Biochemistry

UPMC SORBONNE UNIVERSITÉS

Bibliothèque Universitaire

Subscriber access provided by BUPMC - Bibliothèque Universitaire Pierre et Marie Curie

Article

Hydrophobic Collapse of the Intrinsically Disordered Transcription Factor MAX

Gönül Kizilsavas, Karin Ledolter, and Dennis Kurzbach

Biochemistry, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.7b00679 • Publication Date (Web): 07 Sep 2017

Downloaded from http://pubs.acs.org on September 8, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

Biochemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Hydrophobic Collapse of the Intrinsically Disordered Transcription Factor MAX

Gönül Kizilsavas[a,‡], Karin Ledolter[b,‡] and Dennis Kurzbach[c,d,]*

[a] *Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

[b] *University Vienna, Department for Structural and Computational Biology, Max F. Perutz Laboratories, Campus Vienna BioCenter 5, 1030 Vienna, Austria*

[c] *Département de Chimie, Ecole Normale Supérieure, PSL Research University, UPMC Univ Paris 06, CNRS, Laboratoire des Biomolécules (LBM), 24 rue Lhomond, 75005 Paris, France*

[d] *Sorbonne Universités, UPMC Univ Paris 06, Ecole Normale Supérieure, CNRS, Laboratoire des Biomolécules (LBM), Paris, France*

*corresponding author: DK, email: kurzbach@ens.fr, phone: +33 1 44323265

ABSTRACT The conformational space of the proto-oncogenic transcription factor MAX (Myc associated factor X) comprises a dynamic equilibrium between a stably folded coiled-coil homodimer and an intrinsically disordered ensemble of states. We show by means of nuclear magnetic resonance spectroscopy that the intrinsically disordered ensemble samples structures that are even as compact as the folded dimer. These extremely dense, hydrophobically collapsed globules might be of importance for interconversion between different conformations of intrinsically disordered proteins.

KEYWORDS Intrinsically Disordered Proteins, NMR, Paramagnetic Relaxation Enhancement, Polymer Models, Collapsed Globule.

INTRODUCTION

The MYC-associated factor X (MAX) is a proto-oncogenic transcription factor that heterodimerizes with its partner protein MYC to develop transcriptional activity.[1-2] MYC and MAX are driving forces in almost all human cancers as they regulate the transcription of proteins involved in DNA replication and cell proliferation and are therefore involved in high activity.[3] For both, MYC and MAX the central structural element is a basic/helix-loop-helix/leucine zipper ($b/H_1LH_2/LZ$) conformation[4-5] - a motif that is also found important for MYC:MAX heterodimerization.

In the absence of MYC, MAX populates a heterogeneous conformational space including intrinsically disordered conformations as well as a stably folded coiled-coil homodimer that adopts the $b/H_1LH_2/LZ$ conformation.[6] The dynamic sampling of such heterogeneous conformational spaces includes often the transient population of conformations that are devoid of any secondary or tertiary structure elements, but show varying compactness along the polypeptide chain.[7-8] Such conformations are frequently involved in substrate interactions (conformational selection, entropic compensation) or folding processes where they may act as folding nuclei and rate regulating entities.[7,8] Understanding these states might help to tackle the still not resolved problem of protein folding and conformational averaging as well as to elucidate the relation between the complicated conformational dynamics of the MYC-MAX proto-oncoprotein network and the development of cellular transcriptional activity.

We here report by means of nuclear magnetic resonance (NMR) a novel intrinsically disordered state of MAX, which features a very high compactness. It can be described as hydrophobically collapsed globule of high density comparable to that of the stably folded coiled-coil $b/H_1LH_2/LZ$

structure of the homodimer (denoted as MAX_2) in contrast to the monomer (denoted as MAX). The described intrinsically disordered state might be of importance for the formation of the transcriptionally active MYC:MAX dimer that depends on the concentration of ready-to-react monomeric MAX in the cellular environment.

MATERIALS/EXPERIMANTEL DETAILS

NMR

Heteronuclear single quantum coherence (HSQC) spectra were recorded at 35 °C using a Varian Direct Drive 800 MHz spectrometers. Spectra were recorded in the States-TPPI/PFG sensitivity-enhanced mode for quadrature detection with carrier frequencies for $^{1}H^{N}$ and ^{15}N of 4.73 and 120.0 ppm, respectively. The samples contained 1.2 mM MAX, 25 mM MES, and 25 mM NaCl (pH 5.5) in a 90% H_2O/10% D_2O mixture. Intramolecular paramagnetic relaxation enhancement (PRE) of MTSL labeled Cys mutants were carried out in a similar fashion. Reference measurements were performed after reduction of the *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) label with an excess of ascorbic acid.

All NMR spectra were processed and analyzed using NMRPipe, SPARKY and home-written scripts based on the MATLAB program package. A squared and 60° phase-shifted sine bell window function was applied in all dimensions for apodization. Time domain data were zero-filled to twice the data set size, prior to Fourier transformation. The applied pulse sequences are sensitivity-enhanced and use gradients for coherence selection and water suppression. NMR resonance assignment (~80%) was achieved by a combination of several three-dimensional techniques (HNN, HN(C)N[9] and HNCACB[10] spectroscopy.

Relaxation Measurements.

^{15}N longitudinal relaxations experiments for evaluating T_1 were performed at 800 MHz Larmor frequency with relaxation delays of 0, 50, 150 and 250 ms using a recycling delay of 2 s.

^{15}N transverse relaxations experiments for evaluating T_2 were performed at 800 MHz Larmor frequency with relaxation delays of 0, 16, 32, 64, 128, 192, and 256 ms using a recycling delay of 1 s.

In both cases, the sequence of cross-peak intensities in this series of spectra was fitted to a two-parameter exponential decay of the form

$$I(t) = A \exp[(-\tau/T_2)]$$

where $I(t)$ represents the peak intensity and τ the delay time.

Heteronuclear steady-state NOE $^{15}N\{^1H^N\}$ attenuation factors were derived at 14.1 T from the $\eta = I_{NOE}/I_{REF}$ ratio, where I_{NOE} and I_{REF} denote the peak intensities in the experiments with and without proton saturation. In the case of spectra without saturation, a net relaxation delay of 5 s was employed whereas a relaxation delay of 2 s prior to a 3 s proton presaturation period was applied for the NOE spectra. Acquisition parameters were identical to those of the PFG sensitivity-enhanced 2D ^{1}H-^{15}N HSQC experiments.

Reference PRE Measurements

Intermolecular paramagnetic relaxation enhancement (PRE) reference measurements were carried out at different concentrations between starting from 0.2 mM wt-^{15}N-MAX + 0.2 mM MTSL-labeled ^{14}N-MAX to 0.4 mM wt-^{15}N-MAX + 0.4 mM MTSL-labeled ^{14}N-MAX (mutants

R5C, G35C and R55C) using the HSQC method as indicated above. Likewise reference measurements were carried out with 1 mM Mn^{2+}·EDTA or in the presence of 1 mM free MTSL in the solution.

Calculation of PRE rates

As pointed out by Wagner and co-workers, the intensity ratio *V* between oxidized and reduced state of an MTSL label of a protein sample can be expressed as:[14]

$$V = R_2 \exp(-\Gamma_2 t) / (R_2 + \Gamma_2) \qquad (1)$$

Where Γ_2 denotes the PRE rate and R_2 the intrinsic relaxation rate (absence of spin label) of any $^1H^N$ nucleus. *t* denotes the time that the proton magnetization is transverse through the HSQC experiment (9 ms). Assuming a simple spherical motion with a single correlation time τ_e the PRE rate is given by:[15]

$$\Gamma_2 = 1/15(\mu_0\gamma_H g\mu_B/4\pi)2S(S+1)[4J(0)+3J(\omega_L)]\langle r^{-6}\rangle \qquad (2)$$

with

$$J(\omega) = \tau_e/(1+\omega^2\tau_e^2) \qquad (3)$$

ω_L denotes the Larmor frequency of the proton. S denotes the electron spin number. All other symbols have their usual meaning. The average distance, <r>, between the unpaired electron and the nucleus of interest corresponds to the average distance, <*R*>, in eq. 5, obtained from the Flory-type collapsed globule model. Thus, by choosing N in eq. 5 corresponding to the number of residues between the unpaired electron and the amide proton under consideration we can calculate the residue dependence of the PRE rate for a hydrophobically collapsed coil.

Note that the assumption of a single correlation time is most-likely an oversimplification of the model, but here suffices to reproduce the experimentally observed PRE profiles.

DOSY

DOSY was measured on an 800 MHz Bruker HD III spectrometer equipped with a BCU II variable temperature unit using a Z-gradient 5 mm BBO probe. The pulse sequence and analysis method proposed by Wilkins et al.[16] (supplemented with a WATERGATE element to suppress the water signal) was used employing a 128-increment linear gradient profile. The data was processed and fitted using the DOSY toolbox.[17] In a 500 mL sample of 0.4 mM concentration 5 μL 2,4 dioxane were added as internal standard. Data for the disordered state were recorded at 35 ˚C and pH 5.5. Data for the folded state were recorded at 20 ˚C and pH 5.5.[6]

CPMG RD

Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD) NMR data were measured as suggested by Kay and co-workers[11] at 35 °C at 600 and 800 MHz proton Larmor frequency at CPMG frequencies, ν_{CPMG}, of 33, 67, 133, 167, 266, 333, 400, 467, 533, 600 and 667 Hz. The recycling delay was 1 s. The measured effective relaxation rates, $R_{2,eff}$, were fitted globally (considering all residues that showed relaxation dispersion) to the general two side exchange model of Carver and Richards assuming that $R_{2,0}$ is equal for both states A and B :

$$R_{2,eff}(\nu_{CPMG}) = 0.5\{ (R_{2,0} + k_{ex} - \tau_{cp}^{-1}) \operatorname{acosh}[D_+\cosh(\eta_+) - D_-\cos(\eta_-)]\}$$

$$D_\pm = 0.5\{\pm 1 + (\psi + 2(\delta\omega)^2)/[\psi^2 + (2\delta\omega(p_A - p_B)\, k_{ex}^2)^2]^{1/2}\}$$

$$\eta_\pm = \tau_{cp}/\sqrt{2}\ \{\pm\psi + [\psi^2 + (2\delta\omega(p_A - p_B)\, k_{ex}^2)^2]^{1/2}\}^{1/2}$$

$$\psi = [(p_A - p_B)\, k_{ex}]^2 - (\delta\omega)^2 + 4 p_A p_B k_{ex}^2 \qquad (4)$$

p_A and p_B denote the populations of states A and B. $R_{2,0} = 1/T_{2,0}$ denotes the relaxation rate in the absence of exchange.

TRACT

TRACT experiments where performed and analyzed as suggested in reference [18] at 18.8 T magnetic field strength at 35 ˚C.

Protein Expression and Purification

MAX was subcloned into a Pet3d expression vector and transformed into E.coli BL21 pLysS cells. Cells were grown at 37°C in M9 (for ^{13}C and ^{15}N labeling; 1 g/L ^{15}N ammonium chloride; 3 g/L ^{13}C glucose) and induced at an A600 nm of 0.5 with 0.5 mM IPTG at 30 °C over night. Cell pellets were homogenized in 20 mM PBS, 100 mM NaCl and 1 mM EDTA. For protein purification, fractional NH_4SO_4 precipitation (50% and 80% saturation) was carried out and anion exchange chromatography was applied. The final total protein concertation was 0.4 mM except for intermolecular reference measurements where concentrations up to 0.8 mM where tested to assure the absence of any intermolecular PREs.

Cystein mutants and MTSL (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate) labeled proteins were produced according to methods published earlier.[19-20] Excess spin label was removed by dialysis into the buffer used for the NMR experiments. PRE referencing was achieved by reduction of the MTSL label through incubation for 1 h with tenfold excess of ascorbic acid at 35 °C.

RESULTS AND DISCUSSION

Fig. 1a shows the structural motifs of the primary sequence of MAX_2. Each unit of the homodimer can be separated into a rather flexible basic N-terminal domain (NTD) between amino acid (aa) 1 and 16, which houses the primary DNA recognition motif and a very rigid H_1LH_2/LZ segment located between aa 16 and 82.

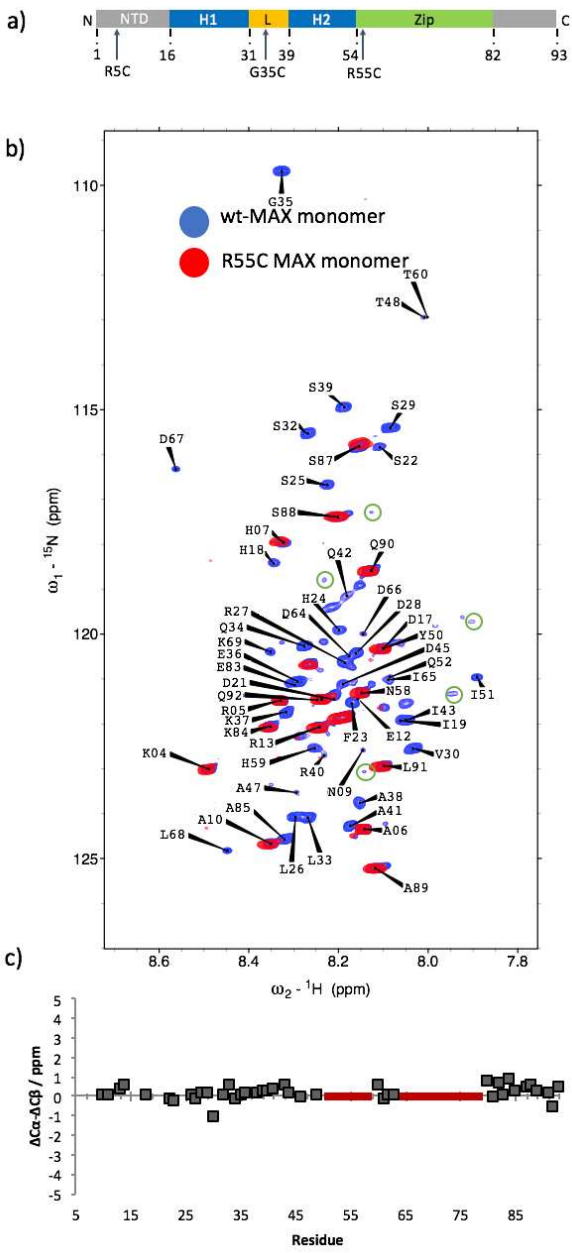


Figure 1. a) Schematic display of the structural constitution of a MAX_2 unit. The MTSL labeling sites R5, G35 and R55 employed here are indicated. b) ^{1}H-^{15}N 2D correlation NMR spectrum (HSQC) of the MAX monomer (blue) and of a spin labeled mutant (red) of the monomeric form. The amino acid assignment is indicated. A tabular form can be found in the Supporting Information. The green circles exemplarily indicate residues that could not be assigned due to

weak signal intensities (see text). c) ^{13}C secondary chemical shifts found for the MAX monomer. The red bar indicates stretches of not assigned residues.

For the study at hand, we chose experimental conditions of pH 5.5 and a temperature of 35 °C. Under these conditions, MAX_2 dissociates and the protein almost entirely populates a monomeric, intrinsically disordered state.[6] Such, conditions are relevant as MAX is residing close to DNA strings in the cellular environment, where pH values can drastically drop as compared to a physiological pH of 7.4.[21] Fig. 1b shows a ^{1}H-^{15}N 2D correlation NMR spectrum (heteronuclear single quantum coherence; HSQC) of the MAX monomer (blue). The chemical shift dispersion in the proton dimension is very low as expected for an intrinsically disordered protein (IDP). The lack of secondary or tertiary structure elements is further corroborated by residue resolved ^{13}C NMR secondary chemical shifts (Fig. 1c) – considered as deviation of Cα and Cβ chemical shifts from values predicted for random coils (denoted Δδ).[10] These shifts ($|\Delta\delta(C\alpha)|$ - $|\Delta\delta(C\beta)|$) are found to be close to zero throughout the entire primary sequence indicating the absence of any significant structural propensities in the MAX monomer, which would entail noteworthy deviations from zero.[10] For the folded homodimer, strong positive shifts ($\Delta\delta(C\alpha)$-$\Delta\delta(C\beta) > 0$) were observed in earlier studies[4] between aa 16 and 82 corresponding to the α-helical elements in the H_1LH_2/LZ domain (see Fig. 2 for the structure of the leucine zipper).[6]

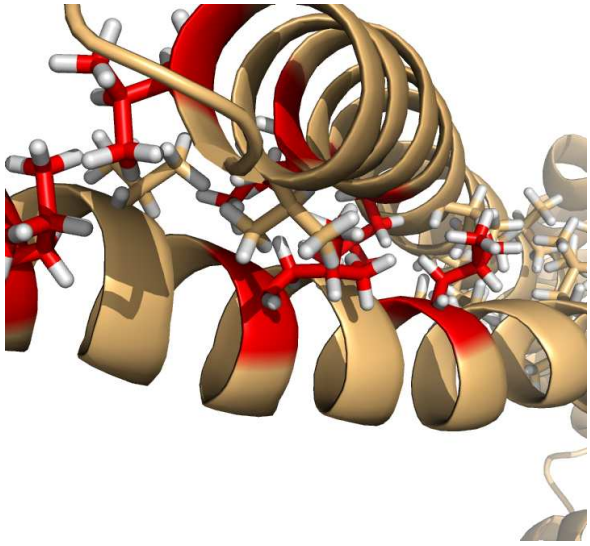


Figure 2. Zoom into the leucine zipper of MAX_2. Hydrophobic side chain contacts guarantee the stability of the coiled-coil motif. Leucine residues are highlighted in red, Leu-attached protons in white.

Interestingly, signal intensities vary significantly between the different cross peaks in the HSQC spectrum in Fig. 1b. This hints towards a relaxation process that reduces the intensity of a subset of the observed NMR signals. Indeed, via relaxation dispersion experiments (see the Supporting Information) we find that MAX undergoes exchange processes with a global frequency of 702±75 Hz that can broaden NMR signals. Unfortunately, for these residues we could not achieve an NMR signal assignment as the corresponding signals were too weak in 3D assignment experiments (especially in HNCA and HNCACB experiments) hampering assignment of (^{13}C) chemical shifts. Using the HNN and HN(C)N[9] experiments this problem could partly be overcome to assign more $^{1}H^{N}$ and ^{15}N resonances.

To cast light on the properties of the intrinsically disordered (monomeric) form of MAX we employed the paramagnetic relaxation enhancement (PRE) technique. To this end, three unique

MAX cysteine mutants were introduced at positions R5C, G35C and R55C (see Fig. 1a and 3a for a graphical display of these sites within the protein construct). The cysteine residues were subsequently labeled with the paramagnetic spin label MTSL (see the Experimental section for details). Short distances between a spin labeled residue and any other amino acid of a protein reduces the amino acid's NMR signal intensity through line broadening. Corresponding signals in the ^{1}H-^{15}N 2D NMR correlation spectra are thus attenuated (cf. Fig. 1b). Signals in direct vicinity of an MTSL label are normally even broadened beyond the detection threshold. Long-range PREs, remote to a labeling site, typically stem from compacted conformations that reduce the average distance, $\langle r \rangle$, between the nucleus of interest and the unpaired electron as the effect depends very steeply on the distance by r^{-6}.[15] Contributions of elongated conformations are, thus, effectively invisible by means of PRE. Fig. 3b indicates the residue-resolved signal reduction. This signal reduction is determined via the ratio, $V = I / I_0$, with I being the signal intensity observed with the active spin label (blue in Fig. 1b), and I_0 being the signal intensity with a deactivated (chemically reduced with ascorbic acid) label (red in Fig. 1b). For the labeling site R5C located in the basic NTD of MAX one observes strong effects ($V \ll 1$) around the labeling site (short-range PRE), but only minor ($V < 1$) or no PREs for the rest of the primary sequence (long-range PRE). In stark contrast, for labeling sites G35C and R55C around 70-80 % of the NMR signals of the central part of the primary sequence are broadened beyond detection (long-range PRE; $V \ll 1$). To the best of our knowledge such a pronounced PRE effect has not been observed so far for any protein, neither for disordered nor for folded ones. The suppression of most of the NMR resonances, as observed for mutants G35C and R55C, indicates a very dense state of between aa 20 and 85 (corresponding to the H_1LH_2/LZ domain) of disordered MAX

which implies close promxity between these residues and the MTSL label. In contrast, the basic NTD remains elongated.

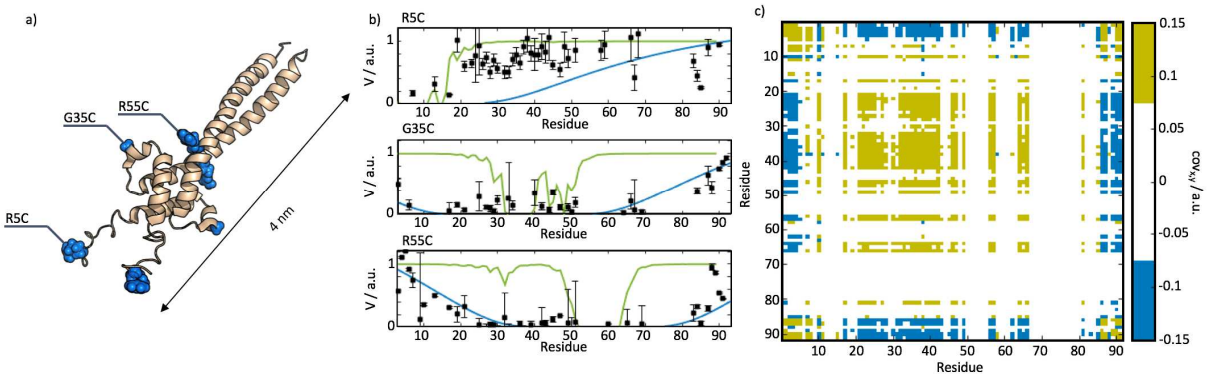


Figure 3. a) NMR solution structure of the MAX_2 homodimer. MTSL labeled residues are highlighted in blue. b) Residue plots of the intensity ratios, *V*, obtained for the three labeling sites in the disordered monomeric state of MAX (black squares). PRE ratios expected for the folded NMR solution structure of the homodimer are shown as green line. PRE ratios expected for a collapsed globule-like conformation are shown as blue line. c) Covariance matrix calculated from the PRE data in b), according to reference 28. Green matrix elements correspond to positive matrix elements, blue to negative elements.

The collapse of the MAX monomer can be rationalized when considering the physics of the hydrophobic collapse of a polymer in a poor solvent. In the framework of scaling laws for IDPs as pioneered by Schuler and co-workers[22-24] the average distance, *R*, between two residues separated by *N* residues in such a hydrophobically collapsed coil-like state is given by

$$\langle R \rangle = |\nu|^{-1/3} b^2 N^{1/3} \qquad (5)$$

, with a Flory-type scaling exponent of 1/3.[25] *v* is the excluded volume, *b* denotes the average amino acid diameter (ca. 0.23 nm). The scaling exponent of 1/3 corresponds to a state of maximal compactness of a random walk-type polypeptide chain[25] and is here, to the best of our knowledge, reported for the first time for a protein.

The blue line in Fig. 3b indicates the PRE-intensity ratios, *V*, that one would expect for the different residues of MAX in a state of full hydrophobic collapse according to eq. 5. The theoretical PRE profiles were calculated by assuming that $\langle r \rangle = \langle R \rangle$ (the average distance between the amino acid and the radical (as relevant for PRE) equals the average distance between two residues in the collapsed globule model). Additionally, we postulated an average global rotational correlation time τ_c of the vectors connecting the labeling site (position 5, 35 or 55) and the amino acids in the MAX backbone for the calculation (considering references [14-15] and equations therein; see the Methods section for details). Note that the rotational correlation time and the excluded volume are the only two variables that determine the shape of the PRE profile calculated for the collapsed globule (cf. Fig. 3b) and that they are correlated and cannot be determined independently. In other words, a similar profile could be obtained with a longer τ_c and a larger *v* or *vice versa*. For an IDP of 93 residues like MAX one may roughly anticipate rotational correlation times of 1 ns < τ_c < 4 ns.[10, 26-27] To reproduce the data in Fig. 3b for this range of τ_c the corresponding excluded volumes varies between -0.12 nm^3 and -0.15 nm^3. Assuming a rigid body spherical spectral density function, we find an average rotational correlation time of 1.6 ± 0.4 ns via TRACT (TROSY for rotational correlation times) measurements (see Fig. 4) according to reference [18] under the assumption of a hard sphere model and a single rotational correlation time. Yet, we want to stress that such an assumption and the corresponding spectral density function of the form $J(\omega) = \tau_e/(1+\omega^2\tau_e^2)$ is an oversimplification

that might yield substantial systematic errors. The value for the rotational correlation time is thus only a rough estimate of the effective average correlation time.

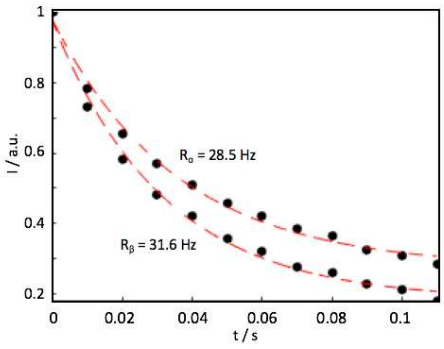


Figure 4. TRACT time traces and corresponding fits. The decay rates for the α and β spin states are indicated in the figure for MAX at 35 ˚C and 18.8 T.

Similar words of caution should be considered when calculating the theoretical PRE profiles in Fig. 3b.

For Fig. 3b we exemplarily assumed an excluded volume of -0.13 nm^3 and a τ_c of 1.6 ns. With these two parameters and the model of a collapsed globule we can reproduce the experimental PREs found for labeling sites G35C and R55C as obvious from the correspondence of the blue reproduction line and the experimental data points. However, with our polymer physical model we cannot reproduce the effect of the MTSL label at position R5C, which is placed in the hydrophilic NTD, and does not take part in the formation of the hydrophobic core, but remains solvent exposed.

Fig. 3b shows in green PREs expected for distances extracted from the NMR-derived solution structure[4] of the folded dimeric form, MAX_2. The folded solution structure resembles the crystal structure[5] of the homodimer indicating the high stability and rigidity of this complex. Strikingly,

the folded dimer shows a narrower dispersion of PREs around the labeling sites G35C and R55C than the disordered state. In other words, the unfolded state appears more compact than the folded state in our PRE measurements.

To rationalize this finding, we consider that the diameter along the main symmetry axis of MAX_2 indicated in Fig 3a is about 4 nm. For the collapsed globule model, we would expect a diameter between 1.7 nm and 2.1 nm, according to eq. 5, depending on the extend of the excluded volume of -0.12 $nm^3 < v <$ -0.15 nm^3.[25] The collapsed globule can be described by a three-dimensional random walk with maximum packing density with all water molecules being expelled from the hydrophobic core and strong intramolecular attraction. It is thus possible to obtain a structure that is even more compressed than a (classically) folded conformation as the latter might feature more void spaces, extended loops and enclosed solvent molecules. In the case of the coiled-coil MAX_2 - a structure that is asymmetrically elongated - the larger expansion in one direction of space in comparison to the compressed globular state of the monomer will lead to larger distances between the labeling sites G35C and R55C and the other amino acids in the backbone of the IDP.

Summing up, the PRE profiles displayed in Fig. 3b indicate an extended, disordered basic NTD and a densely collapsed globule for the rest of the monomeric form of MAX. This distinction between NTD and the rest of the protein is further corroborated by a covariance analysis of our PRE data. This was performed as demonstrated in detail in reference 28 and the result is shown in Fig. 3c. We compute covariance matrix elements between pairs of residues from the three PRE data sets (labeling sites) such that the resulting matrix has the dimension and order of the primary sequence of MAX. Positive matrix elements indicate pairs of residues that are correlated in the formation of the compacted state, while negative elements indicate residues that are

excluded from the compact conformation. The matrix displays a strongly correlated patch between residues 20 and 85 indicating that these residues collectively take part (i.e., they are correlated) in the formation of a compacted state that embraces the entire H_1LH_2/LZ segment. In contrast, the NTD (and the C-terminus) are anti-correlated to this region. In other words, they are not partaking in the formation of the hydrophobic core, but are expelled from this compacted region and remain solvent exposed. Considering the hydrophobic residues that form the internal surface of the coiled-coil motif in the dimer (Fig. 2), a picture arises in which these residues cluster in the core of the collapsed globule expelling all hydration water from its core, while this state is stabilized in the aqueous environment by MAX's hydrophilic residues comprised in the basic NTD and the C-terminal region of the primary sequence.

Note that the experimentally observed PRE profiles could be constituted by a single collapsed conformation or by an ensemble of structures for which we would observe an ensemble-average in our PRE measurements. *A priori,* the polymer physical description of a collapsed globule does not exclude the formation of different sub-states, instead it describes the ensemble average distance considering all the possible conformations of the polypeptide chain in a poor solvent.[25] In the latter case the abovementioned diameters between 1.7 nm and 2.1 nm should be regarded as ensemble averaged values.

Yet, to verify the observation of a highly compressed disordered state of MAX we performed PFG-DOSY (pulse field gradient - diffusion ordered spectroscopy) NMR measurements according to the method proposed by Wilkins et al.[16] The results are shown in Fig. 5 (see the Methods section for details). We find a hydrodynamic radius of the folded MAX_2 of R_h = 1.9±0.2 nm (corresponding to a diffusion coefficient, D, of $1.13 \cdot 10^6$ cm^2s^{-1} at 20 °C), which corresponds well to the extension of the elongated coiled-coil structure along its principle axis

(cf. Fig. 3a). For the unfolded state, we find R_h = 2.3±0.3 nm (corresponding to a diffusion coefficient, D, of $1.31 \cdot 10^6$ cm^2s^{-1} at 35 ˚C). The hydrodynamic radius of the unfolded conformation is thus within the error margin, similar to the radius of the folded state, indicating the presence of a compacted sub-ensemble that reduces the ensemble-averaged hydrodynamic radius of the disordered state. In contrast, one would anticipate an R_h of 2.9 nm for a completely unfolded protein (e.g., in a solution of 8 M urea) of the length of MAX according to empirical relations as described in reference 16 highlighting the here reported compactness of the conformational ensemble of MAX.

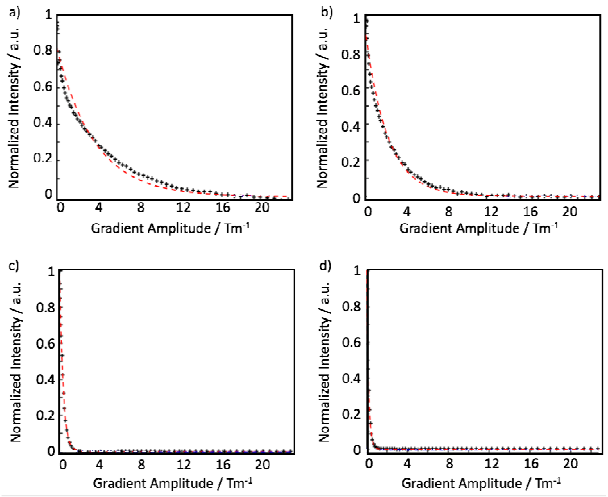


Figure 5. a) DOSY decay trace for MAX_2 at 20 ˚C. The red solid line was fitted to the data as a monoexponential function. b) DOSY decay trace for 2,4-dixoane at 20 ˚C. c) DOSY decay trace for MAX at 35 ˚C. d) DOSY decay trace for 2,4-dixoane at 35 ˚C.

Comparing the hydrodynamic radii found for the folded state (1.9±0.2 nm), the disordered ensemble (2.3±0.3 nm) and anticipated for the entirely unfolded state (2.9 nm)[16] we deduce that the intrinsically disordered state of MAX cannot be described by an ideal random coil picture as it is on average more compact than the hypothetically completely unfolded state. From

theoretical considerations one would expect a Flory scaling exponent of 0.588 for this case with aqueous buffers considered as good solvents for urea denatured proteins. This expansion would entail an increased R_h, which is reflected in a larger average end-to-end distance in the polymer physical description. Indeed, one would find $\langle R \rangle = b(v/b^3)^{0.18}N^{0.588}$ to describe an unstructured coil in a good solvent, i.e., with stabilizing interactions between the solvent and polypeptide (cf. eq. 5, where we find $N^{1/3}$ to describe the reduced coil expansion).[25] Yet, the high degree of compactness of the collapsed globule must be considered to understand the smaller hydrodynamic radius of the intrinsically disordered state as the buffer becomes poor solvent for the hydrophobic amino acids of MAX in the absence of a denaturation agent such that the intramolecular attractions become stronger than the stabilizing interaction with the solvent. These findings are in line with the recent insights that IDPs cannot simply be considered as unfolded polypeptide chains, but instead sample heterogeneous conformational spaces that contain conformations that can deviate from the simple picture of random coils.[29] To this picture we add here another state of possible conformations of IDPs, which comprises an unusually high compactness despite the absence of any secondary or tertiary structure motifs.

Note that we carefully excluded intermolecular interactions that can bias the PRE measurements by performing experiments on a mixture of ^{15}N-labeled (but not spin-labeled) wildtype (wt) MAX and ^{14}N and spin-labeled MAX (mutants R5C, G35C and R55C; see the Supporting Information). In none of these experiments we detected any significant intermolecular PRE under our experimental conditions. Additionally, solvent PREs due to the presence of paramagnetic salts or free MTSL present in the solution have been excluded.

These results are shown in the Supporting Information. No significant intermolecular PRE can be detected. Likewise, no PRE effect was detected in the presence of 1 mM Mn^{2+}·EDTA or in the presence of 1 mM free MTSL in the solution.

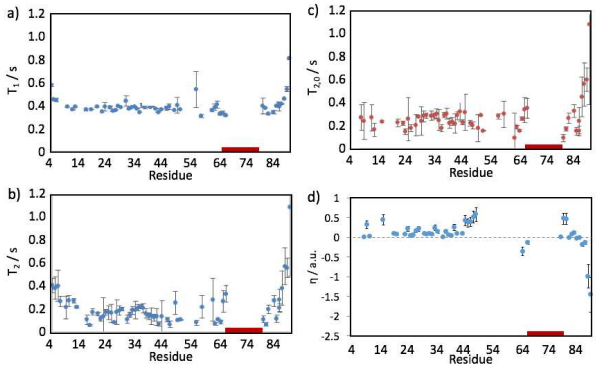


Figure 6. a) Residue-resolved ^{15}N-T_1 values for the monomeric form of MAX at 35˚C and 18.8 T (800 MHz). The red bar indicates a patch of residues that could not be assigned due to weak signal intensities. b) Residue-resolved ^{15}N-T_2 values for the monomeric form of MAX at 35˚C and 18.8 T (800 MHz). c) T_2 values detected with a CPMG field to suppress the exchange contribution (denoted $T_{2,0}$). d) Heteronuclear ^{15}N-{$^1H^N$} Overhauser enhancement (NOE, η) as a function of residue position obtained at 35˚C and a magnetic field of 14.1 T (600 MHz). For some residues, η could not unambiguously determined due to the lower field and signal overlap.

To further characterize the conformational sampling of monomeric MAX, we determined heteronuclear longitudinal ^{15}N-T_1 and transverse ^{15}N-T_2 relaxation times as well as transverse ^{15}N-$T_{2,0}$ times, i.e., ^{15}N-T_2 in the presence of a CPMG field that suppresses exchange

contributions to transverse relaxation (see Fig. 6 a-c). These measurements were performed at 18.8 T magnetic field strength. Displaying a rather flat profile, the ^{15}N-T_1 data shown in Fig. 6a do not indicate significant variations in local backbone flexibility in the central domain between aa 20 and 85. Structural elements would cause variations in local backbone dynamics, which in return would cause significant fluctuations in the ^{15}N-T_1 values in dependence of the residue position. This is not observed here. Only the termini show elevated values due to their intrinsically higher mobility. This is in accordance with the above-deduced picture of a collapsed random-coil structure between aa 20 and 85 void of any significant structural elements. Under our experimental conditions, a short relaxation time generally indicates slow backbone dynamics, while a long relaxation indicates the opposite. The average ^{15}N-T_1 value of 0.40 s is quite low for an IDP[10] thus suggesting a relatively low conformational flexibility and slow local tumbling (on the nanosecond timescale) of residues in the disordered state of MAX, again corroborating the idea of a high level of compression due to locally reduced flexibility in the densely collapsed globule. For comparison, we can anticipate for a fully extended protein, e.g., denatured with 8 M urea, of the same length as MAX a ^{15}N-T_1 on the order of 1 s [7, 10] due to the faster internal dynamics in this hypothetical extended polypeptide. Surprisingly, for the dimer transverse relaxation times of ^{15}N-T_2 > 1.5 s (at 14.1 T) have been reported, which are astonishingly long.[5] However, the authors of this study suggest that these values are to be analyzed with caution due to the strongly anisotropic rotation of MAX_2, which might introduce significant biases. Yet, the short ^{15}N-T_1 for the intrinsically disordered state in comparison to the ^{15}N-T_1 of the dimer corroborates the idea of reduced flexibility and, hence, compactness of the monomeric state.

In Fig. 6b ^{15}N-T_2 values are shown. As for ^{15}N-T_1, it can be observed that ^{15}N-T_2 values are on average rather short for an IDP, especially between aa 20 and 85 spanning the H_1LH_2/LZ domain where we find an average ^{15}N-T_2 of 0.15 s. However, while ^{15}N-T_1 is not influenced by any exchange contributions, ^{15}N-T_2 values are. Consequently, to separate the intrinsic relaxation from the exchange contribution we detected ^{15}N-$T_{2,0}$ values in the presence of a 1 kHz CPMG field, which suppresses the exchange contribution. Fig. 6c displays the ^{15}N-$T_{2,0}$ data showing that ^{15}N-$T_{2,0}$ amounts on average to 0.24 s in the region spanning the H_1LH_2/LZ domain (aa 20-85), which is significantly longer than ^{15}N-T_2. The conformational exchange observed via CPMG RD thus appears to primarily affect the H_1LH_2/LZ motif in accordance with the observation of strong relaxation dispersion in this area (see the Experimental section). The H_1LH_2/LZ region folds upon homodimerization and constitutes the anchor between the two monomer units and thus enables the formation of the coiled-coil motif. Such, the conformational exchange between folded and unfolded state shortens the effective transverse relaxation times leading to the observed CPMG profiles. Note that for both, ^{15}N-$T_{2,0}$ and ^{15}N-T_2, the residue-resolved profiles are quite flat between aa 20 and 85 indicating like ^{15}N-T_1 the absence of significant local structures in the collapsed globule.

Furthermore, it should be noted that an average ^{15}N-$T_{2,0}$ of 0.24 s is again at the lower edge of the spectrum of possible transverse relaxation times expected for an IDP. For an IDP of the size of MAX one would expect values between 0.25 and 0.5 s,[7, 26, 30-31] such that a value of 0.24 s indicates a high degree of compactness and a consequent restriction of local dynamics that reduces relaxation times. For the dimer an average ^{15}N-T_2 of only ~0.04 s has been reported.[5] This value is very short in comparison to the reported ^{15}N-T_1 of > 1.5 s due to the strongly anisotropic rotation and rigidity of the dimer and therefore needs to be analyzed with caution.

It should be noted that from the average value of ^{15}N-$T_{2,0}$/^{15}N-T_1 an average rotational correlation time of 1.7 ns can be estimated via the relation $\tau_c = (6^{15}\text{N-}T_{2,0}/^{15}\text{N-}T_1 - 7)^{1/2}/4\pi\nu(^{15}\text{N})$ (see reference [32]) under the assumption of a simple spherical rotation for the spectral density function, while we found 1.6 ns via our TRACT experiments using the same model (see the Experimental section) indicating the consistency of the NMR data. However, from the hydrodynamic radius obtained via PFG-DOSY for the intrinsically disordered state a significantly larger value of 8.3±2.6 ns can be determined using the hydrodynamic relation for hard spheres $\tau_c = 4\pi\eta R_h^3/3kT$. This discrepancy between the value obtained from PFG-DOSY or from relaxation times indicates that a simple spherical model is not suited for a precise description (neither hydrodynamic nor spectroscopic) of the dynamics of IDPs. Instead the hard sphere model serves here only as a simplest tool to obtain a rough estimate of correlation times in the low ns range. At present, no model exists that accurately considers all aspects of IDP dynamics and the deduction of more complex models for the description of MAX is beyond the scope of this work. Currently, we are working on a detailed analysis of the relaxation data with more sophisticated dynamics models

Finally, we determined heteronuclear ^{15}N-{^{1}H^N} Overhauser enhancements (NOE, η) as a function of residue position (see Fig. 6d). η is indicative for fast backbone vibrations on the picosecond time scale. Such vibrations have been observed to be important for substrate interactions in IDPs.[7] In a simplified picture, positive values can be interpreted as slow fluctuations on this time scale, while negative values indicate the opposite. We find positive values with $0 < \eta < 0.5$ for the largest part of the backbone spanning the entire b/H_1LH_2/LZ domain indicating the absence of fast vibrations on the picoseconds time scale in this region. Only a few values with $\eta > 0.5$ could be observed around position 50. Negative NOEs could be

observed at the C-terminus due to its intrinsically high motional freedom. The overall flat NOE profile in Fig. 6d implies homogeneous internal dynamics of the collapsed globule of the MAX monomer, which is in good agreement with the flat $^{15}N\text{-}T_{2,0}$ and $^{15}N\text{-}T_1$ profiles that are representative for ns-timescale motions. Note that the polymer physical model of a collapsed globule likewise assumes the absence of any structural element and thus predicts a homogeneous, i.e., a self-similar (fractal) internal configuration of the polypeptide chain.

This match between our NMR data and the simple polymer physical model is interesting in regard of the frequent observation that IDP dynamics take place on a range of different timescales and cannot be described by simple models (as evident from the PFG-DOSY data). Instead, the assumption of a distribution of different correlation times for the different time regimes appears to be necessary to describe IDP dynamics.[26-27] For the case of MAX, the homogeneous η, $^{15}N\text{-}T_{2,0}$ and $^{15}N\text{-}T_1$ profiles in Fig. 6 suggest that the distribution of correlation times is quite constant for the different positions along the polypeptide chain since a variation of the latter would cause residue position-dependent variations for at least one relaxation parameter.

CONCLUSION

Summing up, the PRE profiles in Fig. 3 supplemented with DOSY as well as T_1, T_2 and NOE data indicate the presence of a strongly compacted state spanning the H_1LH_2/LZ domain of the intrinsically disordered state of MAX. The formation of this strongly compacted structure might be involved in the conformational transition between the folded and disordered states of MAX since it spans the H_1LH_2/LZ domain which constitutes the central structural motif of the MAX_2 dimer and the anchor point between the two unimer units. It has been reported along these lines

that strongly compacted states of IDPs can either serve to aid subsequent folding by reducing the loss in configurational entropy upon folding[33] or to prevent the formation of secondary/tertiary structures by increasing the internal friction within the polypeptide chain.[23]

Such, folded globular proteins cannot unhesitatingly be considered as the compact counterpart of IDPs. The here observed compact conformation of the monomeric form of the Myc-associated factor X is, to the best of our knowledge, the most compact IDP conformation with a degree of compression comparable to folded proteins. This hints towards the importance of compacted, yet unfolded structures in heterogeneous conformational spaces of IDPs.

ASSOCIATED CONTENT

Supporting Information. CPMG RD Data, NMR Resonance Assignment

AUTHOR INFORMATION

Corresponding Author

DK, email: kurzbach@ens.fr phone: +33 1 44323265

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Funding Sources

DK acknowledges a Feodor Lynen scholarship by the Alexander von Humboldt-Foundation. This work was supported by the Austrian Science Foundation (FWF) (W1221 and I844) and by the European Research Council (ERC contract 'dilute para water').

ACKNOWLEDGMENT

The authors want to thank Prof. Robert Konrat and Prof. Geoffrey Bodenhausen for granting access to their laboratories.

REFERENCES

1. Blackwood, E. M.; Eisenman, R. N., Max - a Helix-Loop-Helix Zipper Protein That Forms a Sequence-Specific DNA-Binding Complex with Myc. *Science* **1991,** *251*, 1211-1217.

2. Littlewood, T. D.; Amati, B.; Land, H.; Evan, G. I., Max and C-Myc Max DNA-Binding Activities in Cell-Extracts. *Oncogene* **1992,** *7*, 1783-1792.

3. Conacci-Sorrell, M.; McFerrin, L.; Eisenman, R. N., An Overview of MYC and Its Interactome. *Csh. Perspect. Med.* **2014,** *4*, 1-24.

4. Sauve, S.; Tremblay, L.; Lavigne, P., The NMR solution structure of a mutant of the max b/HLH/LZ free of DNA: Insights into the specific and reversible DNA binding mechanism of dimeric transcription factors. *J. Mol. Biol.* **2004,** *342*, 813-832.

5. Brownlie, P.; Ceska, T. A.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D., The crystal structure of an intact human Max-DNA complex: New insights into mechanisms of transcriptional control. *Structure* **1997,** *5*, 509-520.

6. Fieber, W.; Schneider, M. L.; Matt, T.; Krautler, B.; Konrat, R.; Bister, K., Structure, function, and dynamics of the dimerization and DNA-binding domain of oncogenic transcription factor v-Myc. *J. Mol. Biol.* **2001,** *307*, 1395-1410.

7. Kurzbach, D.; Schwarz, T. C.; Platzer, G.; Höfler, S.; Hinderberger, D.; Konrat, R., Compensatory Adaptations of Structural Dynamics in an Intrinsically Disordered Protein Complex. *Angew. Chem. Int. Ed.* **2014,** *53*, 3840-3843.

8. Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J., Intrinsically disordered proteins: regulation and disease. *Curr. Opinion Struct. Biol.* **2011,** *21*, 432-440.

9. Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V., Improved 3D triple resonance experiments, HNN and HN(C)N, for H-N and N-15 sequential correlations in (C-13, N-15) labeled proteins: Application to unfolded proteins. *J. Biomol. NMR* **2001,** *20* (2), 135-147.

10. Rule, G. S.; Hitchens, T. K., *Fundamentals of Protein NMR Spectroscopy*. Springer: Dordrecht, 2006.

11. Korzhnev, D. M.; Kay, L. E., Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Accounts Chem. Res.* **2008,** *41*, 442-451.

12. Carver, J. P.; Richards, E. R., A General Two-Site Solution for the Chemical Exchange Produced Dependence of T2 Upon Carr-Purcell Pulse Separation. *J. Magn. Reson.* **1972,** *6*, 89-105.

13. Luz, Z.; Meiboom, S., Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution-Order of the Reaction with Respect to Solvent. *J. Chem. Phys.* **1963,** *39*, 366-370.

14. Battiste, J. L.; Wagner, G., Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **2000,** *39*, 5355-5365.

15. Clore, G. M.; Iwahara, J., Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **2009,** *109*, 4108-4139.

16. Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999,** *38*, 16424-16431.

17. Nilsson, M., The DOSY Toolbox: a new tool for processing PFG NMR diffusion data. *J. Magn. Reson.* **2009,** *200*, 296-302.

18. Lee, D.; Hilty, C.; Wider, G.; Wuthrich, K., Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* **2006,** *178*, 72-76.

19. Kurzbach, D.; Platzer, G.; Schwarz, T. C.; Henen, M. A.; Konrat, R.; Hinderberger, D., Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **2013,** *52*, 5167-5175.

20. Platzer, G.; Schedlbauer, A.; Chemelli, A.; Ozdowy, P.; Coudevylle, N.; Auer, R.; Kontaxis, G.; Hartl, M.; Miles, A. J.; Wallace, B. A.; Glatter, O.; Bister, K.; Konrat, R., The Metastasis-Associated Extracellular Matrix Protein Osteopontin Forms Transient Structure in Ligand Interaction Sites. *Biochemistry* **2011,** *50*, 6113-6124.

21. Geist, L.; Henen, M. A.; Haiderer, S.; Schwarz, T. C.; Kurzbach, D.; Zawadzka-Kazimierczuk, A.; Saxena, S.; Zerko, S.; Kozminski, W.; Hinderberger, D.; Konrat, R., Protonation-dependent conformational variability of intrinsically disordered proteins. *Protein Science* **2013,** *22*, 1196-1205.

22. Nettels, D.; Muller-Spath, S.; Kuster, F.; Hofmann, H.; Haenni, D.; Ruegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B., Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Nat. Acad. Sci.* **2009,** *106*, 20740-20745.

23. Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Muller-Spath, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B., Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Nat. Acad. Sci.* **2012,** *109*, 17800-17806.

24. Aznauryan, M.; Nettels, D.; Holla, A.; Hofmann, H.; Schuler, B., Single-Molecule Spectroscopy of Cold Denaturation and the Temperature-Induced Collapse of Unfolded Proteins. *J. Am. Chem. Soc.* **2013,** *135*, 14040-14043.

25. Rubinstein, M.; Colby, R., *Polymer Physics*. Oxford University Press: 2003.

26. Khan, S. N.; Charlier, C.; Augustyniak, R.; Salvi, N.; Dejean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F., Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys. J.* **2015,** *109,* 988-999.

27. Charlier, C.; Khan, S. N.; Marquardsen, T.; Pelupessy, P.; Reiss, V.; Sakellariou, D.; Bodenhausen, G.; Engelke, F.; Ferrage, F., Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry. *J, Am. Chem. Soc.* **2013,** *135*, 18665-18672.

28. Kurzbach, D.; Beier, A.; Vanas, A.; Flamm, A. G.; Platzer, G.; Schwarz, T. C.; Konrat, R., NMR probing and visualization of correlated structural fluctuations in intrinsically disordered proteins. *Phys, Chem, Chem, Phys,* **2017,** *19*, 10651-10656.

29. Konrat, R., The meandering of disordered proteins in conformational space. *Structure* **2010,** *18,* 416-419.

30. Salvi, N.; Abyzov, A.; Blackledge, M., Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *J. Phys. Chem. Lett.* **2016,** *7* (13), 2483-2489.

31. De Biasio, A.; de Opakua, A. I.; Cordeiro, T. N.; Villate, M.; Merino, N.; Sibille, N.; Lelli, M.; Diercks, T.; Bernado, P.; Blanco, F. J., p15(PAF) Is an Intrinsically Disordered Protein with Nonrandom Structural Preferences at Sites of Interaction with Other Proteins. *Biophys. J.* **2014,** *106*, 865-874.

32. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **1989,** *28*, 8972-8979.

33. Cheung, J. K.; Shah, P.; Truskett, T. M., Heteropolymer collapse theory for protein folding in the pressure-temperature plane. *Biophys. J.* **2006,** *91*, 2427-2435.

For Table of Contents Use Only

