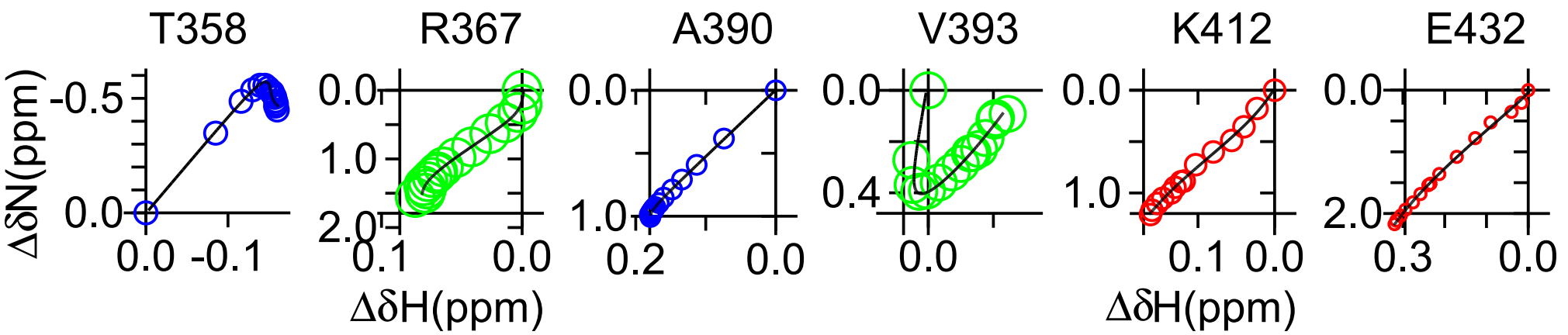
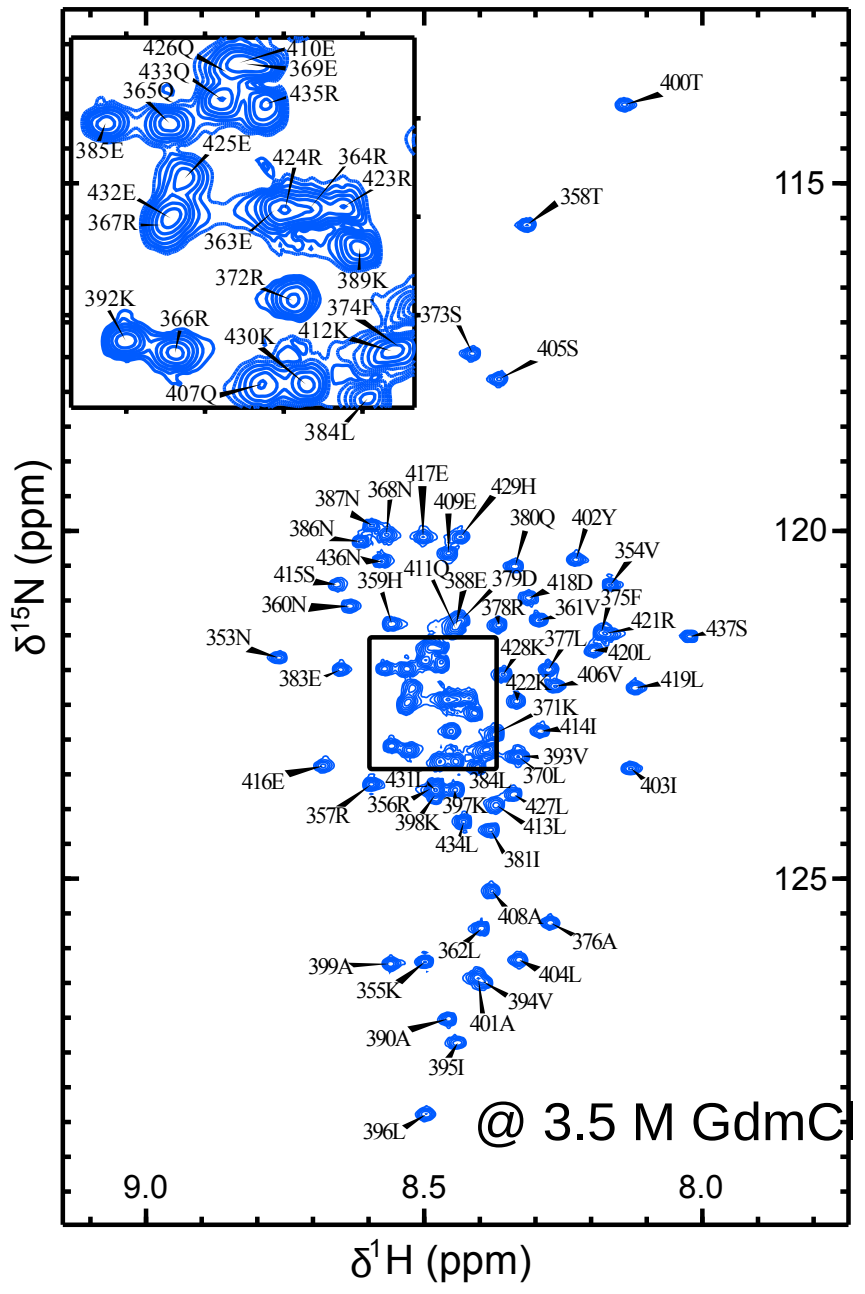

Article:
Panova, S., Cliff, M.J., Macek, P. et al. (6 more authors) (2019) Mapping hidden residual structure within the Myc bHLH-LZ domain using chemical denaturant titration. Structure, 27 (10). 1537-1546.e4. ISSN 0969-2126

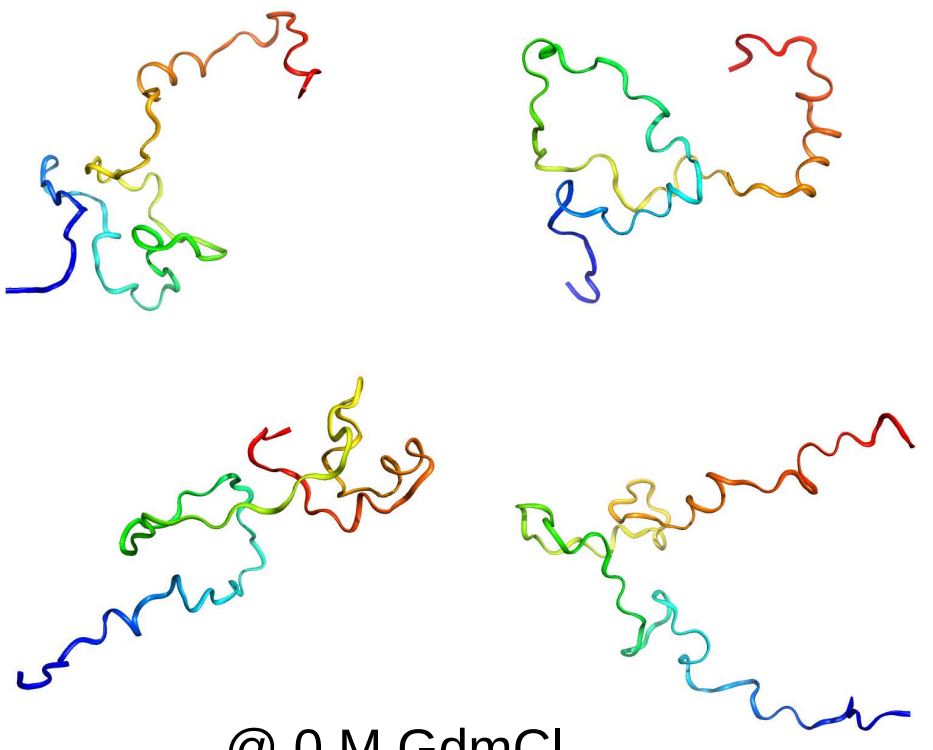
https://doi.org/10.1016/j.str.2019.07.006

Graphical Abstract



Chemical denaturant titration





Title:

Mapping hidden residual structure within the Myc bHLH-LZ domain using chemical denaturant titration

Authors: Stanislava Panova[1], Matthew J. Cliff[1], Pavel Macek[2,3], Martin Blackledge[4], Malene Ringkjøbing Jensen[4], J. Willem M. Nissink[5], Kevin J. Embrey[2], Rick Davies[2], & Jonathan P. Waltho[1,6,7*]

Author affiliations:

[1] Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK.

[2] Discovery Sciences, IMED Biotech Unit, AstraZeneca, Alderley Park, SK10 4TG, UK.

[3] NMR-Bio, Institut de Biologie Structurale, Grenoble Cedex 9, France

[4] Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

[5] Oncology, IMED Biotech Unit, AstraZeneca, Cambridge, CB2 0AA, UK.

[6] Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK

[7] Lead contact

*Correspondence: j.waltho@manchester.ac.uk

Summary (150 words)

Intrinsically disordered proteins (IDPs) underpin biological regulation and hence are highly desirable drug-development targets. NMR is normally the tool of choice for studying the conformational preferences of IDPs but the association of regions with residual structure into partially collapsed states can lead to poor spectral quality. The bHLH-LZ domain of the oncoprotein Myc is an archetypal example of such behaviour. To circumvent spectral limitations, we apply chemical denaturant titration (CDT) NMR, which exploits the predictable manner in which chemical denaturants disrupt residual structure and the rapid exchange between conformers in IDP ensembles. The secondary structure propensities and tertiary interactions of Myc are determined for all bHLH-LZ residues, including those with poor NMR properties under native conditions. This reveals conformations that are not predictable using existing crystal structures. The CDT-NMR method also maps sites perturbed by the prototype Myc inhibitor, 10058-F4, to areas of residual structure.

Keywords: Intrinsically Disordered Proteins, Myc, Guanidinium Chloride, Molten globule, Solution NMR, paramagnetic relaxation enhancement.

Introduction

Intrinsically disordered regions of proteins or entire intrinsically disordered proteins (IDPs) are extremely prevalent in higher eukaryotes, and are involved in a wide range of biologically important processes, such as extracellular communication, intracellular signalling, DNA replication and transcription (Babu et al., 2011; Habchi et al., 2014; Oldfield and Dunker, 2014; Wright and Dyson, 2015). The conformational ensemble defined as disordered does not necessarily represent a purely random coil state, and different IDP sequences exhibit different behaviours, ranging from rapidly-rearranging disordered coils, to more collapsed states with long-range contacts and persistent secondary structure elements. These latter IDPs can be classified as having molten globule -like behaviour, characterised by a loose core without the precise packing of folded proteins (van der Lee et al., 2014). Molten globule -like behaviour is observed in a range of states from native-like folds with dynamic interiors (e.g. *apo*-myoglobin), to ensembles with low hydrodynamic radii but no persistent structure. IDPs that display this behaviour are frequently involved in molecular recognition and some adopt conventional globular structures when in complex with a binding partner. The basic-helix-loop-helix-leucine zipper (bHLH-LZ) domain of the oncoprotein Myc is an archetypal example of such behaviour.

Myc is an important transcription factor for cell growth, metabolism and apoptosis, and its overexpression is associated with many cancers (Santarius et al., 2010). Transcriptional activity of Myc requires hetero-dimerization with the protein Max. Both proteins have disordered bHLH-LZ domains, which become ordered upon heterodimer formation, leading to the recognition and binding of the E-box DNA sequence (Blackwood and Eisenman, 1991; Nair and Burley, 2003; Prendergast et al., 1991). This makes small molecule inhibitors that disrupt the Myc-Max interaction attractive candidates to be used as anticancer agents (Follis et al.,

2009; Metallo, 2010), but their development is limited by a lack of appropriate characterisation of Myc in the IDP state adopted when isolated from Max.

Conventional structural biology techniques are not well-suited for IDPs, because any tertiary structure is transient. NMR spectroscopy provides a number of informative measurements, including chemical shifts, residual dipolar couplings and paramagnetic relaxation enhancement (PRE) (Bhowmick et al., 2017; Jensen et al., 2014; Sormanni et al., 2017). PREs report on interactions up to 30 Å, and on rarely populated states (<5%) (Baldwin and Kay, 2009; Clore, 2013; Salmon et al., 2010). However, IDPs present challenges to standard NMR techniques; low structural complexity results in poor signal dispersion, although this is ameliorated by the intense resonances observed for fully denatured proteins, and exposed amide groups are subject to signal loss through solvent exchange, although this can sometimes be ameliorated using ^{13}C-detection (Bermel et al., 2012; Goradia et al., 2015; Wiedemann et al., 2015). IDPs that have molten globule -like behaviour are more problematic, because the underlying conformational exchange typically occurs on timescales that result in severely attenuated NMR resonances. The signal attenuation is proposed to arise from averaging between many conformers with a large range of barrier heights defined by a rough protein conformation energy landscape (Milanesi et al., 2012).

Here, the structure propensity of the isolated bHLH-LZ domain from Myc has been extensively characterised using chemical denaturant titration NMR (CDT-NMR). Under native-like conditions, the molten globule -like behaviour of Myc results in residues of the leucine zipper region producing no detectable resonances. We resolve this problem by shifting the solution equilibrium towards a monomeric, less collapsed state (McParland et al., 2002; Reed et al., 2006); titration with increasing guanidinium chloride (GdmCl) induces a cooperative transition of Myc to a more disordered state. The GdmCl dependence is used to extrapolate chemical shifts back to native conditions and to analyse the PREs of three singly MTSL-labelled cysteine

variants. The data reveal considerable helical structure under native conditions, especially in part of the leucine zipper region. There is also significant tertiary contact between residues in the helix 2-leucine zipper boundary region and those of helix 1 that is quite different to what is observed in the Myc-Max crystal structure (Nair and Burley, 2003). The CDT-NMR approach also allows the interaction with the prototype Myc inhibitor, 10058-F4, to be identified as specifically affecting this tertiary contact in the molten globule -like state of Myc.

Results

Assignment of bHLH-LZ domain NMR spectra.

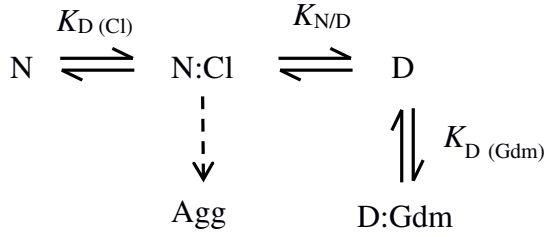
The ^{15}N-^{1}H HSQC spectra of the bHLH-LZ domain of Myc (residues 352-437) in native-like conditions (20 mM phosphate, 0.1 M NaCl, pH 7.4, 298K) shows only 46 out of 83 possible cross-peaks (**Fig. 1a**). The detected peaks are broad, while retaining the poor dispersion expected of a fully disordered protein. Therefore, residual order is slowing motion to a timescale that causes the intensity of some resonances to be completely attenuated by NMR relaxation. Changes in NaCl concentration, pH and acquisition temperature failed to increase the number of detected peaks (described in Methods). In contrast, introduction of chemical denaturant substantially improved the spectra. All 83 expected backbone amide signals are observable at 0.6 M GdmCl and above, and all resonances have narrow linewidths (<14 Hz for ^{1}H) at 3.2 M GdmCl (**Fig. 1b**).

Residue-specific assignment of all backbone amide resonances was completed at 3.2 M GdmCl using conventional triple resonance experiments, and an additional (H)N(CA)NNH experiment for regions where the C_α dispersion is poor. The $^{1}H_N$ and $^{15}N_H$ chemical shifts were observed to change continuously in a titration from 3.2 to 0 M GdmCl, allowing the assignment to be transferred to other GdmCl concentrations. HNCACB spectra were recorded at 2.4, 1.6 and 0 M of GdmCl to verify the assignment at these GdmCl concentrations. The resonances that are missing in the absence of GdmCl correspond to the C-terminal residues of Myc (400-437), and

constitute part of the second helix (H2) and the whole leucine-zipper (LZ) region in the Myc-Max crystal structure (secondary structure shown in **Fig**. **2a**).

Three categories of response to denaturant

The responses of the NMR resonances to GdmCl were categorized into three distinct behaviour types (**Fig. 2a & b**). For some residues (Blue category **Fig**. **2a & b**), the largest chemical shift changes occur at low GdmCl concentration (<0.6 M) where both $\delta^{1}H_{N}$ and $\delta^{15}N_{H}$ have the same hyperbolic relationship with GdmCl concentration. These transitions correspond well with the behaviour when GdmCl is replaced by NaCl in the titrations (**Fig. S1**). Hence, the effect is independent of the cation present and therefore is ascribed to interaction with Cl^{-} ions. The simplest model of a binding interaction is adequate to describe the data but the interaction with Cl^{-} may not be a direct binding event, and may also involve formation of Myc homodimers (Blackwood and Eisenman, 1991). For a second group of residues (Red category **Fig. 2a & b**), the $\delta^{1}H_{N}$ and $\delta^{15}N_{H}$ changes have a sigmoidal dependence on GdmCl concentration with a mid-point at ~1.2 M. There is no equivalent transition with NaCl and so the effect is ascribed to the Gdm^{+} ions and is indicative of a folding event with some cooperativity. A third group of residues (Green category **Fig. 2a & b**) are subject to both effects and, in contrast to the other categories, these ^{1}H-^{15}N crosspeaks follow a significantly curved path as the denaturant concentration changes. Additionally, all residues show a weak linear dependence on GdmCl concentration, which persists at high denaturant concentrations (>2.5M). This likely reflects a weak interaction between Gdm^{+} and the protein backbone (Plaxco et al., 1997), analogous to the well-established effect of urea (Huang et al., 2012; Meier et al., 2007). The total reaction scheme for the interaction between Myc and GdmCl is described by *Scheme 1*:



Scheme 1

where N stands for native disordered state of Myc, D for denatured and N:Cl and D:Gdm indicate the bound forms. The dotted arrow represents a very slow precipitation of Myc in the presence of elevated Cl^- concentrations, which is countered by Gdm^+ ions.

The parameters defining the Cl^- interaction and the cooperative folding event were fitted globally to *Equation 1* using δ^1H_N and $\delta^{15}N_H$ values for a subset of residues with large chemical shift changes, and then fixed when fitting the chemical shift changes for all residues. All residues are fitted satisfactorily with global parameters for the underlying physical processes (examples in **Fig. 2b**). Fitting the chemical shift changes using the global parameters allows the extrapolation of the incomplete datasets to 0 M GdmCl, thereby providing estimates of δ^1H_N and $\delta^{15}N_H$ values where none could be determined experimentally. The accuracy of these estimates was determined to be 0.03 ppm for δ^1H_N and 0.3 ppm for $\delta^{15}N_H$ (see Methods). Using $\Delta\delta^1H_N$ and $\delta^{15}N_H$ values in the global fitting procedure rather than the standard chemical shift perturbation formula ($(\Delta\delta^1H_N^2 + (\Delta\delta^{15}N_H/6)^2)^{0.5}$) produced more consistent fit parameter values, particularly for residues in the green category (**Fig. 2b)**, because the root - sum of squares function does not account for the direction of chemical shift changes.

The sequence distribution of the resulting chemical shift changes corresponding to each transition is shown in **Fig. S2**. The derived K_D value for the interaction with Cl^- (410 mM) is consistent with weak, non-specific binding, which is focussed at hotspots around the KRR sequence (residues 355-357) and around residue A390, which in the Myc-Max crystal structure correspond to the N-terminal basic H1 region and the turn region, respectively. The signal

attenuation observed at low denaturant concentration precluded analysis of the Cl^- interaction for much of the C-terminal region. The best fit parameters for the cooperative folding transition indicate weak stability (–6.2 $kJ.mol^{-1}$) and a low Gdm^+ *m*-value of 2.8 (equivalent to 6.5 $kJ.mol^{-1}.M^{-1}$(Clarke and Waltho, 1997; Myers et al., 1995)) for the more folded species; values typical of partially ordered states (Cliff et al., 2009; Reed et al., 2006; Scholtz et al., 2009). The final two heptad repeats of the LZ (residues 426-436) in the Myc-Max crystal structure show the largest chemical shift changes for this transition, and are in the region with large signal attenuations in native-like conditions. The gradients of the weak linear dependence of chemical shift on GdmCl concentration, visible at high denaturant concentrations, are also largest for the C-terminal residues.

α-Helical structure is populated at low denaturant concentrations

The changes in chemical shift with denaturant suggest that the conformational distribution of the polypeptide chain is changing. Whilst some of these changes can be ascribed to weak interactions with Cl^- or Gdm^+, the cooperative transition is consistent with structural changes that alter the exposure of hydrophobic surface area (Scholtz et al., 2009). The backbone ^1H, ^{15}N and ^{13}C chemical shifts were analysed to determine whether these changes corresponded to a change in secondary structure propensity (SSP). Full data sets (δ^1H_N, $\delta^{15}N_H$, $\delta^{13}C\alpha$, $\delta^{13}C\beta$) were available at 1.6, 2.4 and 3.2 M GdmCl, with additional data at 0 M for residues 352-395. The $\Delta\delta^{13}C$ values have a strong correlation with $\Delta\delta^{15}N_H$ values for the same residue (**Fig. S3**) and so $\delta^{13}C\alpha$ and $\delta^{13}C\beta$ were extrapolated back to 0 M GdmCl using the fitted $\Delta\delta^{15}N_H$ values from the analysis above, which allowed estimation of SSPs for residues 400-437. The accuracy of the extrapolations was determined to be 0.23 ppm for both $\delta^{13}C\alpha$ and $\delta^{13}C\beta$ (see Methods).

Two helical clusters can be distinguished (residues 359-373 and 400-436), which dissolve upon addition of GdmCl (**Fig. 3**). In particular, the region from residues 416-422 (SEEDLLR) is predicted to be 90% helical in the absence of denaturant. These clusters are also helical in the

Myc-Max crystal structure. However, the reverse is not always true; for example the residues of the first turn of H2 in the Myc-Max crystal structure (393-396, VVIL) have a significant strand or PPII-helix propensity in isolated Myc at 0 M GdmCl. The chemical shifts for several residues do not reach random coil values at 3.2 M GdmCl, with up to 30% helical content remaining in the LZ region. It is notable that the region of greatest helix content determined from $\delta^{13}C$ values does not co-locate to the region of greatest $\delta^{15}N_H$ perturbation by GdmCl (**Fig. S2**), which lies between residue Q426 and the C-terminus.

Paramagnetic relaxation enhancements detect a tertiary contact at low denaturant

The secondary structure information present in chemical shifts is complemented by tertiary contact information from paramagnetic relaxation enhancement (PRE) measurements. A number of single cysteine variants were screened for optimal protein expression, and three were prepared and labelled using the nitroxide spin-label MTSL, namely Q365C-SL, N386C-SL and S405C-SL. The denaturant titration profiles for the derivatised variants are very similar to the equivalent profiles for wild-type Myc (**Fig. S4**), with crosspeaks being readily assigned by direct comparison of spectra. In the absence of denaturant, no data are available for residues 400-437 because of the signal attenuation described above. In order to gain tertiary contact information from the line-broadened peaks, PREs were measured as a function of denaturant concentration, and *Equation 2* was used to extrapolate PRE values for residues with attenuated signals at 0 M GdmCl, using the parameters defined by the analysis of chemical shift changes (Cliff et al., 2009) (see **Fig. 4** for example profiles).

For all three spin-labelled variants, there is a strong denaturant dependence for the measured PREs. At 0 M GdmCl (**Fig. 5a**), the sequence distribution of the resulting intensity ratios (I_{para}/I_{dia}) is broad. The Q365C-SL variant reports significant contacts throughout the 355-390 region, with lower effects up to residue Q407, but little contact with the C-terminal region. The N386C-SL variant reports contacts throughout the same region, with the primary effect

occurring between residues Q365 and A399. In contrast, the S405C-SL variant reports contacts over much more of the sequence, with the primary effects broadly centred around residue F375 and, to a lesser extent, around residue E425. Effects observed for residues before Q365 and between E385 and A390 are lower, indicating there is a preferential contact between S405 and the F375 region compared with the intervening residues. At 3.2 M GdmCl (**Fig. 5b**), the PREs from each spin-labelled variant largely follow the behaviour expected for a fully disordered protein, but with some low PREs (i.e. intensity ratios less than 2 standard deviations below that expected for a random coil) at sequence distant positions. The Q365C-SL variant reports contacts extending to K355 and Q407, while the N386C-SL variant reports contacts to the region between residues Q365 and R378. The S405C-SL variant reports contacts to the region between residues E363 and E385. Hence, the simplest model is that the observed PREs at this GdmCl concentration reflect the rare population of species that resemble the native disordered state.

Overall, the primary long-range contacts detected in the native disordered state are consistent between the three spin-label variants and occur between regions around F375 and around T400, and between around L370 and around E385. The former contact is close to an intramolecular contact within Myc present in the Myc-Max crystal structure, where the C-terminus of H1 interacts with H2. In order to test how consistent the measured PREs are with this folded state of Myc, values were calculated on the basis of an isolated Myc monomer from the crystal structure conformation, and a modelled Myc homodimer (using Max as a template) (solid lines in **Fig**. **5a**). This established that the native disordered state of Myc is a much more dynamic system as the PRE data are, in general, inconsistent with ordered, folded states; the calculated profiles for folded Myc show significant peaks and troughs throughout the sequence (blue line in **Fig 5a**). For the Q365-SL variant, the distribution of PREs is very different to the calculated values, particularly around residues A390 and S415, indicating substantial non-native character.

However, for the S405-SL variant, the distribution, though not the size, of PREs more closely resembles the calculated values from the folded structures.

The NMR properties exhibited by the native disordered state of Myc may reflect transient intramolecular or intermolecular interactions, or a combination of both. Consistent with a self-association component, the broadening of resonances for LZ residues showed a small dependence on protein concentration below 1.2 M GdmCl. Consequently, PREs were measured as a function of protein concentration to determine the extent to which intermolecular contacts contributed to the native disordered state. Changes in intensity ratios of less than 10% were observed upon 10-fold dilution (**Fig. S5**). In addition, MTSL-derivatised ^{14}N Myc had a negligible PRE effect on the NMR spectrum of underivatised ^{15}N-labelled Myc, at the lowest GdmCl concentration where all peaks were visible. (**Fig. S6**). Therefore, the protein concentration dependence of signal intensities of LZ residues is ascribed to viscosity or other solvent effects rather than self-association, and the dominant relaxation enhancements and chemical shift changes result from intramolecular contacts.

The structure of the bHLH-LZ domain is compact and disordered.

In order to visualise the properties of an ensemble that is consistent with the data, and to allow calculation of macroscopic properties like the radius of gyration (R_g), the experimental PREs and chemical shifts were used as input into Flexible Meccano/ASTEROIDS calculations (Ozenne et al., 2012; Salmon et al., 2010). The extensive degrees of freedom available to IDPs vastly outweigh the sparse experimental constraints, so the resulting ensemble does not reliably predict chemical shift and PRE data other than those used as input. An initial pool of 10000 conformers was calculated based on the extrapolated chemical shift values ($^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_N$, ^{1}N) at 0 M GdmCl. Five ensembles of 200 conformers that satisfied both the chemical shift and PRE data were then selected from the initial pool using ASTEROIDS, and combined into a final ensemble.

The calculated values for the final ensemble correspond well to the experimental ones (**Fig. S7**). Per-residue secondary structure propensities in the final ensemble (example Ramachandran plots are shown in **Fig. 6a**) are consistent with the SSP predictions above (**Fig. 3**) and show that two main regions of ψ, ϕ space dominate, α-helical and PPII (**Fig. 6b**). The β-strand region is less populated, with just one residue, I381, predominantly (~60%) in this conformation. The 416-421 region shows almost 100% α-helicity, in close agreement SSP predictions based solely on chemical shift.

The contact map and the distribution of R_g values calculated from the final ensembles (**Fig. 6c & d**) illustrate that residues in the 360-380 region are closer to residues in the 400-410 region than would be expected for a random coil ensemble, consistent with the experimental PRE data (**Fig. 5**). Correspondingly, the R_g distribution peaks at 23 Å, which is 5 Å smaller than in the distribution for the initial pool. The compaction is not as large as for the formation of the Myc-Max crystal structure (R_g = 18 Å), and while there are some preferential conformations, the ensemble is largely disordered. Representative protein conformations (**Fig. 7**) illustrate the distribution of helical segments, and the absence of common tertiary structure.

Structure in relation to ligand binding.

A number of molecules that specifically interact with Myc have been reported (Metallo, 2010), but thus far their mode of interaction has been difficult to define (Follis et al., 2008; Hammoudeh et al., 2009; Harvey et al., 2012; Heller et al., 2017). Hence, we investigated whether the CDT-NMR approach could help elucidate the interactions of a molten globule -like IDP with a ligand. The archetypal Myc-targeting molecule, 10058-F4, has an antiproliferative action in cell cultures that is consistent with interrupting the Myc-Max interaction (Yin et al., 2003). Initial NMR studies under native conditions showed some signs of interaction between the measurable resonances of Myc and 10058-F4, but these effects were small, meaning that other solvent effects such as protein and DMSO concentration variations could not be

discounted. Therefore, the experiments were repeated at Gdm^+ concentrations where there is significant population of the molten globule -like state, but the population of the denatured state produces more favourable NMR relaxation behaviour and therefore higher spectral quality.

For this system, the optimal conditions were 0.5 M GdmCl, where the equilibrium position is 80% molten globule -like and 20% denatured, according to the best fit parameters for the data in **Fig. 2b**. Under these conditions, 10058-F4 induces significant attenuation of ^{1}H-^{15}N HSQC crosspeaks, although only minor chemical shift changes (**Fig. 8a**), consistent with slow intermediate exchange. However, whilst the distribution of affected residues is quite broad (**Fig. 8b**), it has a pattern that resembles the PRE profiles (**Fig. 5**), rather than one that resembles the transition between the denatured and molten globule -like states. The greatest loss of intensity is at a region around T400 (shown as isolated orange peaks in **Fig. 8a**)**,** with a weaker effect at residues 360-380. In contrast, residues from K420 onwards are relatively unaffected. This is consistent with the interaction between 10058-F4 and Myc specifically stabilising the previously identified weak tertiary interaction between the 360-380 and 400-410 regions (**Fig. 5**), rather than stabilising all regions with molten globule -like behaviour. The concentration of 10058-F4 is low (1 mM), so the most probable mechanism is by a specific direct interaction with Myc rather than a solvent effect. The region showing greatest intensity loss corresponds to one that has complete signal attenuation under native conditions (**Fig. 1a**), making it difficult to identify using standard NMR approaches. No interaction is apparent at GdmCl concentrations above 0.8 M, due to the low population of residual structure, indicating that the optimal Gdm^+ concentration to investigate ligand binding in different systems will firstly need to be established using CDT-NMR.

Discussion

Previous work on chemically denatured proteins demonstrated that they rapidly interconvert between denatured states with random coil behaviour and more compact, molten

globule -like states, with the population of the more compact forms increasing as conditions become more native-like (Candotti et al., 2013; Cliff et al., 2009; Schulman et al., 1997). NMR measurements reflect the ensemble average behaviour, and measurements under a range of conditions can allow the contribution of sub-populations to be deconvolved. Molten globule -like states frequently have very poor NMR characteristics, with low resonance dispersion and fast relaxation (leading to broad resonances and signal attenuation), and so the chemical denaturant titration (CDT) method allows the determination of otherwise hidden behaviour. The data here show that the Myc bHLH-LZ domain is an IDP with such molten globule -like behaviour. The per residue folding parameters associated with the regions of Myc with high structure propensity are similar enough that chemical shift data for all residues can be fitted with common values, but the parameters do not necessarily describe a transition between states that is concomitant across the molecule. They are more likely to reflect that individual local clusters have similar hydrophobic burial and stability, but mostly form independently of each other.

Previous NMR studies of Myc behaviour have used viral isoforms (v-Myc (Fieber et al., 2001)), variants with some or all of the bHLH-LZ region missing (B-Myc (Burton et al., 2006)), or short peptides derived from the Myc sequence (Hammoudeh et al., 2009; Lavigne et al., 1998). Such studies can be argued not to fully represent the behaviour of the wild-type protein in conditions most relevant to the cellular environment, and to small molecule based intervention in disease states. In this study, significant helical propensity is seen throughout the Myc sequence, and matches well the helical regions identified in v-Myc (Fieber et al., 2001) and short peptides. Furthermore, the method has also allowed PRE measurements to determine longer range interactions and give a fuller description of the structure propensity of the domain. The ensemble calculation suggests that no particular conformation dominates, but the average R_g is considerably smaller than for a random coil. The tertiary contacts in the ensemble are

consistent with the results of a recent molecular dynamics study, which suggest that Myc has a tendency to form a hairpin-like conformation (Liu et al., 2017).

The structural elements defined for Myc in the absence of denaturant are potentially functionally important. The region with the greatest helical content (>90% helical; residues 416-422, **Fig. 3**) corresponds to the position within the leucine zipper that is important in ensuring specificity for the interaction with Max, and mutation of residues E417, R423 and R424 results in significant homodimerisation (Soucek et al., 1998). The long-range tertiary contact defined by the PREs (between the 360-380 and 400-410 regions) coincides with the phosphorylation sites S373 and T400 (Macek et al., 2018), suggesting that phosphorylation perturbs this conformational ensemble, leading to its role in the mechanism of gene regulation. In addition, mutation of R367 is sufficient to allow homodimerisation (Beaulieu et al., 2012), an effect that is ascribed to electrostatic repulsion, but might be caused by the stabilisation of non-native contacts. Furthermore, the regions of tertiary contacts coincide with the regions that are affected by the small molecule inhibitor, 10058-F4, suggesting that a similar regulatory effect can be induced by pharmaceuticals, and holds hope for the design of further drug candidate molecules.

Non-random coil behaviour in disordered protein chains is well-established, both for conventionally folded proteins (N-PGK (Cliff et al., 2009), Drk-SH 3 (Marsh et al., 2007), Staph Nuclease (Zhang et al., 1997)), and for IDPs. For example, our results broadly resemble those obtained with spin-labelled variants of α-synuclein (Bertoncini et al., 2005; Dedmon et al., 2005), which in combination with residual dipolar couplings (RDCs), defined tertiary interactions to be present in the absence of denaturant, and that contribute even at 8 M Urea. The interactions are between the C-terminus and the aggregation prone NAC region of α-synuclein. Studies with Tau show it to have both compact and extended regions (Schwalbe et al., 2014), whereas the contact map for Myc shows higher level of compactness for most

regions of the construct. On the basis of these studies with Tau and α-synuclein, a link was proposed between polyproline II propensity and aggregation prone precursors for pathogenic β-strand formation, but while there are polyproline II favoured regions in Myc, which coincide with the loop in Myc-Max crystal structure, the NMR data show they are not prone to aggregation.

In conclusion, chemical denaturant titration NMR (CDT-NMR) has allowed us to explore fully the conformational ensemble of a wild-type protein domain containing all the elements required to interact with its binding partners, Max and DNA, despite many resonances being broadened beyond detection in the absence of denaturant, and others having little or no intensity in triple-resonance spectra. The application of the CDT method extends from previous Myc studies by allowing the delineation of native behaviour from that of the chemically denatured state. The utility of the CDT-NMR method for regions that have poor spectral properties in the absence of denaturant has allowed the behaviour of the leucine zipper region to be determined, which has not been possible using previous methods. The signals that are most attenuated by conformational exchange are inherently those with the strongest structure propensity. It follows from the conventional structure-function paradigm that such regions are most likely to be functionally significant (as here, the most structured region defines Myc-Max specificity) and, potentially, the most druggable. Even in fuzzy complexes, where the structure function paradigm does not apply so rigourously, specificity is likely to be defined by small clusters of transient structure. Having defined the denaturant dependence of structured regions, a system can be poised at an equilibrium position where the contribution from more compact conformers is significant, but not sufficient to attenuate the NMR signals deleteriously. This approach can be used, for example, to delineate interactions with inhibitors, making IDPs with molten globule –like behaviour amenable to hitherto prohibited screening of lead drug

compounds. The method should also be compatible with other NMR measurements such as RDCs, and the identification of protein-protein interactions.

Acknowledgements:

SP PhD studentship was funded jointly by Bruker and a President's Doctoral Scholar Award of the University of Manchester, PM was an AZ funded post-doctoral researcher. All data recorded at Manchester Biomolecular NMR facility.

Declaration of Interests

JWMN, KJE and RD are employees of the AstraZeneca and may have stock/stock options in AstraZeneca PLC.

Author Contributions

S.P. produced and purified the protein, performed NMR experiments, analysed the data, and wrote the paper. M.J.C. designed experiments, analysed the data and wrote the paper. P.M. designed protein constructs, carried out NMR experiments and analysed data. M.R.J. and M.B. performed the structural ensemble selections using ASTEROIDS. J.W.M.N., K.J.E. and R.D. designed and coordinated experiments. J.P.W. supervised the project. All authors discussed results and commented on the manuscript.

Figure Legends

Figure 1: (**a**) ^{1}H-^{15}N HSQC spectrum of Myc bHLH-LZ domain under the standard buffer conditions and 0 M GdmCl, **(b)** ^{1}H-^{15}N HSQC spectrum of Myc bHLH-LZ domain under the standard buffer conditions with addition of GdmCl to 3.2 M. Amide cross peaks are labelled with the residue number and single letter code for their assigned amino acid. No peaks are visible in spectrum (**a**) for residues 400-437. Inset in (b) shows the crowded region marked with a box in the centre of the spectrum.

Figure 2. (**a**) Primary structure of Myc bHLH-LZ domain with residues colour coded according to category of behaviour, mostly hyperbolic (blue), mostly sigmoidal (red) and

showing both transitions (green), with the secondary structure boundaries taken from the Myc-Max crystal structure 1nkp (B: basic helix, H1: helix 1, H2, helix 2, LZ; leucine zipper heptad repeats). Circle radii reflect the uncertainty in the chemical shift measurements (0.005 ppm for δ^1H_N and 0.02 ppm for $\delta^{15}N_H$). (**b**) Titration behaviour of amide crosspeaks for a selection of residues. Top row shows positional changes in ^1H-^{15}N HSQC spectra with the axes origin indicating the starting chemical shift (lowest concentration of GdmCl giving a visible peak). Middle row and bottom row show δ^1H_N and $\delta^{15}N_H$ (respectively) behaviour as a function of GdmCl concentration. Solid lines show behaviour expected from the best fit parameters generated by a global fit to *Equation 1* (see **Methods**), where chemical shift related parameters are varied for each residue, but underlying physical constants reflect the whole protein. The best fit global parameters are $K_{N/D}$ = 14 +/- 3 (-6.2 +/- 0.4 kJ.mol^{-1}), m = -2.8 +/- 0.2 (6.5 +/- 0.4 kJ.mol^{-1}.M^{-1}), $K_{D\ (Cl)}$ = 410 +/- 40 mM (14 kJ.mol^{-1}). The mean χ^2 was 12.9 with 24 degrees of freedom.

Figure 3: Secondary structure prediction from SSP on the basis of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts, at the labelled denaturant concentrations. Positive values indicate helix and negative values indicate β-strand or PPII helix. Filled bars show predictions based on measured chemical shifts, whereas predicted values for residues lost through signal broadening are shown as open bars.

Figure 4: Denaturant dependence of PREs for example residues, showing best fit relationship for *Equation* 2 as solid lines. Crosspeak intensity ratios between paramagnetic and diamagnetic samples are shown as circles, with the radii showing the calculated errors.

Figure 5: Sequence distribution of PREs (I_{para}/I_{dia}) at 0 M (**a**) and 3.2 M GdmCl (**b**), for each spin-labelled variant. For residues that experience significant line-broadening (400-417), values at 0 M GdmCl were based on the fits of denaturation profiles to *Equation 2*. In **a**, the solid lines show the behaviour expected for the Myc-Max crystal structure monomer (black) and a

modelled homodimer (blue). In **b**, the red solid line shows the behaviour expected for a self-excluding random coil with the sequence specific secondary structure propensity (an ensemble generated by flexible mecanno/ASTEROIDS).

Figure 6: Structural parameters for the ensemble selected by ASTEROIDS (**a**) Ramachandran plots showing the amino acid conformational potentials for representative residues V354, L377, K398 and L420 in the ensemble. **(b**) Populations of residues in the ensemble with dihedral angles in particular regions of Ramachandran space: α-helix (green), PPII (dark-blue), β-strand (red), with the PPII and regions being summed and defined as negative. The number of conformers populating left-handed helix was negligibly small for all residues (<3%). Definition of regions of Ramachandran space was done as described before (Ozenne et al., 2012). (**c**) Distribution of radius of gyration (R_g) for the random coil (red) and the final PRE, chemical shift-based ensemble (blue). R_g for the random coil state is 29 Å and decreases to 28 Å when the chemical shifts are included in calculations. Inclusion of PRE data reduces this to 23 Å, a compaction of 82%. The compaction expected from the Myc-Max crystal structure is 58%. (**d**) Final contact map showing Myc long-range contacts derived using chemical shift and PRE data. Heat-map represented in terms of log-ratio of the distance between residues in the selected and chemical shift -based pool $\Delta_{ij} = \log(\langle d_{ij}\rangle / \langle d_{ij,ref}\rangle)$. Colours range from red (-0.4, regions in closer contact than pool (see Methods)) to blue (0.0) to violet (0.2, regions farther apart than pool).

Figure 7. Representative structures from the final ensemble selected by ASTEROIDS: a) lowest R_g conformer, b) most frequently selected conformer c) closest to mean R_g conformer d) mode R_g conformer. Structures are represented as a ribbon coloured blue to red, N- to C-termini. Consistent helical region is in orange region.

Figure 8: Ligand binding at 0.5 M GdmCl. (**a**) HSQC spectra of 100 μM Myc recorded at 0.5 M GdmCl and 5% DMSO-d6, in the presence (dark blue) and absence (orange) of 1 mM

10058-F4. (**b**). Ratio of peak height for assigned resonances in the two spectra, showing regions of Myc affected by the compound

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jonathan Waltho (j.waltho@manchester.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental model is recombinant, human Myc protein (RRID:SCR_008608) C-terminal domain (bHLH-LZ) residues 351-437. It was expressed in *E coli* BL21(DE3) Gold from a pET -derived vector with an N-terminal HisTag, which was removed by proteolysis during purification. Standard conditions were defined as 20 mM phosphate buffer, pH=6.5, T=278 K, with protein concentration 5mg/ml (410 μM).

Four single cysteine mutants were prepared for site-specific electron spin-labelling; Q365C, N386, S405C and Q411C.

METHOD DETAILS

The bHLH-LZ domain of Myc (residues 351-437) and the cysteine point-mutants were expressed from a pET derived vector in *E. Coli* BL21 (DE3) Gold cell lines (Agilent Technologies) at 37 °C (310 K) in M9 minimal medium supplemented with 2 g/L of ^{13}C glucose (or 4 g/L of ^{12}C glucose) and 1 g/L ^{15}N ammonium chloride (Sigma-Aldrich) as the only carbon and nitrogen sources. Protein expression was induced by addition of 0.1 mM IPTG to bacterial cultures at OD_{600}=0.9, which were harvested 4 h after induction and frozen at -80 °C. Cell pellets were resuspended in 20 mM phosphate buffer with 8 M urea and protease inhibitor tablet "Complete" (Roche), sonicated and centrifuged for 30 min at 40000 g. Supernatant was loaded on a Talon Co^{2+} affinity column, equilibrated with 3 M GdmCl, 20 mM phosphate, 0.1 M NaCl, pH=7.4, washed by 2.5 mM imidazole and Myc was eluted by addition

of 125 mM imidazole. The His-tag was cleaved off by TEV protease and the resulting His-tag peptides were removed by further Co^{2+}-affinity chromatography. The purity and identity of the proteins were validated by electrospray ionization mass spectrometry (ESI-MS) and SDS-PAGE.

Paramagnetic labelling of Myc was achieved by incubating Myc cysteine mutants with MTSL (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate; Toronto Research Chemicals, Canada) in the presence of 3 M GdmCl and 1 mM DTT at RT in the dark for 8 h. MTSL was present in 5:1 excess over the total thiol concentration. Reaction completeness was confirmed by mass-spectrometry to be more than 95%. Samples were buffer exchanged in 20 mM phosphate, 1 mM EDTA, pH 6.5.

Buffer conditions were screened and optimized to provide best spectral dispersion, highest signal intensity and least visible aggregation for Myc samples. The buffer screen covered pHs from 5 to 7.5 and NaCl concentrations up to 1.2 M, and used either phosphate or Bis-Tris to control pH. The effect of temperature was also studied. Standard conditions were defined as 20 mM phosphate buffer, pH=6.5, T=278 K, with protein concentration 5mg/ml (410 μM).

All samples for NMR spectroscopy were prepared in 20 mM phosphate buffer, pH=6.5 and varying concentration of GdmCl with addition of 10% D_2O and 1 mM TSP (Trimethylsilyl propanoic acid). Except where stated, spectra were recorded at 278 K on a Bruker AVANCE III 800 MHz spectrometer equipped with a TCI cryoprobe (^{1}H-^{13}C/^{15}N with z-gradients) in 3mm tubes. Gradient selective, sensitivity-enhanced HSQC spectra had a spectral width of 20 ppm (1622 Hz) and apparent acquisition time of 60 ms for the indirect dimension. Spectra were processed using Topspin 3.2 software (Bruker Corp).

Protein backbone resonance assignment experiments were recorded for Myc at 3.2 M GdmCl, comprising two-dimensional ^{15}N-^{1}H HSQC and triple resonance experiments (H)N(CA)NNH, HNCACB, HNCA, HNCO, HN(CO)CA and HN(CO)CACB spectra. All spectra were collected

using Echo/Antiecho-TPPI gradient selection, which was efficient at suppressing signals from GdmCl and other buffer components. Non-uniform sampling (NUS) was used to optimize resolution of the indirect dimensions in the available experiment time. NUS acquired data were process using MDD algorithm within Topspin. Spectra were visualised and analysed using CCPN Analysis 2.3.

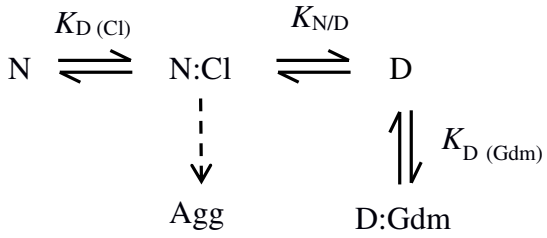
Two-dimensional ^{1}H-^{15}N HSQC spectra were collected at a range of GdmCl concentrations (from 0 to 3.2 M with 0.2 M step). In order to confirm validity of the resonance assignment transfer between denaturant concentrations HNCACB spectra were collected for 2.4, 1.6 and 0 M GdmCl. This led to unambiguous assignment of Myc at native-like conditions even for very low intensity peaks. In addition, a sodium chloride titration of Myc between 0 and 1.2 M (with 0.2 M step) was collected. To characterize the Myc self-association and aggregation, protein concentrations (0, 0.025, 0.05, 0.1, 0.2, 0.4, 0.8 mM) were varied and ^{1}H-^{15}N HSQC spectra recorded, repeated at 0, 0.6, 1 and 2 M GdmCl.

Paramagnetic relaxation enhancement (PRE) data were collected for diamagnetic and paramagnetic samples at a range of GdmCl concentrations (0 to 3.2 M with 0.3 M step). Diamagnetic sample was obtained by addition of a 5-fold excess of sodium L-ascorbate to the spin-labelled protein sample. Relaxation delays for ^{15}N-^{1}H HSQC were 2 s, which was sufficient to allow 90% signal recovery between transients.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were processed in Topspin 3.2 and quantified in CCPN Analysis 2.3. Data fitting was by Levenberg-Marquadt non-linear least squares optimisation using in-house routines interfacing with Numerical Python.

Chemical shifts were analysed as follows: the total reaction scheme for the interaction between Myc and GdmCl can be described by the Scheme 1:



Scheme 1

where N stands for native disordered state of Myc, D for denatured and N:Cl and D:Gdm indicate the bound forms. The dotted arrow represents a very slow precipitation of Myc in the presence of elevated Cl^- concentrations, which is countered by Gdm^+ ions. Because the aggregation process is slow and irreversible, it doesn't contribute to chemical shifts. In addition, the dissociation constant for Gdm^+ binding to the unfolded state is so high (>4M) that the effect can be approximated by a linear function. Thus, the chemical shift data can be analysed according to the following simplified scheme:

$K_{D(Cl)}$ is the binding constant for chloride and $K_{N/D}$ is equilibrium constant between the denatured and the native disordered states. Proton chemical shift perturbation dependence by GdmCl was fitted to equation:

$$\delta H_i = \delta H_0 + \frac{\frac{[GdmCl]}{K_D} * (\Delta\delta H_{Cl} + (\Delta\delta H_{Gdm} + slope_H * [GdmCl]) * K)}{1 + \frac{[GdmCl]}{K_D} * (1 + K)} \text{ with } K = K_0 * e^{-m.[GdmCl]}$$

Equation 1

δH_0 is the initial value of the proton chemical shift difference for residues, $\Delta\delta H_{Cl}$ is the maximum chemical shift perturbation of the proton chemical shift by chloride and $\Delta\delta H_{Gdm}$ is the maximum effect of Gdm. K_0 is the value of K at 0 M GdmCl, m is the denaturant concentration dependence of the free energies. ^{15}N chemical shifts were fitted to the same equation. Initial fitting was to a subset of residues with clearly defined behaviour, fitting a K_D, K_0 and m as global parameters, and chemical shifts on a per residue basis. Subsequently, the values of K_D, K_0 and m were fixed in fits of all residues. No significant difference between the

fit-lines and the data points resulted. The efficacy of using the global parameters to allow extrapolation to zero denaturant for residues with significant signal attenuation under such conditions was tested by omitting data points below 0.6 M GdmCl for complete data sets and repeating the fitting procedure. The RMS deviation between the fitted values and the recorded chemical shifts was 0.03 ppm for δH and 0.3 ppm for δN. These uncertainties were propagated into the estimates of ^{13}C chemical shifts.

$^{13}C_\alpha$ and $^{15}N_H$ chemical shifts changes for each residue were found to correlate, and so the maximum ^{15}N shift defined by data fitting was used to calculate the $^{13}C_\alpha$ shifts at 0 M GdmCl used for SSP. For residues with complete datasets, omitting the 0 M GdmCl point and repeating the process gives a RMS difference from the recorded value of 0.37 ppm for C_α (0.23 after excluding outliers H359 and F375) and similarly 0.75 ppm for C_β (0.23 after excluding the same outliers). Neither of these outlier residues are in regions with strong secondary structure propensity.

PRE effects were calculated from the intensity ratio of HSQC peaks between paramagnetic and diamagnetic samples for each residue, and fitted to the following equation (Cliff et al., 2009).

$$\frac{I_{ox}}{I_{red}} = \frac{R_2'}{R_2'+R_P'}\, e^{-(R_P'.t)} \; where \; R_x' = \frac{R_x^D + R_x^N * K}{(1+K)} \text{ with } K = K_0 * e^{-m.[GdmCl]} \qquad \textit{Equation 2}$$

The exchange between conformers is fast relative to the chemical shift timescale, and therefore assumed to be fast on the ^1H-e relaxation timescale, and therefore the apparent transverse relaxation rate in diamagnetic samples (R_2') and the additional paramagnetic relaxation rate (R_P') are population weighted averages between the rates in the native, denatured state (N) populated at 0 M GdmCl, and the denatured state (D) populated at high GdmCl concentrations. The time t is the amount of time protons are transverse in the HSQC pulse

sequence, amounting to 10.6 ms. R_2^D was fixed at 7 Hz (×2π) and residue specific R_2^N values were estimated from the guanidinium dependence of the intensities in the diamagnetic sample. These needed correcting for the change in sensitivity of the probe with Cl^- concentration, which is proportional to the change in proton pulse-length (p_D/p_{0M}).

$$I_{red,D} = I_{red,0M} \times \frac{p_D}{p_{0M}} \times \frac{R_2' e^{-\left(R_2' - R_2^U\right).t}}{R_2^U} \quad \textit{Equation 3}$$

with R_2' calculated as in *Equation 2*.

Ensemble calculation

An original ensemble created by *flexible meccano* (Ozenne et al., 2012; Salmon et al., 2010) comprised 10,000 structures with the phi/psi angles corresponding to random coils. 50 random ensembles with 200 structures each were created and χ^2 values were calculated for each ensemble. Then using ASTEROIDS (Salmon et al., 2010) genetic algorithm 5 ensembles with 200 structures each were selected on the basis of the best fit to the experimental chemical shift data. SPARTA (Shen and Bax, 2007) was used for prediction of chemical shifts. From these structures, 1000 phi/psi angles were extracted for each residue, which were used as a library to build 8500 structures for the next iteration. Next iteration started with 8500 structures from the previous calculation and 1500 structures created from the random coil library. This process was repeated 5 times to avoid being trapped in local minima.

These final 10000 structures were used as a starting pool to fit the PRE and chemical shift data simultaneously. ASTEROIDS was used to produce equivalent ensembles containing 100 structures (5000 evolution steps were used in the genetic algorithm). PREs were calculated from ^1H R_2 values estimated from the ^1H-e^- distance in each conformation, the INEPT delay, the conformational sampling of the spin label relative to the backbone and the estimated correlation time of the dipole interaction (5ns), as described elsewhere.(Salmon et al., 2010)

Long-range order was assessed in the final ensembles by calculating distance values normalized against the average distance values calculated for the chemical shift (CS) based ensemble:

$$\Delta_{ij} = log\left(\frac{d_{ij}}{d_{ij}^{0}}\right)$$

where d_{ij} is the distance between residues i and j in the final (CS-PRE) calculated ensemble and d_{ij}^{0} is distance between residues i and j for the CS calculated ensemble. These values were used to plot contact map.

DATA AND SOFTWARE AVAILABILITY

NMR chemical shifts have been deposited with the BioMagResBank with accession codes 27701, 27702, 27703 and 27704.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
BL21 (DE3) Gold *E coli.*	Agilent tecnologies	CAT# 230132
Chemicals, Peptides, and Recombinant Proteins		
Human Myc protein (RRID:SCR_008608) C-terminal domain (bHLH-LZ) residues 351-437	In-house	
^{13}C glucose	Sigma Aldrich	CAT# 389374
^{15}N ammonium chloride	Sigma Aldrich	CAT# 299251
Guanidinium Chloride	Sigma Aldrich	CAT#G4505
TEV protease	In-house	
Talon Co2+ affinity column	Sigma Aldrich	CAT#GE28-9575-02
Complete Protease inhibitors	Roche	CAT# 11697498001
MTSL (1-Oxyl-2,2,5,5-tetramethyl-?3-pyrroline-3-methyl) Methanethiosulfonate	Toronto Research Chemicals	CAT#O875000

Deposited Data		
Backbone 1H, 13C, and 15N Chemical Shift Assignments for the Myc bHLH-LZ domain in presence of 3.2 M GdmCl	This paper	BMRB: 27701
Backbone 1H, 13C, and 15N Chemical Shift Assignments for the Myc bHLH-LZ domain in presence of 2.4 M GdmCl	This paper	BMRB: 27702
Backbone 1H, 13C, and 15N Chemical Shift Assignments for the Myc bHLH-LZ domain in presence of 1.6 M GdmCl.	This paper	BMRB: 27703
Backbone 1H, 13C, and 15N Chemical Shift Assignments for the Myc bHLH-LZ domain	This paper	BMRB: 27704
Crystal structure of Myc-Max recognizing	(Nair and Burley, 2003)	PDB 1nkp
Structural ensemble of Myc bHLH-LZ domain consistent with NMR data	This paper	DOI: 10.17632/xhdwd26fdc.1
Backbone 1H, 15N Chemical Shift Assignments for the Myc bHLH-LZ domain as a function of GdmCl concentration.	This paper	DOI: 10.17632/xhdwd26fdc.1
Backbone 1H, 15N Chemical Shift Assignments for the Myc bHLH-LZ domain as a function of NaCl concentration.	This paper	DOI:10.17632/xhdwd26fdc.1
PREs as used in structure model selection	This paper	DOI: 10.17632/xhdwd26fdc.1
Backbone 1H, 13C, 15N Chemical Shift Assignments for the Myc bHLH-LZ domain as used in structure model selection	This paper	DOI: 10.17632/xhdwd26fdc.1
PREs for 3 variants (365-SL, 386-SL and 405-SL) as a function of GdmCl concentration.	This paper	DOI: 10.17632/xhdwd26fdc.1
Experimental Models: Organisms/Strains		
Oligonucleotides		
Recombinant DNA		
6xHis-Tagged-Myc-351-437	This paper	N/A
Software and Algorithms		

Topspin 3.5	Bruker	N/A
CCPN Analysis 2.3	CCPN	N/A
Numeric python routines		N/A
Flexible Meccano	Ozenne *et al.* 2012	N/A
Asteroids	Salmon *et al.* 2010	N/A
SPARTA	Chen & Bax 2007	N/A
Other		

References

Babu, M.M., van der Lee, R., de Groot, N.S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. Current Opinion in Structural Biology *21*, 432-440.

Baldwin, A.J., and Kay, L.E. (2009). NMR spectroscopy brings invisible protein states into focus. Nature Chemical Biology *5*, 808-814.

Beaulieu, M.E., McDuff, F.O., Frappier, V., Montagne, M., Naud, J.F., and Lavigne, P. (2012). New structural determinants for c-Myc specific heterodimerization with Max and development of a novel homodimeric c-Myc b-HLH-LZ. Journal of molecular recognition : JMR *25*, 414-426.

Bermel, W., Bertini, I., Chill, J., Felli, I., Haba, N., Kumar, M., and Pierattelli, R. (2012). Exclusively heteronuclear (13) C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). Chembiochem *13*, 2425-2432.

Bertoncini, C.W., Jung, Y.-S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M., and Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. Proceedings of the National Academy of Sciences of the United States of America *102*, 1430-1435.

Bhowmick, A., Brookes, D.H., Yost, S.R., Dyson, H.J., Forman-Kay, J.D., Gunter, D., Head-Gordon, M., Hura, G.L., Pande, V.S., Wemmer, D.E., *et al.* (2017). Finding our way in the dark proteome. Journal of the American Chemical Society *138*, 9730-9742.

Blackwood, E.M., and Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science *251*, 1211.

Burton, R.A., Mattila, S., Taparowsky, E.J., and Post, C.B. (2006). B-Myc: N-Terminal Recognition of Myc Binding Proteins. Biochemistry *45*, 9857-9865.

Candotti, M., Esteban-Martín, S., Salvatella, X., and Orozco, M. (2013). Toward an atomistic description of the urea-denatured state of proteins. Proc Natl Acad Sci U S A *110*, 5933-5938.

Clarke, A.R., and Waltho, J.P. (1997). Protein folding pathways and intermediates. Current Opinion in Biotechnology *8*, 400-410.

Cliff, M.J., Craven, C.J., Marston, J.P., Hounslow, A.M., Clarke, A.R., and Waltho, J.P. (2009). The Denatured State of N-PGK Is Compact and Predominantly Disordered. Journal of Molecular Biology *385*, 266-277.

Clore, G.M. (2013). Generating accurate contact maps of transient long-range interactions in intrinsically disordered proteins by paramagnetic relaxation enhancement. Biophys J, *104*, 1635-1636.

Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C.M. (2005). Mapping Long-Range Interactions in α-Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. Journal of the American Chemical Society *127*, 476-477.

Fieber, W., Schneider, M.L., Matt, T., Krautler, B., Konrat, R., and Bister, K. (2001). Structure, Function, and Dynamics of the Dimerization and DNA-binding Domain of Oncogenic Transcription Factor v-Myc. J Mol Biol *307*, 1395-1410.

Follis, A.V., Hammoudeh, D.I., Daab, A., and Metallo, S.J. (2009). Small-molecule perturbation of competing interactions between c-Myc and Max. Bioorg Med Chem Lett *19*, 807-810.

Follis, A.V., Hammoudeh, D.I., wang, H., Prochownik, E.V., and Metallo, S.J. (2008). Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules. Chemical Biology *15*, 1149-1155.

Goradia, N., Wiedemann, C., Herbst, C., Görlach, M., Heinemann, S.H., Ohlenschläger, O., and Ramachandran, R. (2015). An approach to NMR assignment of intrinsically disordered proteins. . Chemphyschem *16*, 739-746.

Habchi, J., Tompa, P., Longhi, S., and Uversky, V.N. (2014). Introducing protein intrinsic disorder. Chemical Reviews *114*, 6561–6588.

Hammoudeh, D.I., Follis, A.V., Prochownik, E.V., and Metallo, S.J. (2009). Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc. Journal of the American Chemical Society *131*, 7390-7401.

Harvey, S.R., Porrini, M., Stachl, C., MacMillan, D., Zinzalla, G., and Barran, P.E. (2012). Small-Molecule Inhibition of c‑MYC:MAX Leucine Zipper Formation Is Revealed by Ion Mobility Mass Spectrometry. J Am Chem Soc *134*, 19384−19392.

Heller, G.T., Aprile, F.A., Bonomi, M., Camilloni, C., De Simone, A., and Vendruscolo, M. (2017). Sequence Specificity in the Entropy-Driven Binding of a Small Molecule and a Disordered Peptide. Journal of Molecular Biology *429*, 2772-2779.

Huang, J.-r., Gabel, F., Jensen, M.R., Grzesiek, S., and Blackledge, M. (2012). Sequence-Specific Mapping of the Interaction between Urea and Unfolded Ubiquitin from Ensemble Analysis of NMR and Small Angle Scattering Data. Journal of the American Chemical Society *134*, 4429-4436.

Jensen, M.R., Zweckstetter, M., Huang, J.-R., and Blackledge, M. (2014). Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. Chemical Reviews *114*, 6632-6660.

Lavigne, P., Crump, M.P., Gagné, S.M., Hodges, R.S., Kay, C.M., and Sykes, B.D. (1998). Insights into the mechanism of heterodimerization from the 1H-NMR solution structure of the c-Myc-Max heterodimeric leucine zipper11Edited by P. E. Wright. Journal of Molecular Biology *281*, 165-181.

Liu, J., Dai, J., He, J., Niemi, A.J., and Ilieva, N. (2017). Multistage modeling of protein dynamics with monomeric Myc oncoprotein as an example. Phys Rev E *95*, 032406.

Macek, P., Cliff, M.J., Embrey, K.J., Holdgate, G.A., Nissink, W.M., Panova, S., Waltho, J.P., and Davies, R.A. (2018). Myc phosphorylation in its basic helix–loop–helix region destabilizes transient α-helical structures, disrupting Max and DNA binding (*in press). J Biol Chem 293*, -.

Marsh, J.A., Neale, C., Jack, F.E., Choy, W.Y., Lee, A.Y., Crowhurst, K.A., and Forman-Kay, J.D. (2007). Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. Journal of Molecular Biology *367*, 1494-1510.

McParland, V.J., Kalverda, A.P., Homans, S.W., and Radford, S.E. (2002). Structural properties of an amyloid precursor of β2-microglobulin. Nature Structural Biology *9*, 326.

Meier, S., Grzesiek, S., and Blackledge, M. (2007). Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. Journal of the American Chemical Society *129*, 9799-9807.

Metallo, S.J. (2010). Intrinsically disordered proteins are potential drug targets. Current Opinion in Chemical Biology *14*, 481–488.

Milanesi, L., Waltho, J.P., Hunter, C.A., Shaw, D.J., Beddard, G.S., Reid, G.D., Dev, S., and Volk, M. (2012). Measurement of energy landscape roughness of folded and unfolded proteins. Proceedings of the National Academy of Sciences *109*, 19563.

Myers, J.K., Pace, C.N., and Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein science : a publication of the Protein Society *4*, 2138-2148.

Nair, S.K., and Burley, S.K. (2003). X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors. Cell *112*, 193-205.

Oldfield, C.J., and Dunker, A.K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem *83*, 553-584.

Ozenne, V., Schneider, R., Yao, M.X., Huang, J.R., Salmon, L., Zweckstetter, M., Jensen, M.R., and Blackledge, M. (2012). Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. Journal of the American Chemical Society *134*, 15138-15148.

Plaxco, K.W., Morton, C.J., Grimshaw, S.B., Jones, J.A., Pitkeathly, M., Campbell, I.D., and Dobson, C.M. (1997). The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. J Biomol NMR *10*, 221-230.

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell *65*, 395-407.

Reed, M.A.C., Jelinska, C., Syson, K., Cliff, M.J., Splevins, A., Alizadeh, T., Hounslow, A.M., Staniforth, R.A., Clarke, A.R., Craven, C.J.*, et al.* (2006). The denatured state under native

conditions: A non-native-like collapsed state of N-PGK. Journal of Molecular Biology *357*, 365-372.

Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M.R., Zweckstetter, M., and Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. Journal of the American Chemical Society *132*, 8407-8418.

Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. Nat Rev Cancer *10*, 59-64.

Scholtz, J.M., Grimsley, G.R., and Pace, C.N. (2009). SOLVENT DENATURATION OF PROTEINS AND INTERPRETATIONS OF THE M VALUE. In Methods in Enzymology, Vol 466: Biothermodynamics, Pt B, M.L. Johnson, J.M. Holt, and G.K. Ackers, eds., pp. 549-565.

Schulman, B.A., Kim, P.S., Dobson, C.M., and Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. Nat Struct Biol *4*, 630-634.

Schwalbe, M., Ozenne, V., Bibow, S., Jaremko, M., Jaremko, L., Gajda, M., Jensen, Malene R., Biernat, J., Becker, S., Mandelkow, E.*, et al.* (2014). Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution from NMR and Small Angle Scattering. Structure *22*, 238-249.

Shen, Y., and Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR *38*, 289-302.

Sormanni, P., Piovesan, D., Heller, G., Bonomi, M., Kukic, P., Camilloni, C., Fuxreiter, M., Dosztanyi, Z., Pappu, R., Babu, M.M.*, et al.* (2017). Simultaneous quantification of protein order and disorder. Nature Chemical Biology *13*, 339-342.

Soucek, L., Helmer-Citterich, M., Sacco, A., Jucker, R., Cesareni, G., and Nasi, S. (1998). Design and properties of a myc derivative that efficiently homodimerizes. Oncogene *17*, 2463-2472.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., *et al.* (2014). Classification of Intrinsically Disordered Regions and Proteins. Chemical Reviews *114*, 6589-6631.

Wiedemann, C., Goradia, N., Häfner, S., Herbst, C., Görlach, M., Ohlenschläger, O., and Ramachandran, R. (2015). HN-NCA heteronuclear TOCSY-NH experiment for (1)H(N) and (15)N sequential correlations in ((13)C, (15)N) labelled intrinsically disordered proteins. J Biomol NMR *63*, 201-212.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol *16*, 18-29.

Yin, X., Giap, C., Lazo, J.S., and Prochownik, E.V. (2003). Low molecular weight inhibitors of Myc–Max interaction and function. Oncogene *22*, 6151.

Zhang, O.W., Kay, L.E., Shortle, D., and FormanKay, J.D. (1997). Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta 131 Delta, using NMR methods with improved resolution. Journal of Molecular Biology *272*, 9-20.

Figure 1

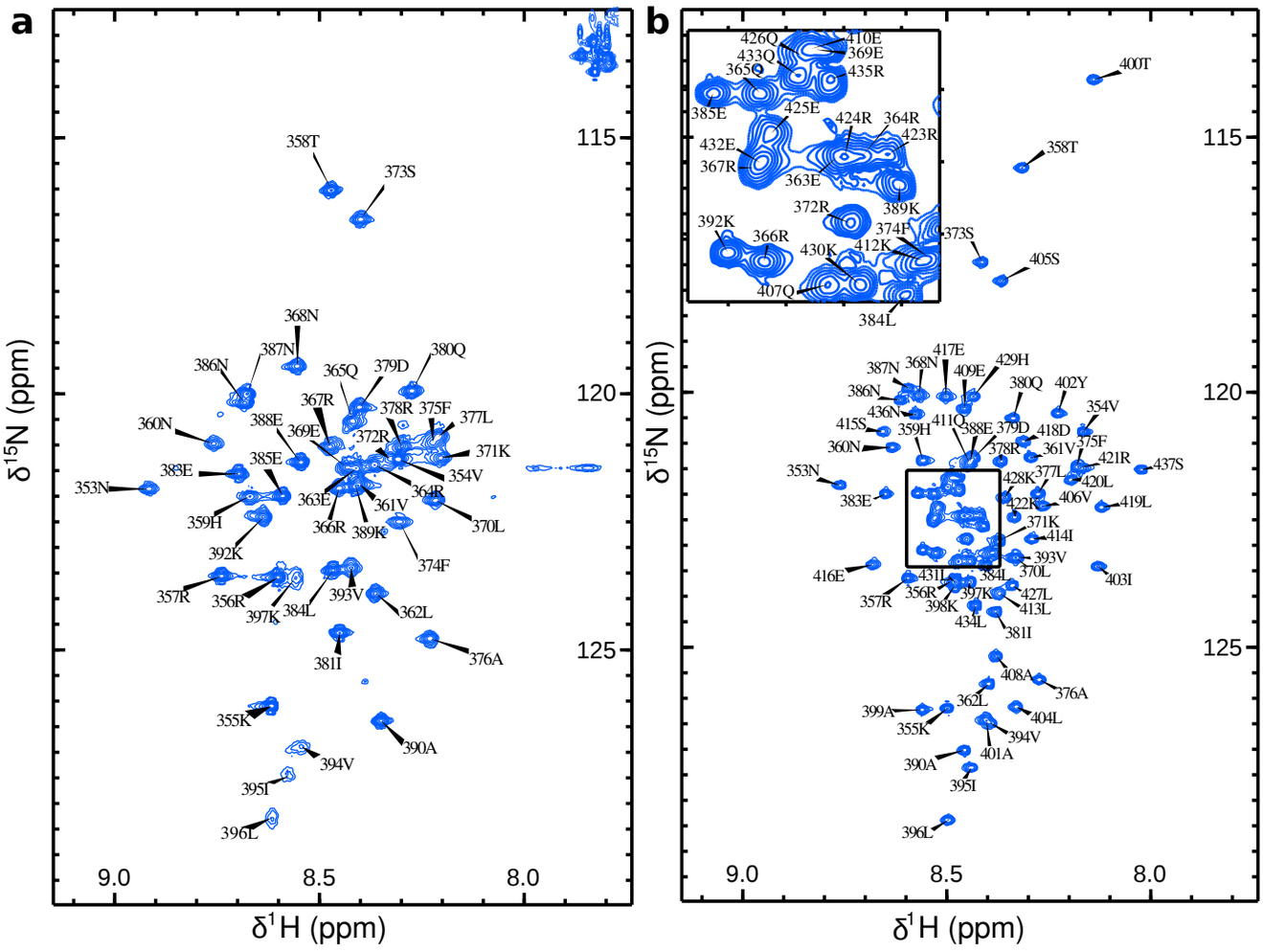

a
b
δ15N (ppm)
δ1H (ppm)
115
120
125
9.0
8.5
8.0
358T
373S
368N
387N
386N
365Q
379D
380Q
367R
378R
375F
377L
360N
388E
369E
372R
371K
385E
383E
354V
353N
363E
364R
361V
359H
366R
389K
370L
392K
374F
357R
356R
393V
397K
384L
362L
381I
376A
355K
390A
394V
395I
396L
426Q
410E
433Q
369E
365Q
435R
385E
425E
424R
364R
423R
432E
367R
363E
389K
372R
392K
374F
373S
366R
430K
412K
407Q
384L
400T
358T
405S
417E
429H
387N
368N
409E
380Q
402Y
386N
354V
436N
415S
411Q
359H
388E
379D
418D
360N
378R
361V
375F
421R
437S
353N
428K
377L
420L
383E
422R
406V
419L
371K
414I
393V
416E
431L
384L
370L
403I
357R
356R
397K
427L
398K
413L
434L
381I
408A
376A
399A
362L
404L
355K
394V
401A
390A
395I
396L

Figure 2

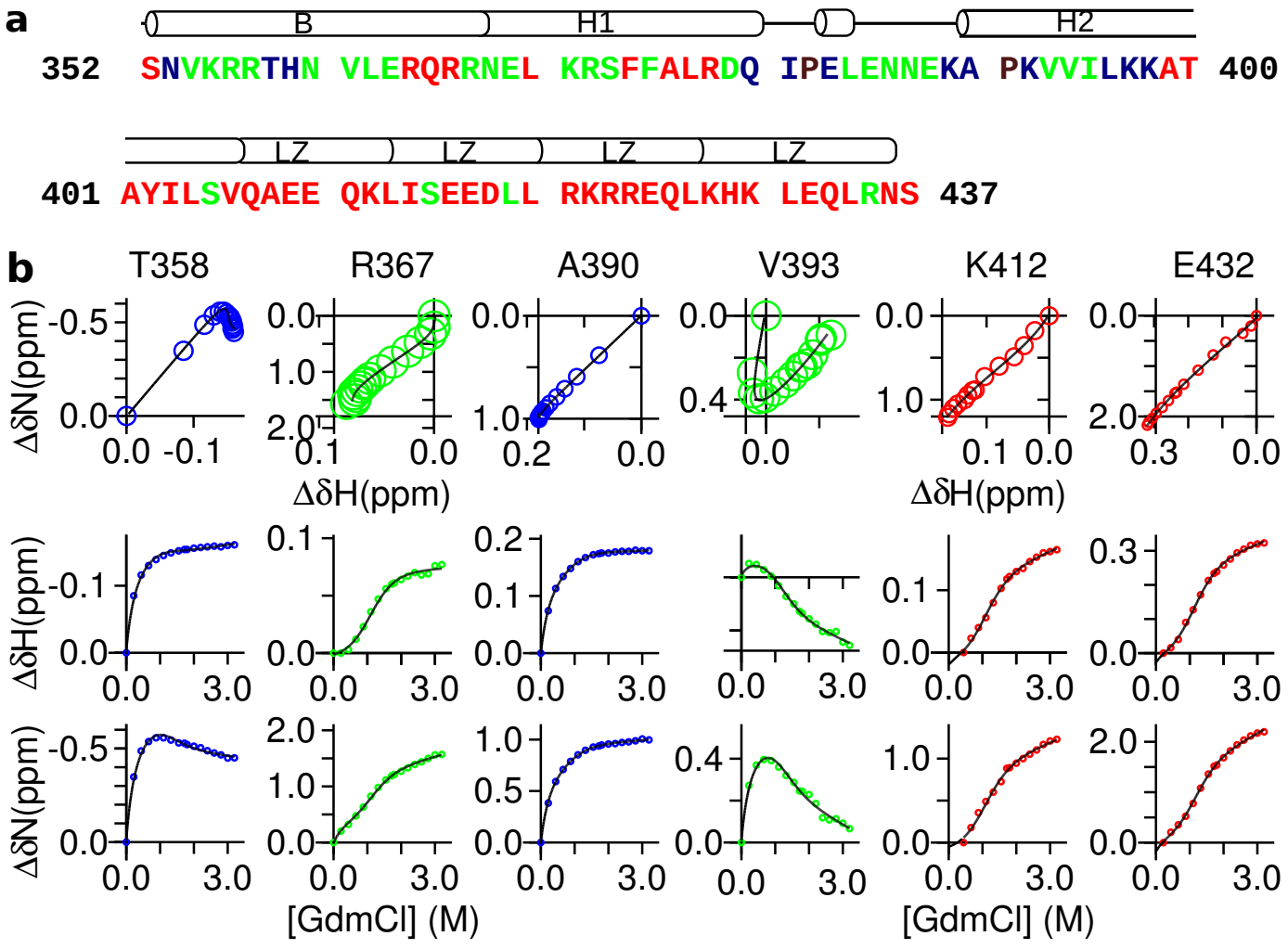

a
B
H1
H2
352 SNVKRRTHN VLERQRRNEL KRSFFALRDQ IPELENNEKA PKVVILKKAT 400
LZ
LZ
LZ
LZ
401 AYILSVQAEE QKLISEEDLL RKRREQLKHK LEQLRNS 437
b
T358
R367
A390
V393
K412
E432
ΔδN(ppm)
ΔδH(ppm)
ΔδH(ppm)
ΔδN(ppm)
[GdmCl] (M)
[GdmCl] (M)

Figure 3

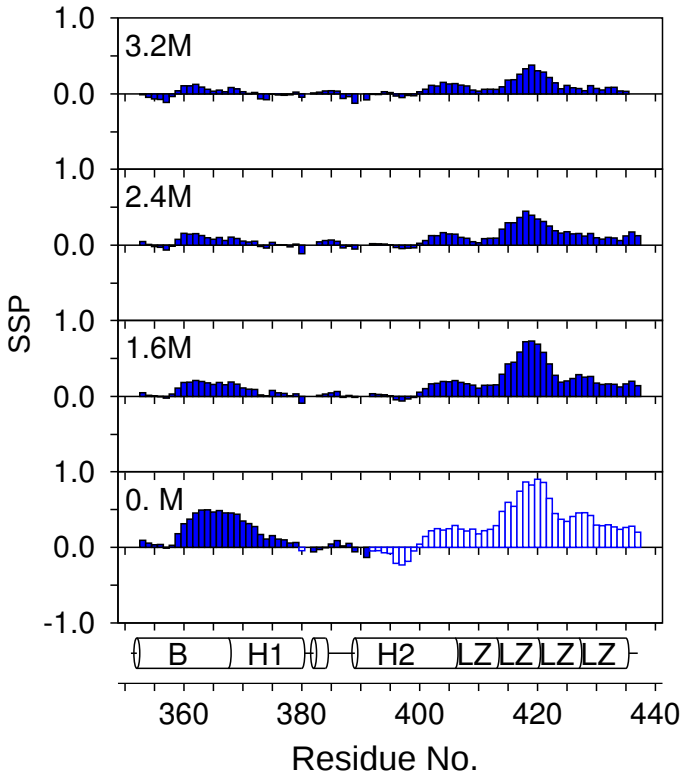

1.0
3.2M
0.0
1.0
2.4M
0.0
1.0
1.6M
0.0
1.0
0. M
0.0
-1.0
SSP
B
H1
H2
LZ
LZ
LZ
LZ
360
380
400
420
440
Residue No.

Figure 4

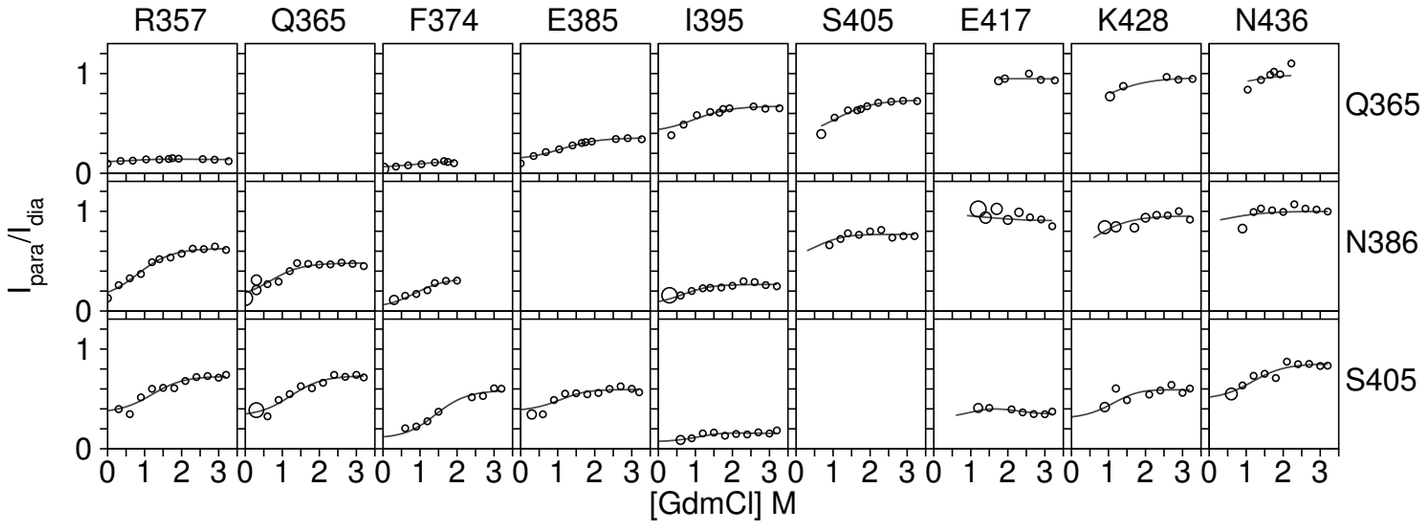

R357
Q365
F374
E385
I395
S405
E417
K428
N436
Q365
N386
S405
I_{para}/I_{dia}
[GdmCl] M

Figure 5

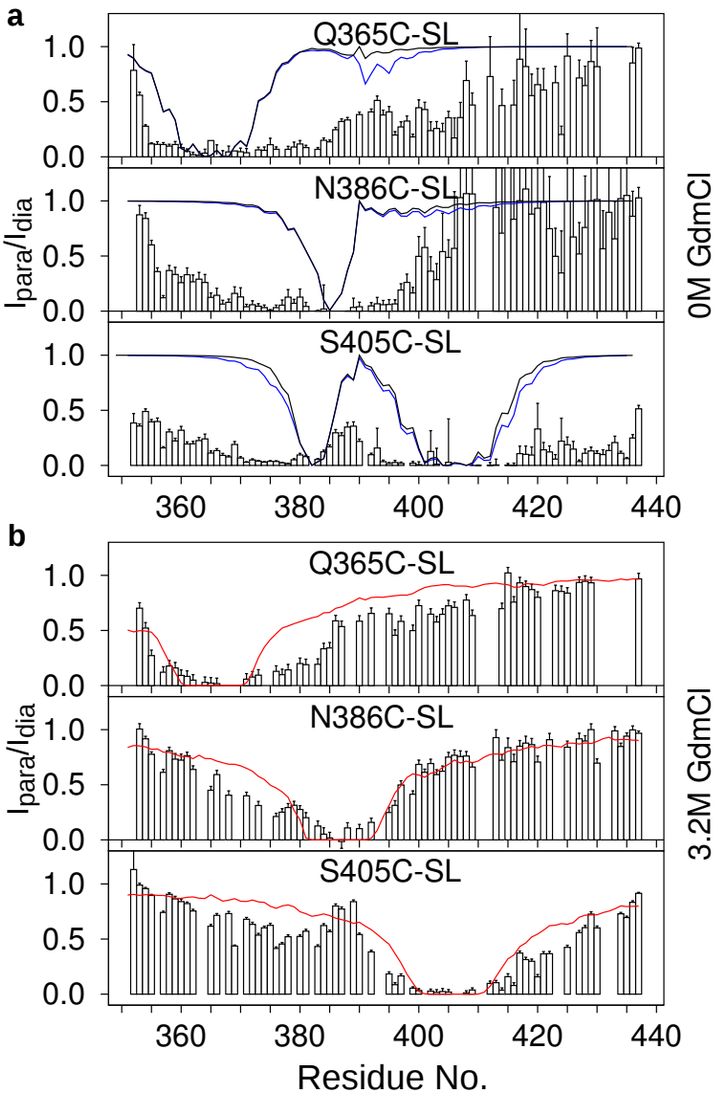

a
Q365C-SL
N386C-SL
S405C-SL
1.0
0.5
0.0
I_{para}/I_{dia}
0M GdmCl
360
380
400
420
440
b
Q365C-SL
N386C-SL
S405C-SL
3.2M GdmCl
Residue No.

Figure 6

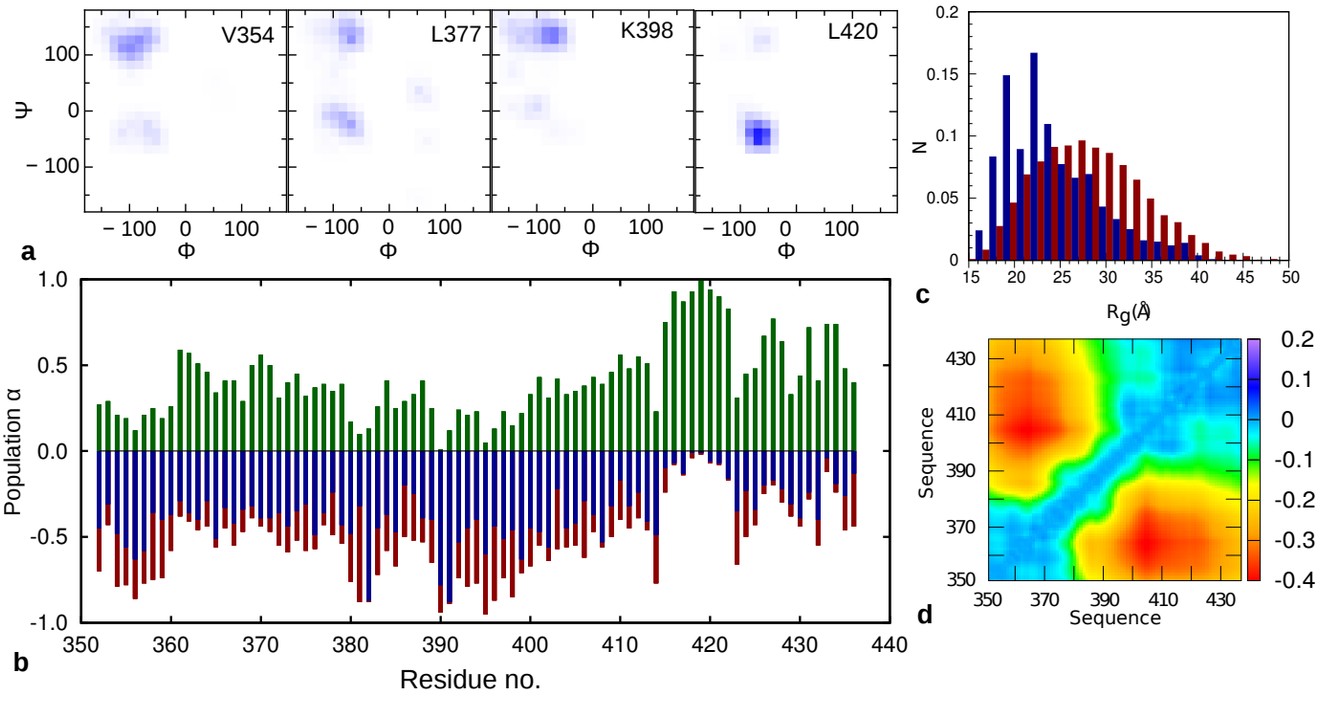

V354
L377
K398
L420
Ψ
Φ
a
b
Population α
Residue no.
c
N
Rg(Å)
d
Sequence
Sequence

Figure 7

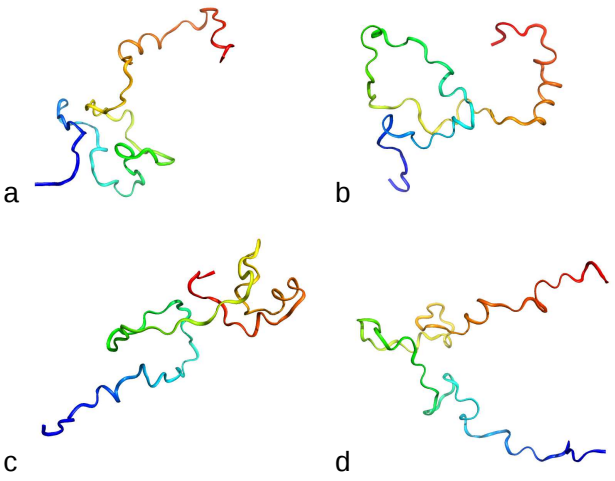
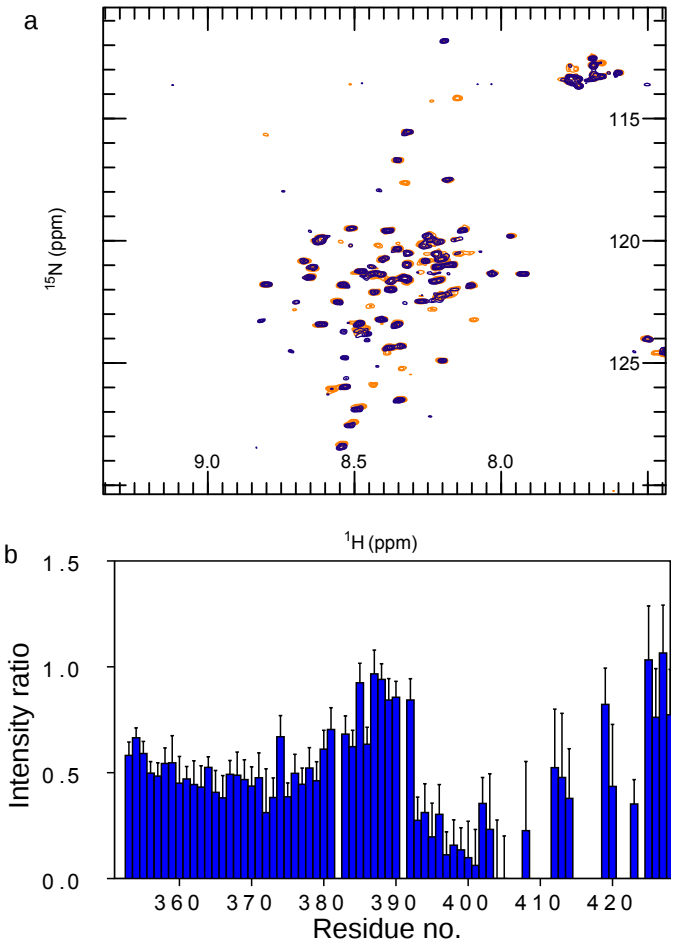


Figure 8


a
^{15}N (ppm)
115
120
125
9.0
8.5
8.0
^{1}H (ppm)
b
Intensity ratio
1.5
1.0
0.5
0.0
360
370
380
390
400
410
420
Residue no.

Supplemental Data

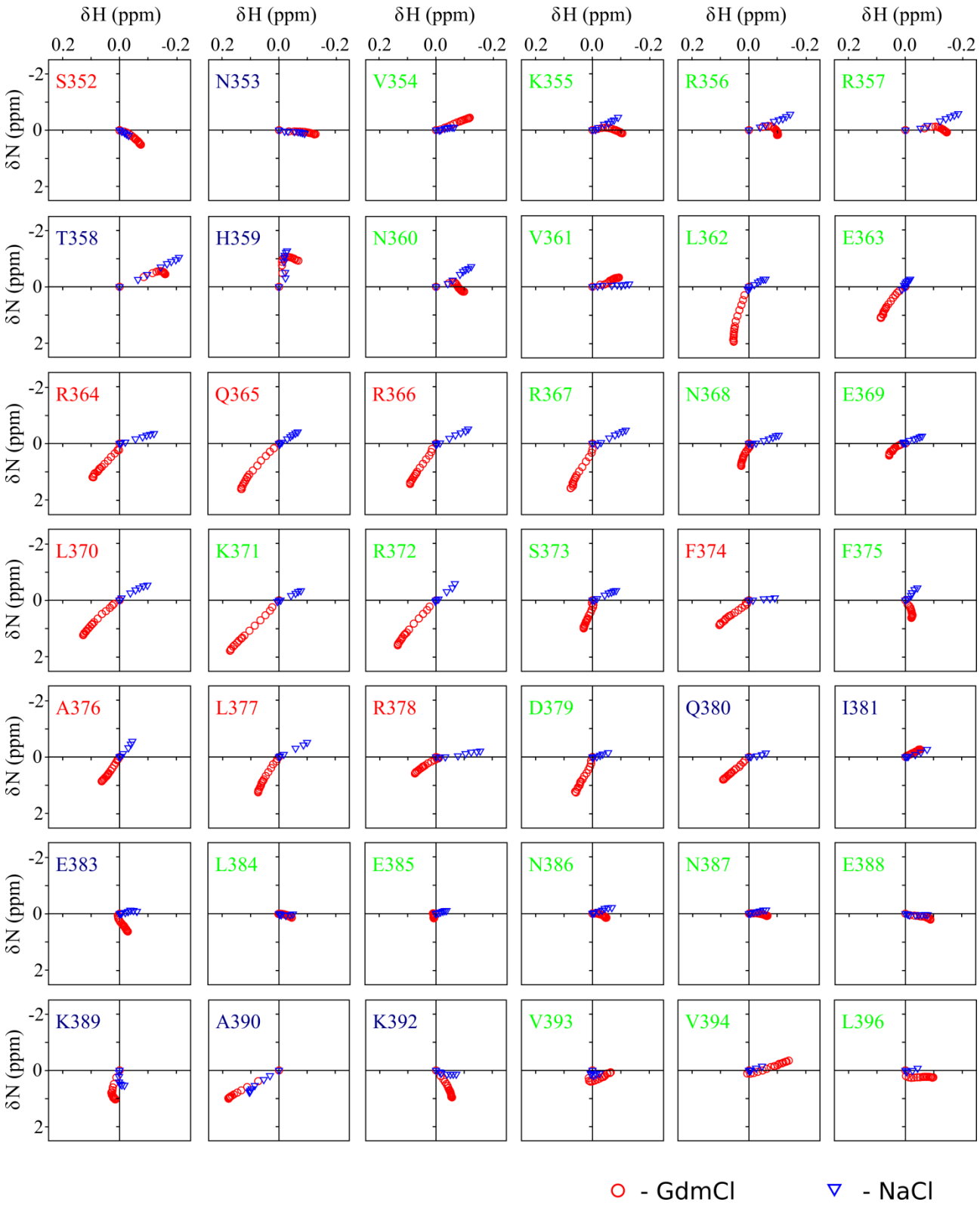


Figure S1: (relates to **Fig. 2**) The behaviour of observable ^{1}H, ^{15}N correlation peaks in HSQC spectra of Myc over a GdmCl titration and a NaCl titration. GdmCl concentration range (red circles): 0.0 to 3.2 M, with 0.2 M steps. NaCl concentration range (blue triangles): 0.0 to 1.2 M, with 0.2 M steps. Residue names are coloured as in Fig. 1.

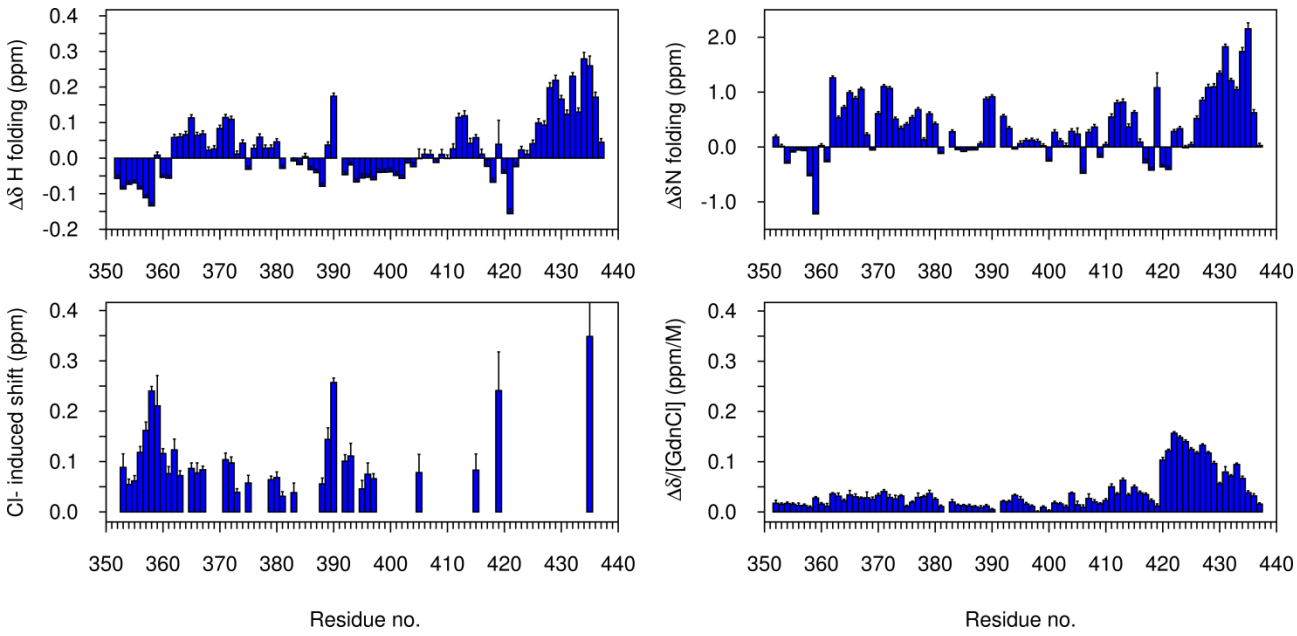


Figure S2 Relates to **Fig 2**. Values of best fit residue-specific parameters for proton (top-left) and nitrogen (top-right) chemical shift for the sigmoidal transition ($\Delta\delta H_{folding}$ and $\Delta\delta N_{folding}$), the Cl^- induced hyperbolic transition (expressed as $((\Delta\delta H_{Cl})^2 + (\Delta\delta N_{Cl}/6)^2)^{0.5}$, bottom-left) and the slope of final shallow transition (bottom-right).

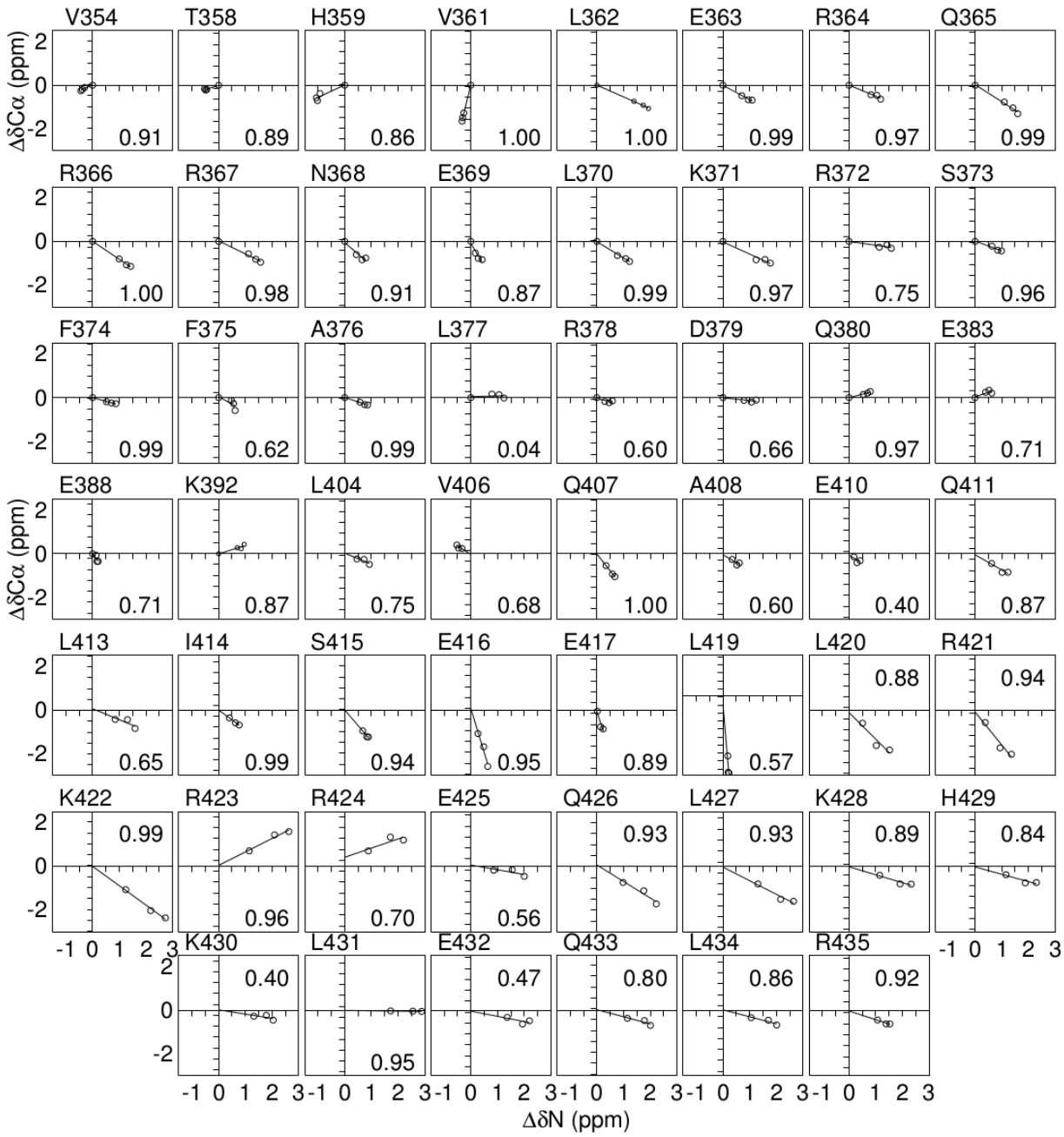


Figure S3: relates to **Fig. 3**, δC_α plotted against $\Delta\delta N$ for representative residues. Correlation coefficients are shown inset for each residue.

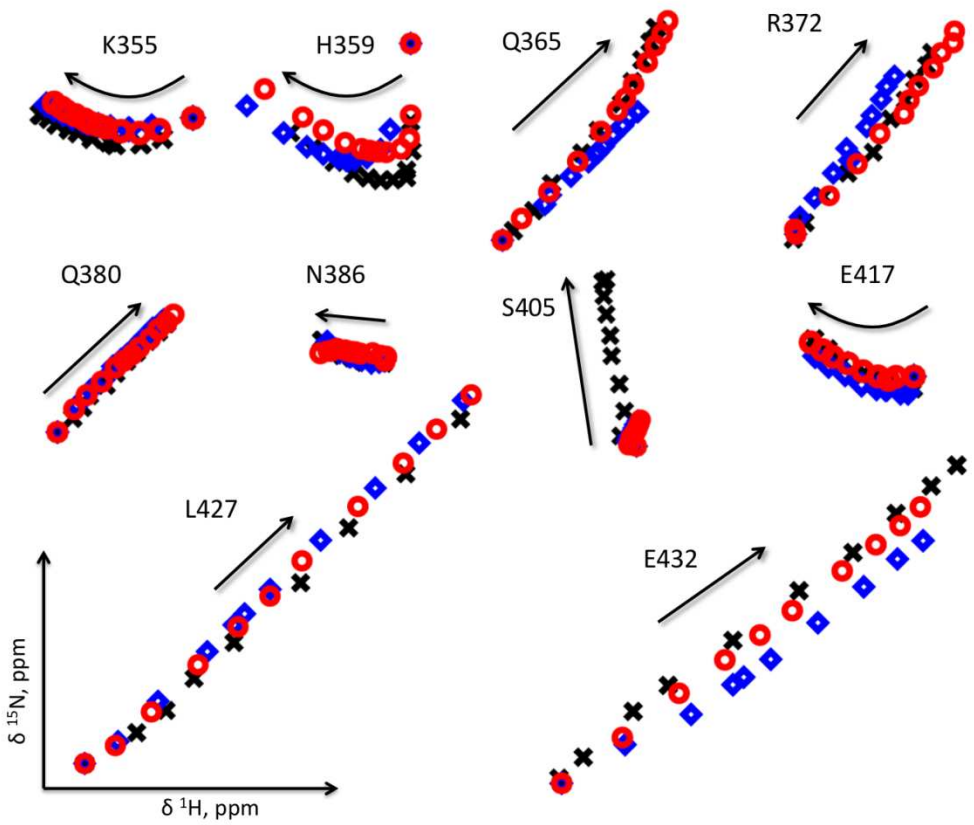


Figure S4: relates to **Fig. 4**. Denaturation profiles mapped as cross-peaks from ^{1}H-^{15}N HSQC spectra for the different Myc cysteine mutants: Q365C (blue diamonds), N386 (red circles), S405 (black crosses). The radii of the circles (dimensions of crosses and squares) correspond to 0.015 ppm in the ^{1}H dimension HSQC spectra. Arrows point in the direction of the cross-peak motion upon increasing GdmCl concentration.

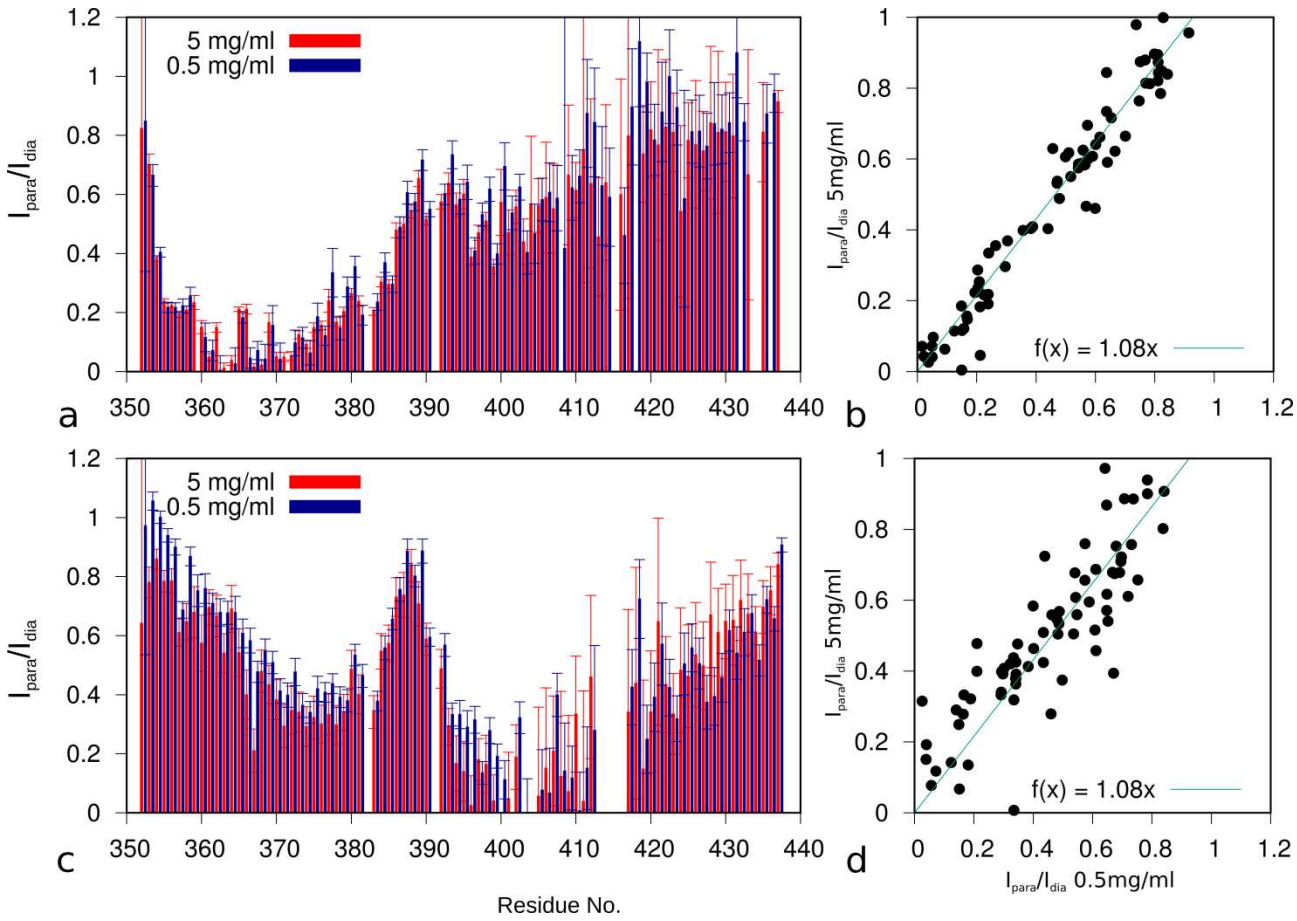


Figure S5: relates to **Fig 5**. Sequence distribution of PREs (I_{para}/I_{dia}) at 0.5 (blue) and 5 (red) mg/ml at 0.6 M GdmCl, for Q365C (A) and S405C (C) spin-labelled variants. PRE ratios of 0.5 mg/ml vs 5 mg/ml were plotted to illustrate correlation between them (B and D) and regression coefficient (slope value).

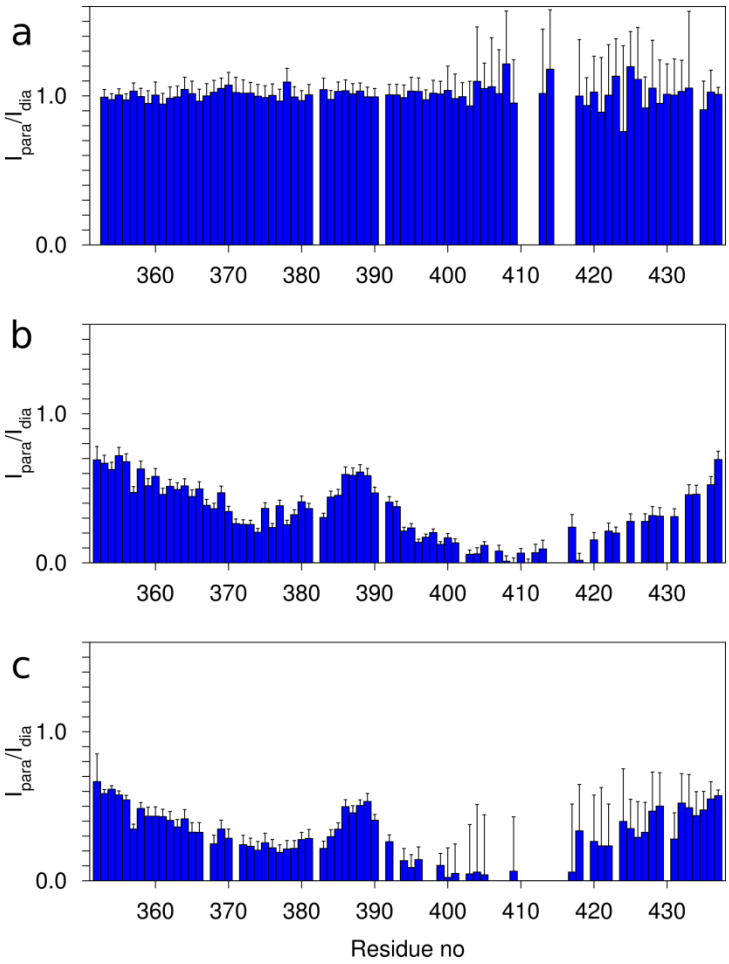


Figure S6: Relates to **Fig. 5.** Intensity ratios of samples with paramagnetic MTSL to diamagnetic MTSL when **a)** ^{14}N -Myc Q411-MTSL is in 1:1 mix with underivatised ^{15}N Myc Q411 at 4 mg/ml **b)** ^{15}N -Myc Q411-MTSL at the same concentration. **c)** ^{15}N -Myc S405-MTSL at the same concentration

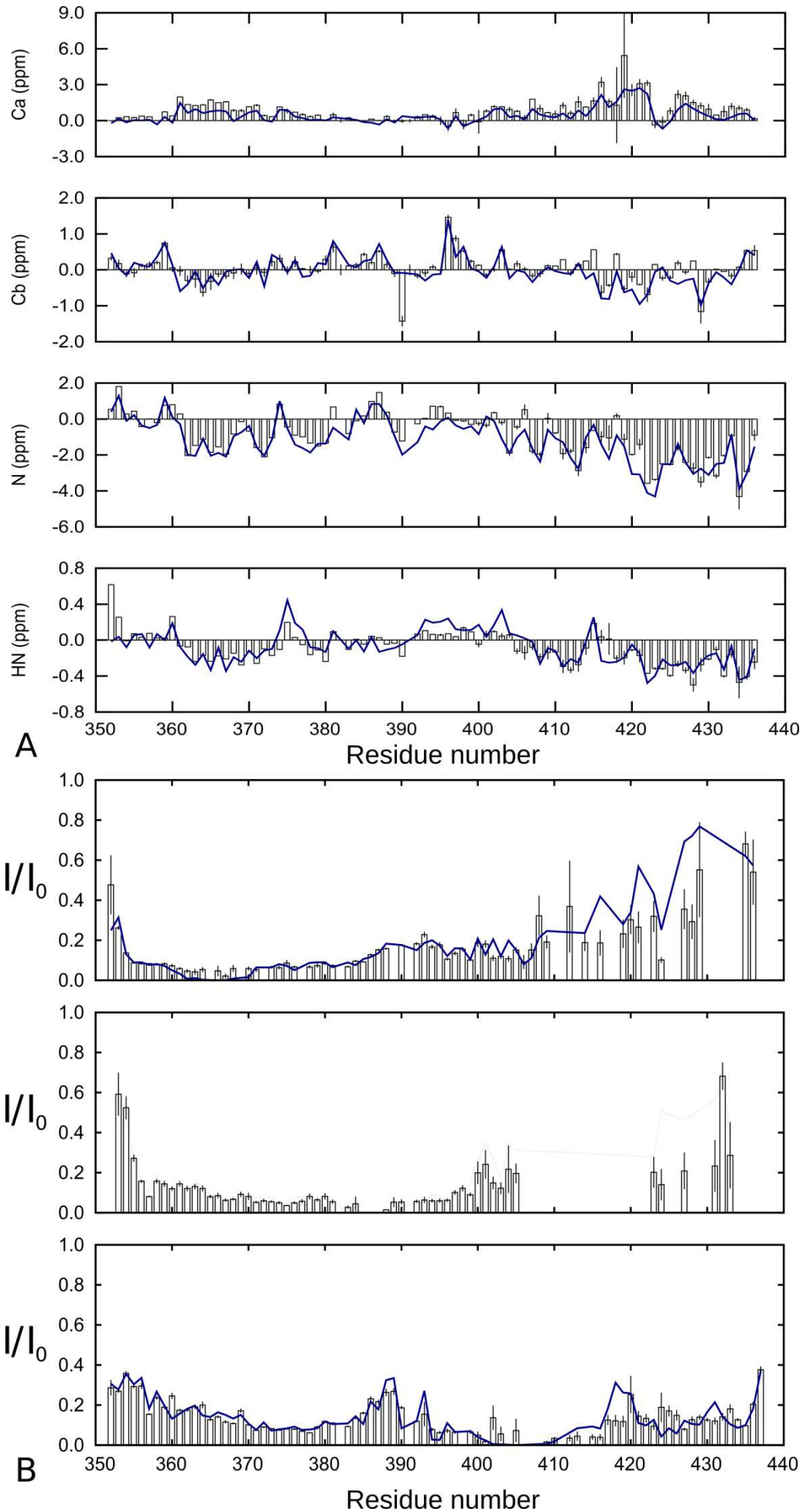

Ca (ppm)
Cb (ppm)
N (ppm)
HN (ppm)
A
Residue number
I/I0
I/I0
I/I0
B
Residue number

Figure S7: Relates to **Fig. 6 & 7**. Comparison of the experimental (empty bars) and ensemble averaged data (blue line) for Myc at 0 M GdmCl. (A) Chemical shift deviation from random coil values back-calculated by SPARTA; (B) Intensity ratios between paramagnetic and diamagnetic samples on a per residue level.