576.

29 W. Shalongo, L. Dugad and E. Stellwagen, Analysis of the thermal transitions of a model helical peptide using 13C NMR, *J. Am. Chem. Soc.*, 1994, **116**, 2500–2507.

30 J. K. Myers, C. N. Pace and J. M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci.*, 1995, **4**, 2138–2148.

31 J. M. Scholtz, G. R. Grimsley and C. N. Pace, Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value, *Methods Enzymol.*, 2009, **466**, 549–565.

32 S. Mitternacht, FreeSASA: An open source C library for solvent accessible surface area calculations, *F1000Research*, 2016, **5**, 1–12.

33 T. Cierpicki and J. Otlewski, Amide proton temperature coefficients as hydrogen bond indicators in proteins, *J. Biomol. NMR*, 2001, **21**, 249–261.

34 P. Privalov and S. Gill, Stability of Protein Structure and Hydrophobic Interaction, *Adv. Protein Chem.*, 1988, **39**, 191–234.

35 J. Seelig and H.-J. Schönfeld, Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding, *Q. Rev. Biophys.*, 2016, **49**, e9.

36 G. S. Kubelka and J. Kubelka, Site-Specific Thermodynamic Stability and Unfolding of a de Novo Designed Protein Structural Motif Mapped by 13C Isotopically Edited IR Spectroscopy, *J. Am. Chem. Soc.*, 2014, **136**, 6037–6048.

37 M. Arai, K. Sugase, H. J. Dyson and P. E. Wright, Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9614–9619.

38 D. T. Jones and D. Cozzetto, DISOPRED3: Precise disordered region predictions with annotated protein-binding activity, *Bioinformatics*, 2015, **31**, 857–863.

39 Z. Dosztányi, V. Csizmok, P. Tompa and I. Simon, IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics*, 2005, **21**, 3433–3434.

40 T. Ishida and K. Kinoshita, PrDOS: Prediction of disordered protein regions from amino acid sequence, *Nucleic Acids Res.*, 2007, **35**, 460–464.

41 A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, JPred4: A protein secondary structure prediction server, *Nucleic Acids Res.*, 2015, **43**, W389–W394.

42 P. Sormanni, C. Camilloni, P. Fariselli and M. Vendruscolo, The s2D method: Simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins, *J. Mol. Biol.*, 2015, **427**, 982–996.

43 V. Munoz and L. Serrano, Elucidating the folding problem of helical peptides using empirical parameters, *Nat. Struct. Mol. Biol.*, 1994, **1**, 399–409.

44 G. Masliah, P. Barraud and F. H. T. Allain, RNA recognition by double-stranded RNA binding domains: A matter of shape and sequence, *Cell. Mol. Life Sci.*, 2013, **70**, 1875–1895.

45 L. Vukovic, H. R. Koh, S. Myong and K. Schulten, Substrate Recognition and Specificity of Double-Stranded RNA Binding Proteins, *Biochemistry*, 2014, **53**, 3457–3466.

46 S. I. Drusin, I. P. Suárez, D. F. Gauto, R. M. Rasia and D. M. Moreno, dsRNA-protein interactions studied by molecular dynamics techniques. Unravelling dsRNA recognition by DCL1, *Arch. Biochem. Biophys.*, 2016, **596**, 118–125.

47 N. Leulliot, S. Quevillon-Cheruel, M. Graille, H. van Tilbeurgh, T. C. Leeper, K. S. Godin, T. E. Edwards, S. T. L. Sigurdsson, N. Rozenkrants, R. J. Nagel, M. Ares and G. Varani, A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III, *EMBO J.*, 2004, **23**, 2468–2477.

48 P. Barraud, S. Emmerth, Y. Shimada, H.-R. Hotz, F. H.-T. Allain and M. Bühler, An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer, *EMBO J.*, 2011, **30**, 4223–4235.

49 A. C. Hausrath and R. L. Kingston, Conditionally disordered proteins: bringing the environment back into the fold, *Cell. Mol. Life Sci.*, 2017, **74**, 3149–3162.

50 D. Shortle, Staphylococcal nuclease: a showcase of m-value effects, *Adv. Protein Chem.*, 1995, **46**, 217–247.

51 E. P. O. Brien, B. R. Brooks and D. Thirumalai, Molecular origin of constant m-values, collapse transition, and residue-dependent transition midpoints in globular proteins, *Biochemistry*, 2009, **48**, 3743–3754.

11246 | *Phys. Chem. Chem. Phys.*, 2018, **20**, 11237–11246

This journal is © the Owner Societies 2018