Just a flexible linker? The structural and dynamic properties of CBP-ID4 revealed by NMR spectroscopy

Alessandro Piai[1], Eduardo O. Calçada[1], Thomas Tarenzi[1], Alessandro del Grande[1], Mihaly Varadi[2], Peter Tompa[2,3], Isabella C. Felli[1*], Roberta Pierattelli[1*]

[1] Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, Florence, Italy;
[2] VIB Structural Biology Research Center (SBRC), Vlaams Instituut voor Biotechnologie and Structural Biology Brussels (SBB), Vrije Universiteit Brussel, Belgium;
[3] Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary.

***Mailing address**
Magnetic Resonance Center
University of Florence
Via Luigi Sacconi 6
50019 Sesto Fiorentino
Italy

Email: felli@cerm.unifi.it
Email: pierattelli@cerm.unifi.it

Running title

The fourth intrinsically disordered region of CBP

Keywords

Intrinsically disordered proteins; IDPs; CREB-binding protein; CBP; prolines

Abstract

Here, we present the structural and dynamic description of CBP-ID4 at atomic resolution. ID4 is the fourth intrinsically disordered linker of the CREB-Binding Protein (CBP). In spite of the largely disordered nature of CBP-ID4, NMR chemical shifts and relaxation measurements show a significant degree of α-helix sampling in the protein regions encompassing residues 2-25 and 101-128 (1852-1875 and 1951-1978 in full-length CBP). Proline residues are uniformly distributed along the polypeptide, except for the two α-helical regions, indicating that they play an active role in modulating the structural features of this CBP fragment. The two helical regions are lacking known functional motifs, suggesting that they represent thus-far uncharacterized functional modules of CBP. The present work provides novel insights regarding the functions of this protein linker, which may exploit its plasticity to modulate the relative orientations of neighboring folded domains of CBP and fine tune its interactions with a multitude of partners.

Introduction

Large proteins are often composed by several folded domains separated in the primary sequence by flexible linkers. While folded domains are generally well characterized at atomic resolution, only little information is available for linkers. These protein segments seldom crystallize, and they are often quite challenging to be characterized by nuclear magnetic resonance (NMR) spectroscopy because of their peculiar amino acid composition and their structural and dynamic properties. However, it is unlikely that linkers play passive connecting roles only, because they often constitute as much as half of the primary sequence of complex proteins. In fact, their amino acid sequences often show functional features such as interaction motifs, post-translational modification sites and conservation in the primary sequence across species (1, 2). To learn more on the role of these protein elements it is important to accomplish their experimental characterization, exploiting recently developed NMR tools (3-14). With this in mind we decided to focus on a well-studied key protein in biological systems, the transcription factor CREB-Binding Protein (CBP).

CBP and its paralog p300 are transcriptional co-regulators that integrate signals from numerous signal transduction pathways, and play critical roles in basic cellular processes ranging from development and differentiation to DNA repair (15). Their biological function is related to their ability to interact with a large number of proteins through multiple protein-interaction domains as well as to their acetyl-transferase activity. There are seven autonomous folded domains in CBP/p300, the 3D structures of which have been determined in recent years either by X-ray crystallography or NMR (16-24). Four of them require zinc(II) ions in order to adopt a stable fold: the transcriptional-adaptor zinc-finger-1 domain (TAZ1), the plant homeodomain (PHD), the zinc-binding domain near the dystrophin WW domain (ZZ) and the transcriptional-adaptor zinc-finger-2 domain (TAZ2); other folded domains are the KID-binding domain (KIX), the bromodomain and the histone acetyl-transferase domain (HAT). Finally, a domain named nuclear-receptor coactivator-binding domain (NCBD) is intrinsically disordered, but folds upon binding to its partner(s).

Regions between CBP folded domains represent more than 50% of the total 2,442 residues of the protein; these are predicted to be mostly intrinsically disordered (25). In fact, the folded domains listed above are spatially separated by five linkers of different length, denoted as CBP-ID# (where ID stands for *Intrinsically Disordered* and # represents the number of the linker), which have not been yet characterized at atomic resolution. Very little is known about their structural and functional roles, which may range from maintaining a specific distance between the various folded domains to fine-tuning and modulation of interaction processes. The second hypothesis is supported by the fact that the primary sequence of CBP linkers is well conserved in evolution.

The amino acid composition of CBP linkers is biased towards disorder-promoting amino acids, as typically found in the case of intrinsically disordered proteins (IDPs) (26-28). Overall, they contain a high proportion of prolines (16%), glutamines (15%), serines (11%), glycines (9%) and alanines (9%). In particular, CBP-ID4 and CBP-ID5 exhibit the most distinctive amino acid composition, since almost 40% of their primary sequences is constituted by prolines and glutamines.

In this work we present the characterization by NMR spectroscopy of CBP-ID4, the 207 residues long linker (CBP residues 1851-2057) located between the TAZ2 and NCBD domains. Among all CBP linkers, ID4 is the one containing the highest percentage of proline residues

(22%) and it is predicted to be intrinsically disordered, two aspects that make it a challenging target for NMR. To obtain its sequence specific assignment we thus used a strategy tailored for IDPs, combining ^{1}H and ^{13}C detection and multidimensional experiments (14, 29, 30 and references therein), which resulted in the complete NMR characterization of this linker.

We believe the findings described in the present work will open new frontiers towards a deeper understanding of the roles of CBP linkers and represent the demonstration that NMR can provide a wealth of information on IDPs difficult to access with other techniques.

Materials and Methods

Protein expression and purification

The recombinant vector pET21a-CBP-ID4 containing the human CBP-ID4 native gene was transformed into *E. Coli* BL21 (DE3) for protein expression. For ^{15}N- and ^{13}C,^{15}N-labeled protein production, a colony from a freshly transformed plate was selected to inoculate in LB medium (50 ml) containing ampicillin and was grown overnight at 37°C and 180 rpm. Cells from the overnight LB growths were diluted 1:100 into 4 l of LB medium and grown at 37°C with constant agitation at 160 rpm. Once an OD600 of 0.7–0.8 was reached, the cells were gently centrifuged at 4500 rpm for 20 min. The pellet was resuspended in 1 l of minimal medium (48.5 mM Na_2HPO_4, 22.0 mM K_2HPO_4, 8.5 mM NaCl, 0.2 mM $CaCl_2$, 2.0 mM $MgSO_4$, 1 mg l^{-1} each of biotin and thiamin, 7.5 mM $(^{15}NH_4)_2SO_4$ and 11.1 mM glucose/$^{13}C_6$-glucose) and grown at 37 °C for 1 hour with constant agitation at 160 rpm, following the Marley's method (31). The cells were induced with 0.5 mM IPTG and allowed to grow for additional 4 hours at 30°C. The culture was then harvested at 8000 rpm for 20 min and the pellet was stored at -20°C. Frozen cells were thawed and suspended in 45 ml of equilibration buffer A (50 mM MES, 10 mM EDTA, 20 mM NaCl, pH 5.5). Cells were disrupted by sonication on ice (at 80% sonication power) with cycles of 6 s with 4 s delay pulses for 25 min. Lysed cells were centrifuged at 40000 rpm for 40 min at 4°C and the supernatant (≈ 40 ml) was warmed at 80°C for 20 min in order to remove contaminants. This solution was spun at 8000 rpm for 20 min at 4°C and the supernatant was filtered through a 0.22 µm pore membrane in preparation for the purification steps. Cation exchange chromatography was performed on a 6 ml Resource S column (GE Healthcare) pre-equilibrated with buffer A. A linear gradient between buffer A and buffer B (buffer A with 0.5 M NaCl) was applied over 20 min at a flow rate of 2.5 ml/min. The CBP-ID4 was eluted with a salt concentration around 300 mM. Fractions were analyzed by SDS–PAGE and those containing the target protein were concentrated until 2 ml. The sample was further purified by gel filtration on Hiload 16100 superdex 75 (GE Healthcare) in 20 mM potassium phosphate buffer at pH 6.5, 100 mM NaCl, 50 µM EDTA. The fractions containing pure CBP-ID4 were concentrated up to 2 ml and stored at 4°C for biophysical analysis. Mass spectrometry (MALDI) was performed to confirm the molecular mass of purified CBP-ID4.

NMR samples

All the multidimensional NMR experiments for sequence-specific assignment were performed on a sample of 0.6 mM uniformly ^{13}C, ^{15}N labeled human CBP-ID4 in 20 mM potassium phosphate buffer, 100 mM KCl at pH 6.5, with 10% D_2O added for the lock. ^{15}N relaxation experiments were acquired on a 0.5 mM uniformly ^{15}N labeled human CBP-ID4, in the same experimental conditions. For all the experiments, 3 mm NMR sample tubes were used to reduce the detrimental effects of high salt concentration.

NMR data acquisition

A set of multidimensional $^{1}H^{N}$ and ^{13}C detected NMR experiments tailored to achieve sequence-specific assignment of IDPs was acquired at 283.0 K. 2D CON-IPAP (32, 33), 3D (H)CBCACON-IPAP (34), 3D (H)CBCANCO-IPAP (34), 4D (HCA)CON(CA)CON-IPAP (35) and 4D (HN)CON(CA)CON-IPAP (35) experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. 2D BEST-TROSY (BT) (11, 36), 3D BT-HNCO (11, 37, 38), 3D BT-HN(CA)CO (11, 38), 3D BT-HNCACB (11, 38), 3D BT-HN(CO)CACB (11, 38), 3D BT-(H)N(COCA)NH (11) and 3D BT (H)N(CA)NNH (39) experiments were performed at 21.1 T on a Bruker Avance spectrometer operating at 898.57 MHz ^{1}H, 225.95 MHz ^{13}C and 91.05 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. 5D BT-(H)NCO(CAN)CONH (40) and 5D BT-HN(COCAN)CONH (40) experiments were collected at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

Heteronuclear ^{15}N relaxation experiments (41-44) for the measurement of ^{15}N R_1, ^{15}N R_2 and ^{15}N-^{1}H NOEs, (CLEANEX-PM)-FHSQC experiments (45) for the estimation of the extent of amide proton exchange with the solvent and the 3D HNHA experiment (46) for the determination of homonuclear $^{3}J_{HN\text{-}H\alpha}$ were measured at 283.0 K at 16.4 T on a Bruker Avance spectrometer operating at 700.13 MHz ^{1}H, 176.05 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

The temperature dependence of CBP-ID4 was investigated by acquiring a series of 2D BEST-TROSY and 2D CON-IPAP spectra in the range 238.0-308.0 K, with steps of 5 degrees, at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

The parameters used for the acquisition of all the experiments are reported in the Supporting Material (Tables S1-S5). All the data sets were acquired using *Bruker TopSpin 1.3* or *3.1* software. 3/4/5D experiments for the sequence-specific assignment were performed using on-grid non-uniform sampling (NUS). The on-grid ''Poisson disk'' sampling scheme (47) was chosen to generate the time schedules with the *RSPack* program [J. Stanek, A. Zawadzka-Kazimierczuk, unpublished]. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with σ=0.5.

NMR data processing and analysis

Conventionally sampled NMR data sets were processed using *Bruker TopSpin 1.3* software. Instead, when NUS was employed, the NMR data were converted with *nmrPipe* (48) and then processed using either the Multidimensional Fourier Transform (MFT) algorithm (for 3D data sets) or the Sparse MFT (SMFT) algorithm (for 4/5D data sets), respectively implemented in *ToASTD* (49) and *reduced* (50, 51) programs. Both programs are available at http://nmr.cent3.uw.edu.pl.

CARA (52) and *Sparky* (53) were used to analyze 3D and 4/5D spectra, respectively, whereas *CcpNmr Analysis* (54) was employed to analyze ^{15}N relaxation data.

The secondary structure propensity from the heteronuclear chemical shifts was determined by using the neighbor corrected structural propensity calculator (ncSPC) tool (55), available online at http://nmr.chem.rug.nl. The Tamiola, Acar and Mulder random coil chemical shift library (56) was chosen for the analysis.

Theoretical helical propensities were calculated using the Agadir algorithm (57-61), available online at http://agadir.crg.es. For the calculation the pH was set to 6.5, the temperature to 283.0 K and the ionic strength to 0.1 M.

Intrinsic protein disorder was predicted by using IUPred (62) and PONDR-FIT (63) tools, available at http://iupred.enzim.hu and http://www.disprot.org.

For the conservation analysis, a comprehensive data set of 39 experimentally validated homologous amino acid sequences was assembled by performing a PSI-BLAST (64) against the non-redundant protein data set of NCBI. The query sequence was the ID4 segment of the human CBP protein (residues 1852-2057) [UniProt ID: Q92793], and the retrieved sequences covered all major vertebrate taxonomic groups from mammals to fish. The multiple sequence alignment was generated using MAFFT (65) and then used as input for the sequence- and disorder conservation analysis. We used the local version of the DisCons tool to analyze the multiple sequence alignment (66). DisCons was used with default parameters, namely predicting disorder with IUPred, and quantifying the sequence conservation with Jensen-Shannon divergence; the maximum allowed fraction of gaps for a position was set to 0.6.

The presence of linear motifs and functional sites was investigated by using the Eukaryotic Linear Motif (ELM) computational biology resource (67), available online at http://elm.eu.org.

Potential serine, threonine and tyrosine phosphorylation sites were predicted exploiting NetPhos 2.0 (68), available online at http://www.cbs.dtu.dk. Potential kinase specific phosphorylation sites were predicted by using using NetPhosK 1.0 (69), available online at http://www.cbs.dtu.dk.

To obtain the ensemble description of the protein linker under investigation 10,000 conformers were generated using the program *flexible-meccano* (70), available online at http://www.ibs.fr. The SSP score calculated from experimentally measured chemical shifts was used as input in the calculation to allow also the generation of conformers whose backbone dihedral angles deviate substantially from those typically found in pure random coils. The ensemble and the underlying data were deposited in the Protein Ensemble Database (http://pedb.vib.be, 71) under the accession ID PED5AAE.

Results and Discussion

NMR sequence-specific resonances assignment

The 2D BEST-TROSY and 2D CON-IPAP fingerprint spectra of CBP-ID4 are reported in Figure 1. The ^{1}H-^{15}N correlation spectrum exhibits all the peculiar NMR features of an IDP, namely the reduced chemical shift dispersion particularly pronounced in the ^{1}H dimension, and the high degree of signal crowding. The 2D CON-IPAP spectrum benefits from the larger spectral width in the ^{15}N dimension, which includes signals due to proline residues nitrogen nuclei, and the fact that the experiment provides inter-residues correlations, which are more resolved with respect to intra-residue ones (72). A series of triple-resonance 3/4/5D NMR experiments (see Materials and Methods Section) specifically designed for IDPs, exploiting

either $^{1}H^{N}$ or ^{13}C detection, was acquired. Non-uniform sampling (NUS) (73) was exploited to reduce the experimental time while preserving high spectral resolution in the indirect dimensions (74).

The ^{13}C-detected 3D and 4D spectra (34, 35) provided the key information to achieve the sequence-specific assignment of this proline-rich IDP. The 3D BEST-TROSY (BT) triple-resonance experiments (11) were acquired to obtain the complementary information needed to complete the assignment, further supported by the 5D BT-(H)NCO(CAN)CONH and 5D BT-HN(COCAN)CONH experiments (40), used to resolve a few ambiguities and to confirm the chemical shift assignment obtained with the 3D spectra.

This set of spectra tailored for the study of IDPs made it possible to assign the resonances belonging to all H^{N}, N, C′, C^{α} and C^{β} nuclei. The chemical shift assignment of CBP-ID4, both at 283.0 and 308.0 K, is deposited in the *BioMagResBank* (BMRB, http://www.bmrb.wisc.edu, (75), entries 26616 and 26639.

NMR characterization

The presence of secondary structure propensity within CBP-ID4 was investigated by comparing experimentally measured heteronuclear chemical shifts (N, C′, C^{α} and C^{β} nuclei) to the corresponding random coil values (55). The obtained differences (Figure 2) are distributed around zero, consistently with the disordered nature of CBP-ID4. However, two regions, encompassing residues 2-25 and 101-128, exhibit significantly positive values suggesting a high propensity to form α-helices in these regions.

The $^{3}J_{HN\text{-}H\alpha}$ coupling constants (46) shown in Figure 3, provide complementary information to detect contingent residual structural propensity along the polypeptide chain, and are therefore often used in structural characterizations of chemically unfolded or intrinsically disordered proteins (76-80). The obtained average value of about 7 Hz is similar to values obtained for flexible proteins and other disordered peptides (79, 80). Only two protein regions (residues 13-16 and 106-118) substantially deviate from this trend, corresponding to those of highest predicted α-helical propensity. Therefore, conformations with values for the angle ϕ towards 60° are increasingly being populated in these protein regions. Interestingly, also residues 92-99 show significantly lower values, suggesting that a certain amount of structural organization may be present there too, as revealed by the SSP analysis (Figure 2). From $^{3}J_{HN\text{-}H\alpha}$ values alone it is not possible to conclude whether this include canonical α-helical or polyproline type II (PPII) structures (81, 82).

Backbone dynamics of CBP-ID4 was investigated through the measurement of ^{15}N R_1, ^{15}N R_2 relaxation rates and heteronuclear ^{15}N-^{1}H NOE values. Relaxation parameters, determined for the majority of the assigned cross-peaks in the 2D ^{1}H-^{15}N HSQC spectrum, are reported in Figure 4. Considerably higher ^{15}N R_2 and ^{15}N-^{1}H NOE values are observed in the regions encompassing residues 2-25 and 101-128, confirming a reduced flexibility in these regions likely related to the observed α-helical propensity (Figure 2). Interestingly, the magnitude of the ^{15}N R_2 values indicates that the transient α-helix located between residues 101-128 is more rigid than that comprising residues 2-25. ^{15}N-^{1}H NOE values reveal that the fragments separating the two partially populated α-helices are highly flexible, with the first fragment (residues 26-103) more flexible than the second one (residues 129-207).

Local solvent accessibility was investigated by monitoring H_2O exchange processes for backbone amide protons (45). The results, reported in Figure 5, show that exchange effects are overall very pronounced, becoming operative for the majority of the residues already with 16 ms of mixing time (τ_m). Notably, the two protein regions exhibiting significant α-helical propensity, in particular the one encompassing residues 101-128, which is more rigid by ^{15}N R_2 measurements, appears to exchange less with the solvent; on the contrary, the other residues of the polypeptide chain are characterized by high-solvent exchange effect.

The temperature dependence of CBP-ID4 was investigated by acquiring a series of 2D BEST-TROSY and 2D CON-IPAP spectra in the range 283.0-308.0 K (shown in Figures S2 and S3 in the Supporting Material). These spectra enabled the transfer of the resonances assignment obtained at 283.0 K to the other temperatures by following the chemical shift changes of the cross-peaks with temperature. A 3D CBCACON experiment was acquired to confirm the signals assignment at 308.0 K.

In Figure 6a the difference of cross-peak intensities in the 2D BEST-TROSY NMR spectra acquired at 308.0 and 283.0 K is reported. The majority of cross-peaks become weaker or broadened with increasing temperature as a consequence of the higher rate of amide proton exchange with the solvent protons. Instead, the signals belonging to residues 101-128, those exhibiting a significant propensity to form α-helix, become more intense at high temperature. This evidence suggests that this protein region results less affected by chemical exchange processes and benefits from an improved correlation time resulting from the faster rotational tumbling. Interestingly, an opposite behavior is observed in the other protein region characterized by high α-helix propensity (residues 2-25): it is destabilized by increasing temperature and experiences a decrease in signal intensities similar to that of the other unstructured regions of the protein.

Correlations between amide proton temperature coefficients ($\Delta\delta H^N/\Delta T$) and hydrogen bonds (83, 84) were also investigated (Figure 6b). Less negative temperature coefficients are clustered in the two protein regions exhibiting higher propensity to form α-helix, in agreement with the results of the other analysis. In particular, in the case of the transient α-helix encompassing residues 101-128, the great majority of amide proton temperature coefficients result less negative than -4.6 ppb/K, a value considered the threshold below which the presence of intramolecular hydrogen bonds can be excluded with a predictive value of 85% (84). Therefore, transient intra-molecular hydrogen bonds are likely to be present in this protein region, which again appears more structured than the region comprising residues 2-25.

Structural and dynamic properties of CPB-ID4

The overall picture resulting from the high resolution experimental characterization of the structural and dynamic properties of CBP-ID4 reveals two well-defined partially populated helical fragments, termed helix I and helix II, which display properties significantly different from one another, separated by segments characterized by high local flexibility. It is interesting to note that prolines, the most abundant amino acid in CBP-ID4 (22%), are distributed quite uniformly throughout the protein primary sequence (Figure 1a) with the exception of the two parts of the protein exhibiting α-helical propensity (Figure 2). Proline residues may be selected against in the α-helical segments because of their helix-breaking properties: their peculiar side chain disrupts the H-bond network and interferes with the backbone helical packing of adjacent residues. Furthermore, it can be noted that 7 proline

residues are clustered in the region immediately preceding the inner transient helix II, suggesting that they are involved in its stabilization. In fact, it has been reported that prolines present in proximity of the so-called pre-structured motifs (PreSMos) (85) of IDPs are very important in promoting and delimiting the PreSMos themselves (86). Therefore, the prolines flanking helix II (residues 101-128) seem to play a specific structural role (87). This would also explain why this partially structured protein region appears to be more stable than that comprising residues 2-25, which in the full-length protein is preceded by a structured domain but is here unrestrained.

The degree of α-helical propensity could have an important role in providing specific patterns for partner recognition through a specific surface provided by the helix itself, as shown in the literature in several different examples (25, 88, 89). On the other hand, the transition from a random-coil, extended conformation of the protein backbone to partially formed helices represents the more drastic local change for backbone dihedral angles, with the helical conformation constituting the most compact one (1.5 Å shortening for each amino acid part of the helix with respect to an elongated conformation). The presence of the secondary structure element can also reduce the degree of freedom of the polypeptide, which cannot freely bend to form compact structures. Therefore, the extent of local helical propensity could be a way to modulate the overall length of a specific polypeptide fragment with local cooperative changes in backbone dihedral angles. In this way, the insertion of a partially populated helical segment in between largely random coil backbone fragments, such as observed for the case of CBP-ID4, could provide a mean to modulate the overall length and orientation of the linker itself.

To visualize the possible conformers populating the ensemble of structures in dynamic equilibrium in solution, the secondary structure propensities determined experimentally for CBP-ID4 were used as input to calculate an ensemble of conformers through the program *Flexible-Meccano* (70), using the statistical coil potential proposed for IDPs. The computed ensemble shows a great variability in the overall radius of gyration (Figure S4 in the Supplementary Material), in agreement with the idea that flexible linkers can indeed fine-modulate the relative distance between folded units (PED ID PED5AAE). The obtained structures can be inspected locally to evaluate the properties of the partially populated helical conformers. In the present case, the two regions characterized by secondary structural propensities feature quite different composition in terms of the constituent amino acids, which has an impact on the properties of the partially populated helices. In particular, helix I (2-25) is characterized by the presence of several charged amino acids, resulting in an overall positive electrostatic potential (Figure 7e), with a quite peculiar methionine-rich patch on one side (3 methionine residues quite close to each other on the helix surface, Figure 7d). Instead, helix II (101-128) has an overall hydrophobic surface (Figure 7i) with some charged residues (three positive arginine residues and three negative glutamic acids residues, Figure 7h). Therefore, the two partially populated helical elements provide significantly different modules for protein-protein interactions. Both helices share a large number of polar amino acids, with a predominance of glutamine residues, which may promote the stability of these two segments.

Although CBP-ID4 has never been characterized in detail before, the crystal structure of the TAZ2 domain of human p300, paralogue of CBP, was solved earlier (19). The construct used in that study, in addition to the TAZ2 domain, included at its C-terminus 21 residues of the subsequent linker, which is homologous to the initial part of CBP-ID4 (16/21 residues are conserved in the two protein sequences). Specifically, these residues correspond to residues

1-21 of CBP-ID4, identified as exhibiting significant α-helical propensity. Interestingly, in the crystal structure these residues assume a α-helix conformation, which extends outside the globular structure of TAZ2 domain. Furthermore, it has been reported that such α-helix contributes actively to the binding of p300 to the transcription factor Myocyte Enhancer Factor 2 (MEF2) (90). This observation fits with the general notion that IDPs often function by binding to partner molecules via structural elements that are dynamically sampled in the disordered ensemble (91). Such motives are conserved structural/functional elements of IDPs, and their presence can be concluded from the structural propensity of certain regions of the IDP chain.

Prediction of functional regions within CBP-ID4

It is interesting to compare the picture resulting from the experimental characterization of CBP-ID4 with all the information that can be predicted on the basis of the primary sequence, starting from the local propensities for local order/disorder, the local secondary structure, all the way to interaction motifs or post-translational modification sites.

The significant local propensity for secondary structure that was observed in the ensemble model can hint at the presence of binding motifs within the helical regions, as often seen in IDPs (95, 96). Furthermore, the potential functional importance of the two helical regions should also manifest themselves in their sequence conservation. Thus, we have collected and aligned 39 CBP sequences and used the DisCons online tool to quantify the conservation of the amino acid sequence and of the structural disorder (66) (Figure 8 and Figure S5). The position-specific conservation scores of the amino acid sequence of regions 2-25 and 101-128 are consistently high, and in fact higher than the conservation of disorder, which is suggestive of their functional importance and might indicate that these segments act as molecular recognition features, which are known to have higher sequence conservation and lower disorder with respect to their flanking regions (66).

We have used the Eukaryotic Linear Motif (ELM) database (67) to identify the presence of Short Linear Motifs (SLiMs) in the sequence, since SLiMs are known to be involved in recognition and targeting processes (97). By definition, they are highly specific short structural/functional elements of 3-10 amino acids, mostly located in intrinsically disordered regions of proteins, which are able to mediate protein-protein interactions without the need of a stable three-dimensional structure. Remarkably, more than 95% of the SLiMs identified by ELM are located in the most disordered regions of CBP-ID4 (Figure S6a in the Supplementary Material).

The position of potential serine, threonine and tyrosine phosphorylation sites of CBP-ID4 were also identified using NetPhos 2.0 (68), an artificial neuronal network-based method for predicting phosphorylation sites. The importance of post-translational modification (PTM) sites derives from their frequent involvement in the modulation of protein functions (98-101). The location of the recognized sites (Figure S6b in the Supplementary Material), similarly to SLiMs, are distributed exclusively along the most flexible regions of the protein, and are excluded from the two α-helical segments. In particular, several identified phosphorylation sites are located in the region 26-103 which, by ^{15}N-^{1}H NOE analysis, is the most flexible region of the entire protein (Figure 4c). Therefore, we can infer that the negligible secondary structure content of this protein region may facilitate the interaction between the phosphorylation sites and the corresponding phosphotransferases (102).

Based on the characterization of CBP-ID4 presented here, which shows that two parts of the polypeptide chain have an intrinsic propensity to adopt a helical conformation in solution, one might speculate on the active role played by this protein regions in the communication between different parts of complex proteins or between different partners. It is feasible that both transient α-helices identified in CBP-ID4 may promote the interactions between CBP and its partners. In addition, the highly flexible parts linking the two partially populated helices may provide complementary functional advantages through their largely exposed and flexible protein backbones.

Conclusions

The NMR description of CBP-ID4 presented here reveals very heterogeneous structural and dynamic properties of this "linker", which combines a great extent of structural disorder together with two protein regions characterized by significant α-helical propensities. These findings lead to a reexamination of the concept of protein linker, traditionally considered as mere connection between folded domains, which instead may modulate and fine-tune protein function through an heterogeneous structural disorder. This study shows that the recently developed NMR methods allow to overcome potential limitations deriving from the peculiar properties of linkers (high proline content, extensive spectral overlap, fast amide protons exchange) and opens new possibilities for the characterization at atomic level of PTMs and of intra- and inter-molecular interactions of CBP-ID4, starting with the partners known to interact with adjacent folded domains of CBP.

Authors contribution

EOC, PT, ICF and RP conceived the research; ICF and RP directed the research; EOC and AdG expressed the protein and prepared the samples; AP and TT performed the NMR experiments and analyzed the data; MV and PT performed the bioinformatics analysis; AP, MV, PT, ICF and RP wrote the paper. All Authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

This work has been supported in part by the European Commission Projects IDPbyNMR (Contract No. 264257), BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252). Simone Kosol and Leonardo Gonnelli are gratefully acknowledged for the stimulating discussions in the early stages of the project.

Conflict of interest

The Authors declare that they have no conflicts of interest with the contents of this article.

Captions to the Figures

Figure 1. Amino acid sequence and fingerprint spectra of CBP-ID4. a) primary sequence of CBP-ID4 in which proline residues are highlighted in bold; b) 2D BEST-TROSY and c) 2D CON-IPAP spectra of CBP-ID4. Both spectra are reported with the same Hz/cm ratio in the direct dimension.

Figure 2. Secondary structure propensity score obtained from experimentally measured N, C', C^{α} and C^{β} chemical shifts. Positive and negative values correspond to α-helical and β-sheet propensities, respectively. The two protein regions which exhibit significant propensity for α-helix are colored in dark gray. Proline residues are highlighted along the protein sequence as black triangles on the top and result clustered in the most disordered regions of the protein.

Figure 3. $^{3}J_{HN\text{-}H\alpha}$ coupling constants for the intrinsically disordered protein domain CBP-ID4. Symbols in the two protein regions characterized by high propensity for α-helix are colored in dark gray.

Figure 4. Dynamic characterization of CBP-ID4. Bars in the two protein regions characterized by higher propensity for α-helix are colored in dark gray. a) ^{15}N R_1 relaxation rates; b) ^{15}N R_2 relaxation rates; c) ^{15}N-^{1}H NOEs. All ^{15}N relaxation measurements were carried out at 16.4 T at 283.0 K.

Figure 5. Extent of amide proton exchange with the solvent. The residues belonging to the two protein regions characterized by high propensity for α-helix are colored in dark gray. $I\tau_m/I_{Ref}$ values measured through the (CLEANEX-PM)-FHSQC experiments are reported as a function of residue number for two different mixing-times, namely 16 ms (filled dots) and 40 ms (empty dots). Error bars have been removed to facilitate the comparison and are provided as Supporting Material (Figure S1).

Figure 6. Temperature dependence of CBP-ID4. The values for the two protein regions experimentally identified as exhibiting high propensity to form α-helix are colored in dark gray. a) cross-peak intensity changes in the 2D BEST-TROSY spectrum acquired at 308.0 and 283.0 K; b) temperature coefficients ($\Delta\delta H^N/\Delta T$) reported as a function of residue number.

The black dashed line at -4.6 ppb/K represents the threshold to discriminate between intra- and inter-molecular hydrogen bonds (with a predictive value of 85%).

Figure 7. Representative ensemble of conformers for CBP-ID4. a) Location of the helical segments within the primary sequence; glutamine, arginine and methionine residues are shown in green, blue and gold. b) Subset of conformers obtained superimposing residues of helix I (residues 2-25). All the helical segments are shown in blue. c) Close-up of a set of superimposed helix I with amino acids atoms represented as sticks. d) Helix I with amino acids atoms represented as spheres; glutamine, arginine and methionine residues are shown in green, blue and gold. e) Electrostatic potential on the surface of helix I; positively charged, negatively charged, and neutral amino acids are represented in blue, red, and white. Panels f-i) contain similar information of panels b-e) for helix II (residues 101-128). The backbone models were generated with *flexible-meccano* (70), while side-chains and hydrogen atoms were added by using Molmol 2.0 (92); the energy of the models was minimized using Chimera (93) and the electrostatic potentials were computed by using APBS (94).

Figure 8. Evolutionary conservation of the amino acidic sequence (black) and disorder tendency (gray) of CBP-ID4. The background of the plot corresponding to the two protein regions characterized by high propensity for α-helix is colored in light gray.

References

1. Habchi, J., P. Tompa, S. Longhi, and V. N. Uversky. 2014. Introducing protein intrinsic disorder. Chem. Rev. 114: 6561-6588.
2. van der Lee, R., M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. Chem. Rev. 114: 6589-6631.
3. Hiller, S., C. Wasmer, G. Wider, and K. Wüthrich. 2007. Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J. Am. Chem. Soc. 129: 10823-10828.
4. Mittag, T., and J. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. 17: 3-14.
5. Hsu, S. T., C. W. Bertoncini, and C. M. Dobson. 2009. Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J. Am. Chem. Soc. 131: 7222-7223.
6. Narayanan, R. L., H. N. Duerr, S. Bilbow, J. Biernat, E. Mendelkow, and M. Zweckstetter. 2010. Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. J. Am. Chem. Soc. 132: 11906-11907.
7. Nováček, J., A. Zawadzka-Kazimierczuk, V. Papoušková, L. Židek, H. Sanderová, L. Krásný, W. Koźmiński, and V. Sklenar. 2011. 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J. Biomol. NMR 50: 1-11.
8. Felli, I. C., and R. Pierattelli. 2012. Recent progress in NMR spectroscopy: towards the study of intrinsically disordered proteins of increasing size and complexity. IUBMB Life 64: 473-481.
9. Harbison, N. W., S. Bhattacharya, and D. Eliezer. 2012. Assigning backbone NMR resonances for full length tau isoforms: efficient compromise between manual assignments and reduced dimensionality. Plos ONE 7: e34679.
10. Zawadzka-Kazimierczuk, A., W. Koźmiński, H. Sanderová, and L. Krásný. 2012. High dimensional and high resolution pulse seqeunces for backbone resonance assignment of intrinsically disordered proteins. J. Biomol. NMR 52: 329-337.
11. Solyom, Z., M. Schwarten, L. Geist, R. Konrat, D. Willbold, and B. Brutscher. 2013. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J. Biomol. NMR 55: 311-321.
12. Parigi, G., N. Rezaei-Ghaleh, A. Giachetti, S. Becker, C. Fernandez, M. Blackledge, C. Griesinger, M. Zweckstetter, and C. Luchinat. 2014. Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. J. Am. Chem. Soc. 136: 16201-16209.
13. Jensen, M. R., M. Zweckstetter, J. R. Huang, and M. Blackledge. 2014. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem. Rev. 114: 6632-6660.
14. Konrat, R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. J. Magn. Reson. 241: 74-85.

15. Goodman, R. H., and S. Smolik. 2000. CBP/p300 in cell growth, transformation, and development. Genes Dev. 14: 1553-1577.
16. Legge, G. B., M. A. Martinez-Yamout, D. M. Hambly, T. Trinh, B. M. Lee, H. J. Dyson, and P. E. Wright. 2004. ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. J. Mol. Biol. 343: 1081-1093.
17. De Guzman, R. N., J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. 2005. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. Biochemistry 44: 490-497.
18. Liu, X., L. Wang, K. Zhao, P. R. Thompson, Y. Hwang, R. Marmorstein, and P. A. Cole. 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature 451: 846-850.
19. Miller, M., Z. Dauter, S. Cherry, J. E. Tropea, and A. Wlodawer. 2009. Structure of the Taz2 domain of p300: insights into ligand binding. Acta Crystallogr. D Biol. Crystallogr. 65: 1301-1308.
20. Kjaergaard, M., K. Teilum, and F. M. Poulsen. 2010. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. Proc. Natl. Acad. Sci. U. S. A. 107: 12535-12540.
21. Lee, C. W., M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. 2010. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. Biochemistry 49: 9964-9971.
22. Wang, F., C. B. Marshall, K. Yamamoto, G. Y. Li, G. M. Gasmi-Seabrook, H. Okada, T. W. Mark, and M. Ikura. 2012. Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. Proc. Natl. Acad. Sci. U. S. A. 109: 6078-6083.
23. Filippakopoulos, P., S. Picaud, M. Mangos, T. Keates, J. P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Müller, T. Pawson, A. C. Gingras, C. H. Arrowsmith, and S. Knapp. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149: 214-231.
24. Plotnikov, A. N., S. Yang, T. J. Zhou, E. Rusinova, A. Frasca, and M. M. Zhou. 2014. Structural insights into acetylated-histone H4 recognition by the bromodomain-PHD finger module of human transcriptional coactivator CBP. Structure 22: 353-360.
25. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6: 197-208.
26. Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, and E. C. Garner. 2001. Intrinsically disordered protein. J. Mol. Graph. Model 19: 26-59.
27. Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. Proteins 52: 573-584.
28. Hansen, J. C., X. Lu, E. D. Ross, and R. W. Woody. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. J. Biol. Chem. 281: 1853-1856.
29. Felli, I. C., and R. Pierattelli. 2014. Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J. Magn. Reson. 241: 115-125.

30. Nováček, J., L. Zídek, and V. Sklenar. 2014. Toward optimal-resolution NMR of intrinsically disordered proteins. J. Magn. Reson. 241: 41-52.
31. Marley, J., M. Lu, and C. Bracken. 2001. A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR 20: 71-75.
32. Bermel, W., I. Bertini, L. Duma, L. Emsley, I. C. Felli, R. Pierattelli, and P. R. Vasos. 2005. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew. Chem. Int. Ed. 44: 3089-3092.
33. Bermel, W., I. Bertini, I. C. Felli, and R. Pierattelli. 2009. Speeding up ^{13}C direct detection Biomolecular NMR experiments. J. Am. Chem. Soc. 131: 15339-15345.
34. Bermel, W., I. Bertini, L. Gonnelli, I. C. Felli, W. Koźmiński, A. Piai, R. Pierattelli, and J. Stanek. 2012. Speeding up sequence specific assignment of IDPs. J. Biomol. NMR 53: 293-301.
35. Bermel, W., I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, and A. Zawadzka-Kazimierczuk. 2013. High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J. Biomol. NMR 57: 353-361.
36. Pervushin, K., B. Vogeli, and A. Eletsky. 2002. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. J. Am. Chem. Soc. 124: 12898-12902.
37. Salzmann, M., K. Pervushin, G. Wider, H. Senn, and K. Wüthrich. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc. Natl. Acad. Sci. U. S. A. 95: 13585-13590.
38. Lescop, E., P. Schanda, and B. Brutscher. 2007. A set of BEST triple resonance experiments for time-optimized protein resonance assignment. J. Magn. Reson. 187: 163-169.
39. Weisemann, R., H. Rüterjans, and W. Bermel. 1993. 3D triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins. J. Biomol. NMR 3: 113-120.
40. Piai, A., T. Hošek, L. Gonnelli, A. Zawadzka-Kazimierczuk, W. Koźmiński, B. Brutscher, W. Bermel, R. Pierattelli, and I. C. Felli. 2014. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. J. Biomol. NMR 60: 209-218.
41. Barbato, G., M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax. 1992. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. Biochemistry 31: 5269-5278.
42. Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33: 5984-6003.
43. Peng, J. W., and G. Wagner. 1992. Mapping of spectral density function using heteronuclear NMR relaxation measurements. J. Magn. Reson. 98: 308-332.
44. Peng, J. W., and G. Wagner. 1994. Investigation of protein motions via relaxation measurements. Methods Enzymol. 239: 563-596.
45. Hwang, T. L., P. C. M. Van Zijl, and S. Mori. 1998. Accurate quantification of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange

(CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. J. Biomol. NMR 11: 221-226.

46. Vuister, G. W., and A. Bax. 1993. Quantitative J correlation: a new approach for measuring homonuclear three-bond J($H^N H^\alpha$) coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. 115: 7772-7777.
47. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2008. Optimization of random time domain sampling in multidimensional NMR. J. Magn. Reson. 192: 123-130.
48. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX Pipes. J. Biomol. NMR 6: 277-293.
49. Kazimierczuk, K., A. Zawadzka, W. Koźmiński, and I. Zhukov. 2006. Random sampling of evolution time space and Fourier transform processing. J. Biomol. NMR 36: 157-168.
50. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2009. Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling. J. Magn. Reson. 197: 219-228.
51. Kazimierczuk, K., A. Zawadzka-Kazimierczuk, and W. Koźmiński. 2010. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J. Magn. Reson. 205: 286-292.
52. Keller, R. L. J. 2004. The Computer Aided Resonance Assignment Tutorial, Cantina Verlag.
53. Goddard, T. D., D. G. Kneller, SPARKY 3, University of California, San Francisco.
54. Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, Llinas M, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. 2005. The CCPN data model for NMR spectrooscopy: development of a software pipeline. Proteins: Struct. , Funct. , Bioinf. 59: 687-696.
55. Tamiola, K., and F. A. Mulder. 2012. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem. Soc. Trans. 40: 1014-1020.
56. Tamiola, K., B. Acar, and F. A. A. Mulder. 2010. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J. Am. Chem. Soc. 132: 18000-18003.
57. Muñoz, V., and L. Serrano. 1994. Elucidating the folding problem of helical peptides using empirical parameters. Nat. Struct. Biol. 1: 399-409.
58. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J. Mol. Biol. 245: 275-296.
59. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. J. Mol. Biol. 245: 297-308.
60. Muñoz, V., and L. Serrano. 1997. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41: 495-509.

61. Lacroix, E., A. R. Viguera, and L. Serrano. 1998. Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. J. Mol. Biol. 284: 173-191.
62. Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21: 3433-3434.
63. Xue, B., R. L. Jr. Dunbrack, R. W. Williams, A. K. Dunker, and V. Uversky. 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804: 996-1010.
64. Camacho, C., G. Couloris, V. Avagyan, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.
65. Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30: 772-780.
66. Varadi, M., M. Guharoy, F. Zsolyomi, and P. Tompa. 2015. DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. BMC Bioinformatics 16: 153.
67. Dinkel, H., K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Krüger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sánchez, F. Diella, and T. J. Gibson. 2014. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. 42: D259-D266.
68. Blom, N., S. Gammeltoft, and S. Brunak. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294: 1351-1362.
69. Blom, N., T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, and S. Brunak. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 4: 1633-1649.
70. Ozenne, V., F. Bauer, L. Salmon, J. R. Huang, M. R. Jensen, Segard S., P. Bernadó, Charavay C., and M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28: 1463-1470.
71. Varadi, M., S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. L. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. S. Wishart, P. E. Wright, and P. Tompa. 2014. pE-DB: a database of structural ensemble of intrinsically disordered and of unfolded proteins. Nucleic Acids Res. 42: D326-D335.
72. Bermel, W., M. Bruix, I. C. Felli, V. M. V. Kumar, R. Pierattelli, and S. Serrano. 2013. Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J. Biomol. NMR 55: 231-237.
73. Kazimierczuk, K., J. Stanek, A. Zawadzka-Kazimierczuk, and W. Koźmiński. 2010. Random sampling in multidimensional NMR spectroscopy. Prog. NMR Spectrosc. 57: 420-434.
74. Felli, I. C., A. Piai, and R. Pierattelli. 2013. Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann. Rep. NMR Spectroscop. 359-418.

75. Ulrich, E. L., H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, and J. L. Markley. 2008. BioMagResBank. Nucleic Acids Res. 36: D402-D408.

76. Schwalbe, H., K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson. 1997. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. Biochemistry 36: 8977-8991.

77. Massad, T., J. Jarvet, R. Tanner, K. Tomson, J. Smirnova, P. Palumaa, M. Sugai, T. Kohno, K. Vanatalu, and P. Damberg. 2007. Maximum entropy reconstruction of joint phi, psi-distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from phi and psi-dependent J-couplings. J. Biomol. NMR 38: 107-123.

78. Meier, S., M. Blackledge, and S. Grzesiek. 2008. Conformational distributions of unfolded polypeptides from novel NMR techniques. J. Chem. Phys 128: 052204.

79. Otten, R., K. Wood, and F. A. A. Mulder. 2009. Comprehensive determination of $^{3}J_{HNHa}$ for unfolded proteins using $^{13}C'$-resolved spin-echo difference spectroscopy. J. Biomol. NMR 45: 343-49.

80. Lendel, C., and P. Damberg. 2009. 3D J-resolved NMR spectroscopy for unstructured polypeptides: fast measurement of $^{3}J_{HNH\alpha}$ coupling constants with outstanding spectral resolution. J. Biomol. NMR 44: 35-42.

81. Schweitzer-Stenner, R. 2012. Conformational propensities and residual structures in unfolded peptides and proteins. Mol. Biosyst. 8: 122-133.

82. Shi, Z., K. Chen, Z. Liu, and N. R. Kallenbach. 2006. Conformation of the Backbone in Unfolded Proteins. Chem. Rev. 106: 1877-1897.

83. Baxter, N. J., and M. P. Williamson. 1997. Temperature dependence of ^{1}H chemical shifts in proteins. J. Biomol. NMR 9: 359-369.

84. Cierpicki, T., and J. Otlewski. 2001. Amide proton temperature coefficients as hydrogen bond indicators in proteins. J. Biomol. NMR 21: 249-261.

85. Lee, S. H., D. H. Kim, J. J. Han, E. J. Cha, J. E. Lim, Y. J. Cho, C. Lee, and K. H. Han. 2012. Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. Curr. Prot. Pept. Sci. 13: 35-54.

86. Lee, C., L. Kalmar, B. Xue, P. Tompa, G. W. Daughdrill, V. N. Uversky, and K. H. Han. 2014. Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. Biochim. Biophys. Acta. 1840: 993-1003.

87. Kini, R. M., and H. J. Evans. 1995. A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites. Biochem. Biophys. Res. Commun. 212: 1115-1124.

88. Georgieva, E. R., S. Xiao, P. P. Borbat, J. H. Freed, and D. Eliezer. 2014. Tau binds to lipid membrane surfaces via short amphipathic helices located in its microtubule-binding repeats. Biophys. J. 107: 1441-1452.

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

90. He, J., J. Ye, Y. Cai, C. Riquelme, J. O. Liu, X. Liu, A. Han, and L. Chen. 2011. Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. Nucleic Acids Res. 39: 4464-4474.
91. Tompa, P., and M. Varadi. 2014. Predicting the predictive power of IDP ensembles. Structure 22: 177-178.
92. Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structure. J. Mol. Graphics 14: 51-55.
93. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J Comp Chem 25: 1605-1612.
94. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. Mc Cammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 98: 10037-10041.
95. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338: 1015-1026.
96. Mohan, A., C. J. Oldfield, P. Radivojac, M. S. Cortese, A. K. Dunker, and V. N. Uversky. 2006. Analysis of molecular recognition features (MoRFs). J. Mol. Biol. 362: 1043-1059.
97. Fuxreiter, M., P. Tompa, and I. Simon. 2007. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23: 950-956.
98. Lu, K. P., Y. C. Liou, and X. Z. Zhou. 2002. Pinning down proline-directed phosphorylation signaling. Trends Cell Biol. 12: 164-172.
99. Theillet, F. X., C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M. K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, and P. Selenko. 2012. Cell signaling, post-translational protein modifications and NMR spectroscopy. J. Biomol. NMR 54: 217-236.
100. Amata, I., M. Maffei, A. Igea, M. Gay, M. Vilaseca, A. R. Nebreda, and M. Pons. 2013. Multi-phosphorylation of the intrinsically disordere unique domain of c-Src studied by in-cell and real-time NMR spectroscopy. ChemBioChem 14: 1820-1827.
101. Amata, I., M. Maffei, and M. Pons. 2014. Phosphorylation of unique domains of Src family kinases. Front. Genet. 5: 181.
102. Iakoucheva, L. M., P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker. 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32: 1037-1049.

Figure 1

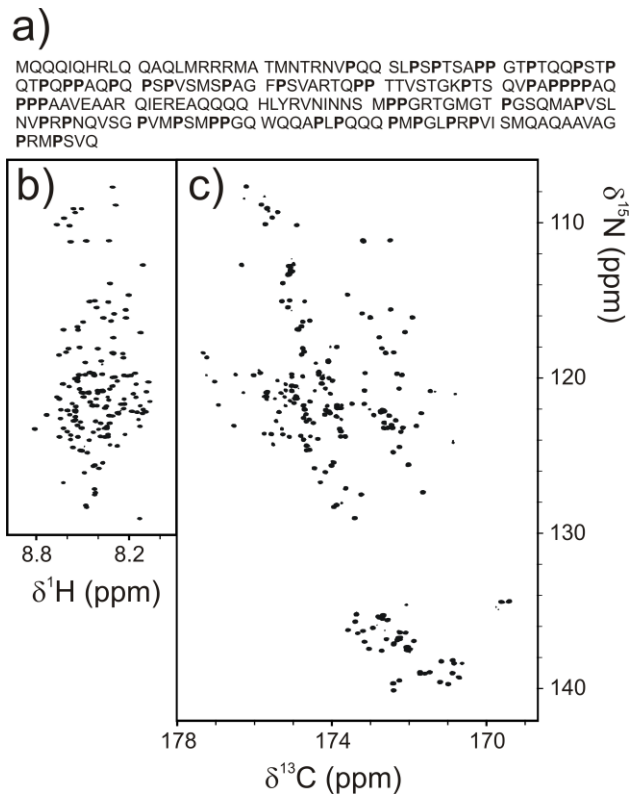


Figure 2

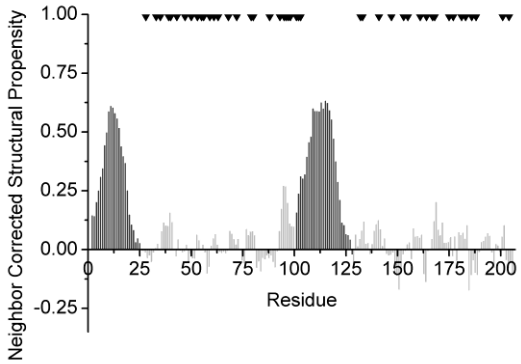


Figure 3

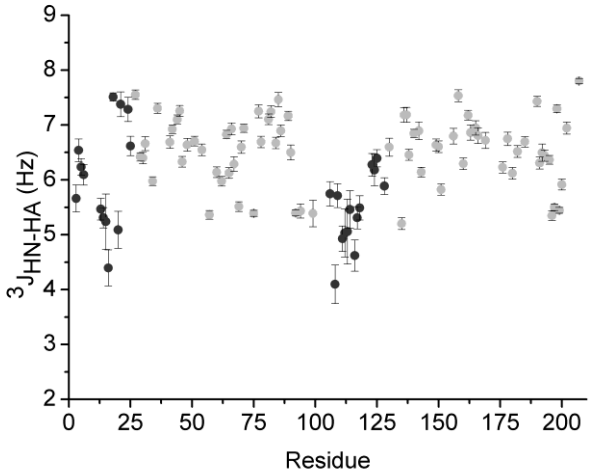


Figure 4

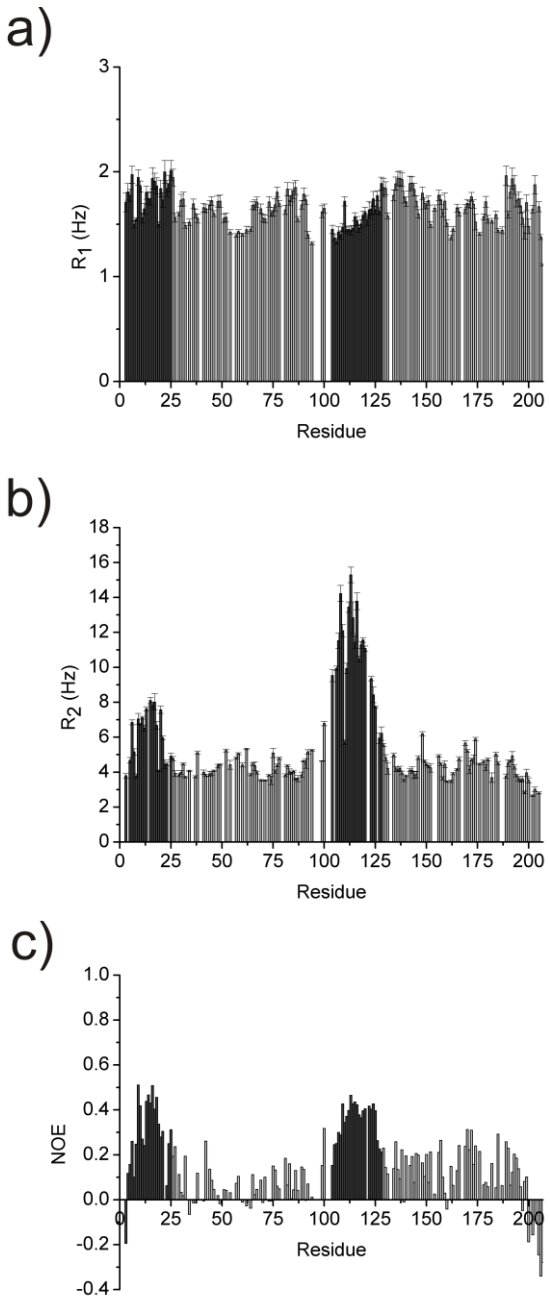


Figure 5

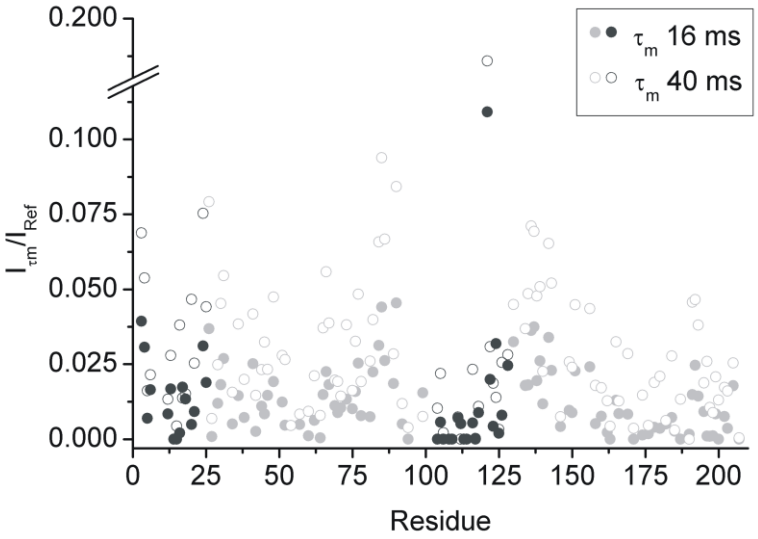


Figure 6

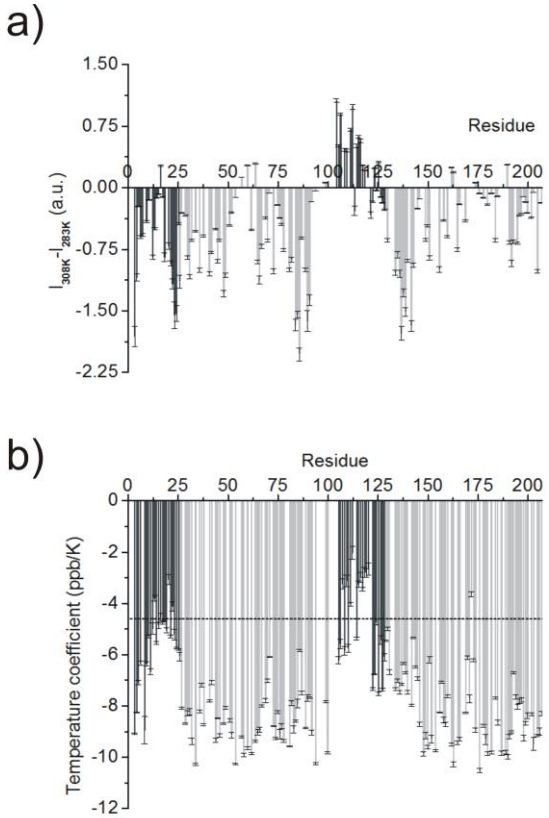


Figure 7

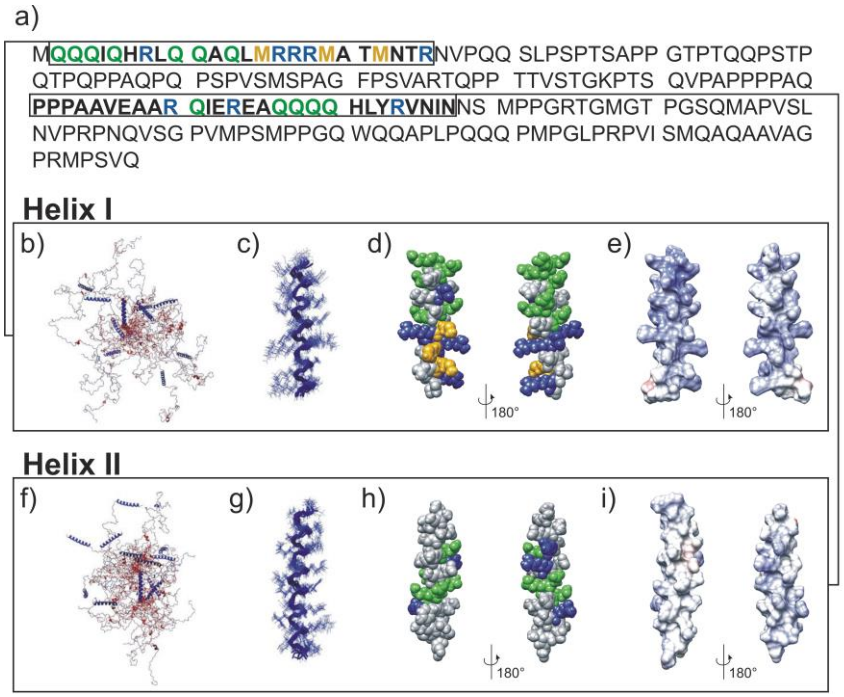
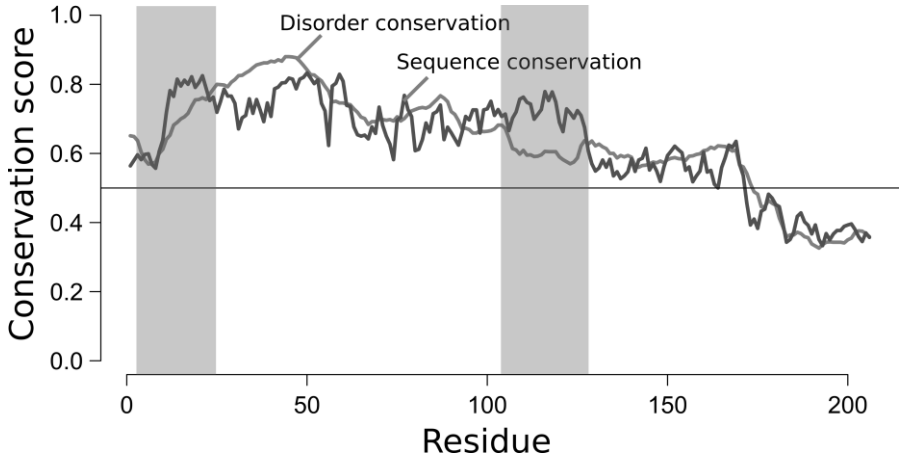


Figure 8



Just a flexible linker? The structural and dynamic properties of CBP-ID4 as revealed by NMR spectroscopy

Alessandro Piai[1], Eduardo O. Calçada[1], Thomas Tarenzi[1], Alessandro del Grande[1], Mihaly Varadi[2], Peter Tompa[2,3], Isabella C. Felli[1], Roberta Pierattelli[1]

1 Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino (Florence), Italy

2 VIB Structural Biology Research Center (SBRC), Vlaams Instituut voor Biotechnologie, 1050 Brussel, Belgium; Structural Biology Brussels (SBB), Vrije Universiteit Brussel, 1050 Brussel, Belgium

3 Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, 1117 Budapest, Hungary.

Supporting Material

Supporting figures

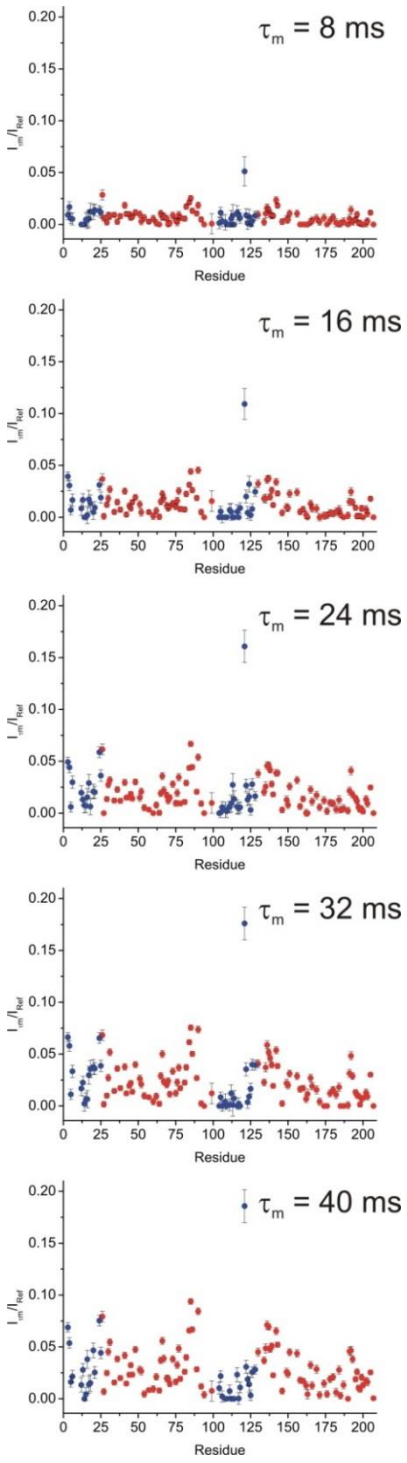


Figure S1. Extent of amide proton exchange with the solvent. $I\tau_m/I_{Ref}$ values measured through the (CLEANEX-PM)-FHSQC experiments are reported as a function of residue number. In the top right of each plot, the length of the mixing time τ_m is indicated. The residues belonging to the two protein regions characterized by higher propensity for α-helix are reported in blue.

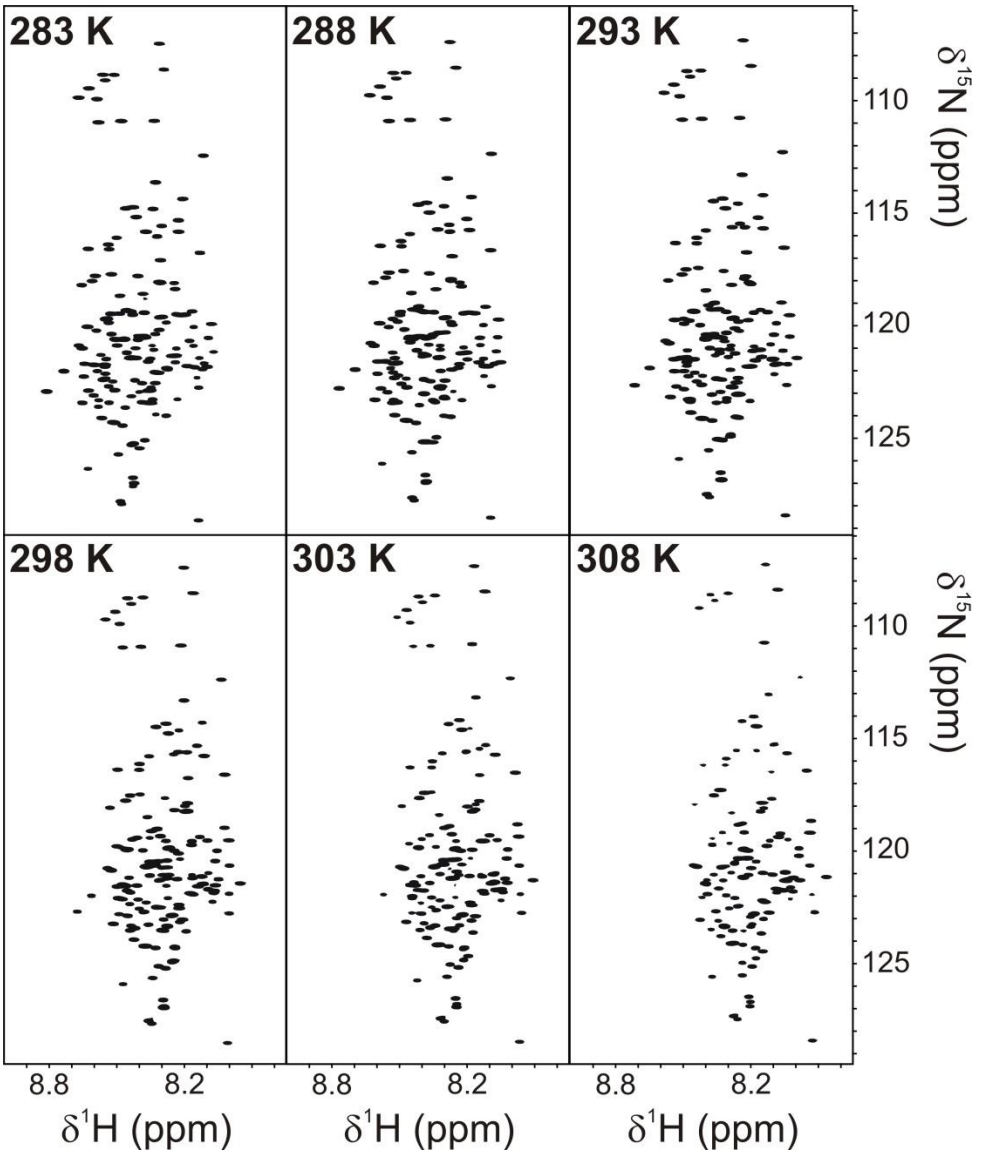


Figure S2. Temperature dependence of CBP-ID4. 2D BEST-TROSY spectra acquired at different temperature (from 283.0 to 308.0 K) are reported. The spectra are plotted with the same contour levels.

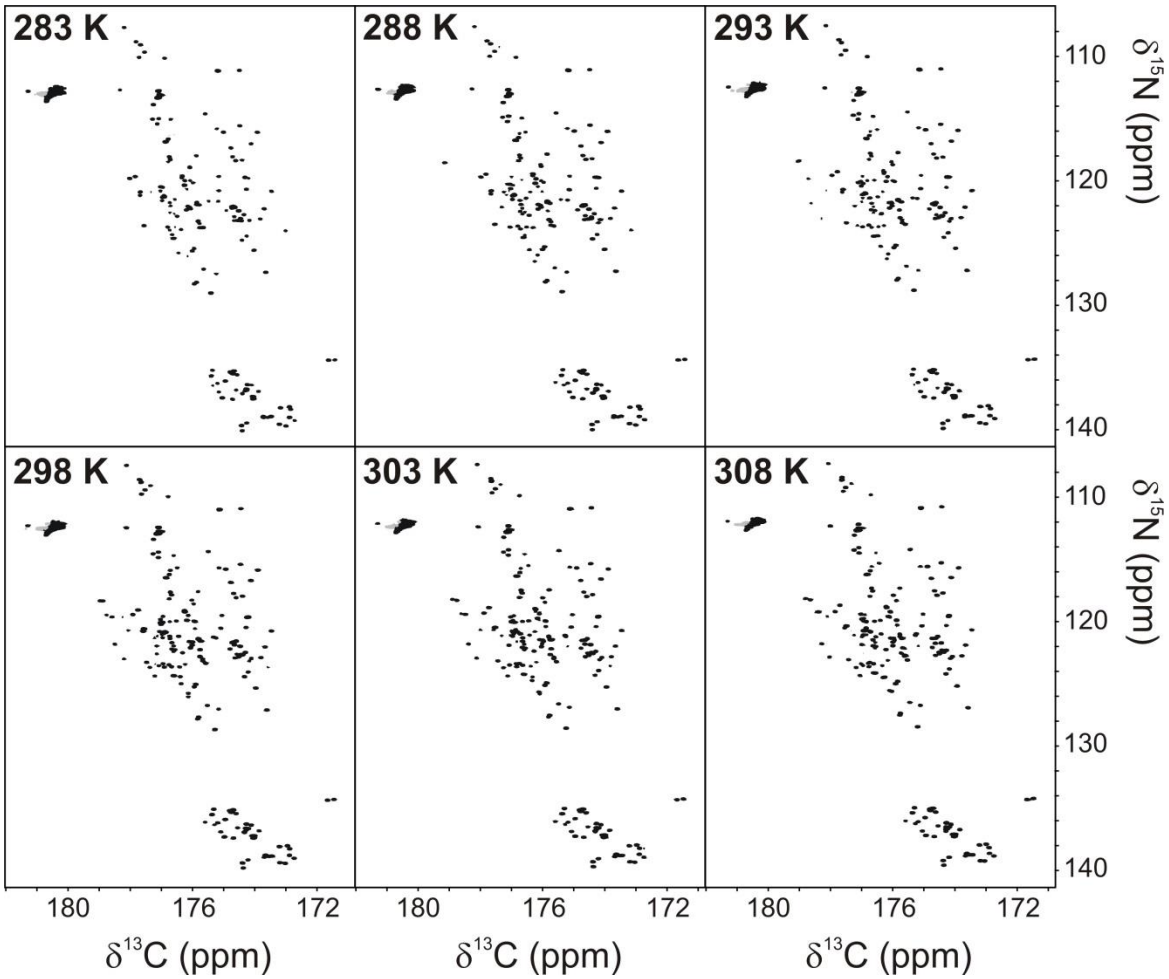


Figure S3. Temperature dependence of CBP-ID4. 2D CON-IPAP spectra acquired at different temperature (from 283.0 to 308.0 K) are reported. The spectra are plotted with the same contour levels.

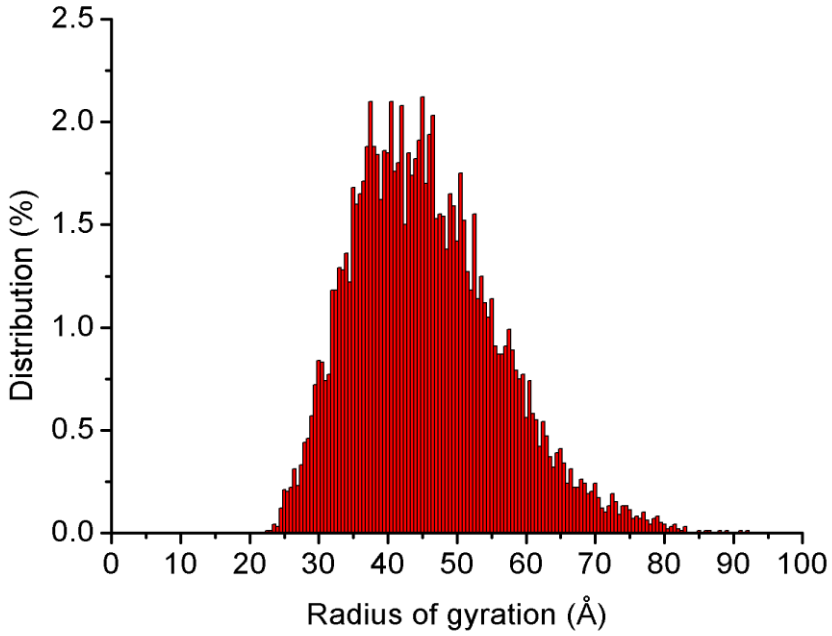


Figure S4. Radius of gyration of the conformers of the computed ensemble. The wide distribution of the radius of gyration reflects the ability of the linker to fine-modulate the relative distance between attached globular domains.

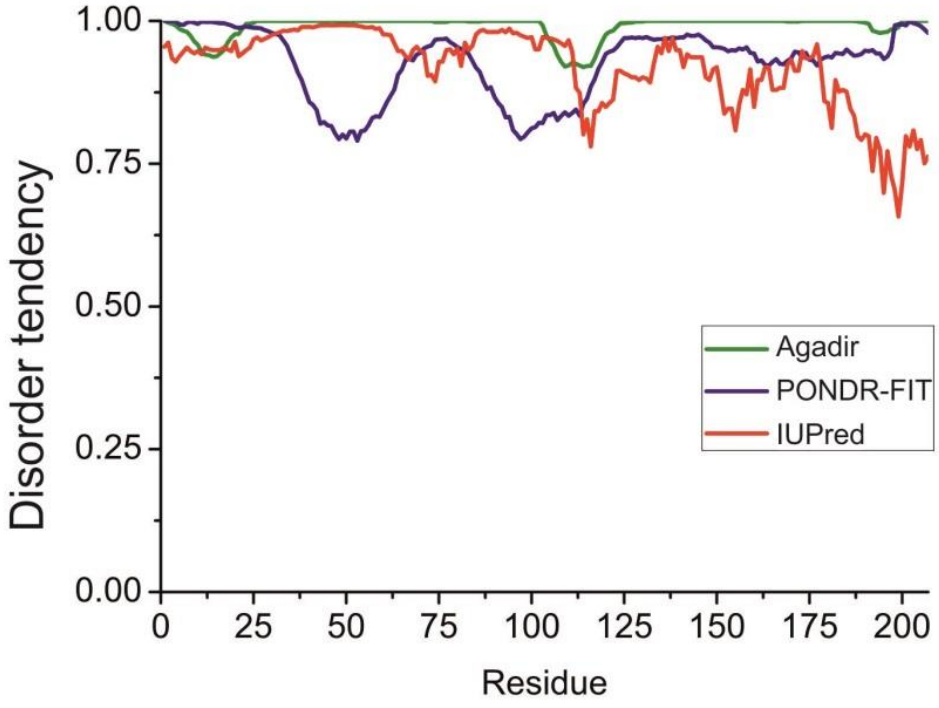


Figure S5. Disorder tendency of CBP-ID4 as predicted by IUPred (red), PONDR-FIT (blue) and Agadir (green).

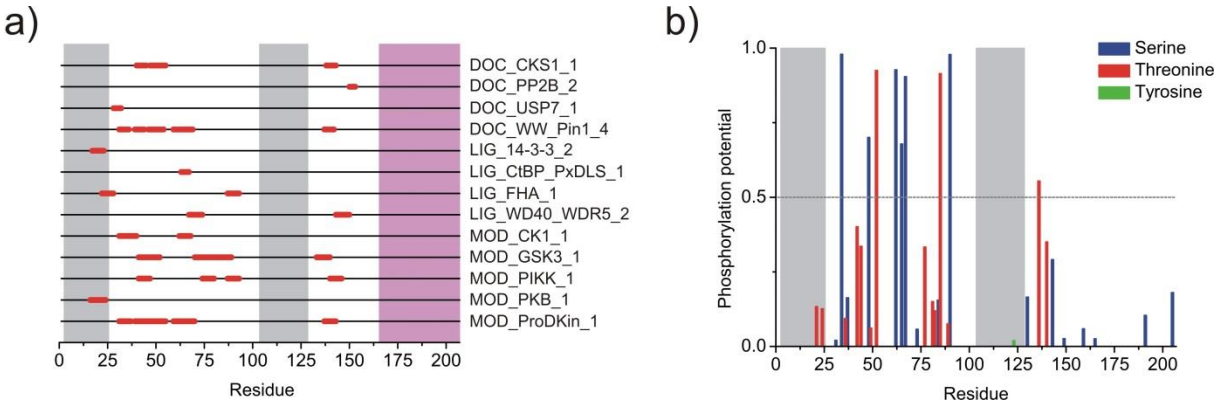


Figure S6. Distribution of potential SLiMs (a) and phosphorylation sites (b) of CBP-ID4. The two protein regions characterized by significant α-helical propensity are highlighted in gray: almost no SLiMs and phosphorylation sites are predicted in these regions (a). The protein region highlighted in magenta was discarded from the analysis since recognized by the protein families database (Pfam) as the interlocking domain of CBP, which forms a 3-helical non-globular array that forms interlocked heterodimers with its target. b) Residues having a phosphorylation potential above the threshold are considered to be highly probable phosphorylation sites. Phosphorylation sites predicted to be kinase specific are the following: Ser 34 (p38MAPK and cdk5 kinases); Thr 52 (p38MAPK, GSK3 and cdk5 kinases); Ser 62 (GSK3 and cdk5 kinases); Ser 67 (p38MAPK and cdk5 kinases); Thr 85 (PKC kinase); Ser 90 (DNAPK kinase).

NMR data acquisition and processing information

3D BEST-TROSY (BT) NMR experiments were performed at 21.1 T on a Bruker Avance spectrometer operating at 898.57 MHz ^{1}H, 225.95 MHz ^{13}C and 91.05 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. PC9 and E-BURP2 (or time reversed E-BURP2) shapes of durations of 1800 and 1270 μs, respectively, were employed for ^{1}H band-selective π/2 flip angle pulses; REBURP shape of duration of 1250 μs was used for ^{1}H band-selective π flip angle pulse; BIP-750-50-20 pulse shapes of duration of 140 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π/2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 274 and 190 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 660 μs). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. All gradients employed had a smoothed square shape.

5D BT-NMR experiments were collected at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. E-BURP2 and REBURP shapes of durations of 1200 and 1180 μs, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes of duration of 200 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π /2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 260 and 161 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 μs) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 μs, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. All gradients employed had a smoothed square shape.

3D and 4D ^{13}C detected experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. For ^{13}C band-selective π /2 and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes of durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 μs, 25 % smoothing, 80 kHz sweep width, 11.3 kHz RF field

strength). The ^{13}C band selective pulses on $C^{\alpha/\beta}$, C^{α}, and C' were given at the center of each region, respectively, and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz16 (1.7 kHz) and garp4 (1.0 kHz) sequences, respectively. All gradients employed had a sine-shape. All experiments employ the IPAP approach to remove the splitting in the direct acquisition dimension caused by the $C^{\square}$–C' couplings. The in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded.

The most relevant experimental parameters used for the acquisition of $^{1}H^{N}$ and ^{13}C detected NMR experiments for sequence-specific assignment are reported in Table S1 and Table S2, respectively. Experimental parameters of ^{15}N relaxation NMR experiments are provided in Table S3. ^{15}N relaxation rates (R_1 and R_2) were determined by fitting cross-peak intensities, measured as a function of a variable delay, as single-exponential decay. ^{15}N-^{1}H NOEs were obtained as the ratio between cross-peak intensities in spectra recorded with and without ^{1}H saturation.

The most relevant experimental parameters for (CLEANEX-PM)-FHSQC and 3D HNHA experiments are given in Table S4. Exchange rates between amide protons and water protons were estimated by analyzing the ratio between cross-peak intensities in spectra recorded with and without CLEANEX-PM mixing period. Homonuclear $^{3}J_{HN\text{-}H\alpha}$ couplings were quantified by measuring the diagonal-peak to cross-peak ratio obtained in the 3D ^{15}N-separated quantitative J-correlation HNHA spectrum.

Finally, parameters related to the NMR experiments for temperature dependence are shown in Table S5.

Table S1 Experimental parameters used for the acquisition of $^{1}H^{N}$ detected NMR experiments for sequence-specific assignment

	Spectral widths and maximal evolution times					No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
2D BEST-TROSY	2300 Hz (^{15}N) 222.6 ms				10800 Hz ($^{1}H^{N}$) 95.1 ms	2	0.200 s	1024	512	15 min	100.0
3D BT-HNCO	2400 Hz (^{13}C') 52.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	2	0.200 s	1024	4500	4 h 20 min	27.5
3D BT-HN(CA)CO	2400 Hz (^{13}C') 52.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	4500	8 h 45 min	27.5
3D BT-HNCACB	14000 Hz ($^{13}C^{\alpha/\beta}$) 24.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	6000	11 h 45 min	13.4
3D BT-HN(CO)CACB	14000 Hz ($^{13}C^{\alpha/\beta}$) 24.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	6000	12 h 30 min	13.4
3D BT-(H)N(COCA)NH	2300 Hz (^{15}N) 28.7 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	8	0.200 s	1024	4500	18 h 20 min	52.5
3D BT-(H)N(CA)NNH	2300 Hz (^{15}N) 33.0 ms		2300 Hz (^{15}N) 33.0 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	16	0.200 s	1024	2400	20 h 30 min	40.5
5D BT-(H)NCO(CAN)CONNH	2600 Hz (^{15}N) 40.4 ms	2600 Hz (^{13}C') 40.4 ms	2600 Hz (^{13}C') 40.4 ms	2600Hz (^{15}N) 40.4 ms	13300 Hz ($^{1}H^{N}$) 77.0 ms	8	0.200 s	1024	2000	1 d 15 h	0.002
5D BT-HN(COCAN)CONNH	2600 Hz (^{15}N) 40.4 ms	2600 Hz (^{13}C') 40.4 ms	2600 Hz (^{13}C') 40.4 ms	2600Hz (^{15}N) 40.4 ms	13300 Hz ($^{1}H^{N}$) 77.0 ms	8	0.200 s	1024	2000	1 d 15 h	0.003

Table S2 Experimental parameters used for the acquisition of ^{13}C detected NMR experiments for sequence-specific assignment

	Spectral widths and maximal evolution times				No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
2D CON-IPAP	2600 Hz (^{15}N) 98.5 ms		5300 Hz (^{13}C) 96.8 ms		16	2.500	512	256	12 h 30 min	100.0
3D (H)CBCACON-IPAP	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 50.0 ms	8800 Hz (^{13}C) 58.3 ms		8	1.000 s	512	1100	1 d	8.8
3D (H)CBCANCO-IPAP	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 31.9 ms	8800 Hz (^{13}C) 58.3 ms		16	1.000 s	512	1100	2 d 4 h	13.7
4D (HCA)CON(CA)CON-IPAP	2400 Hz (^{13}C) 24.2 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 40.0 ms	8800 Hz (^{13}C) 58.3 ms	16	0.900 s	512	850	3 d 1 h	0.2
4D (HN)CON(CA)CON-IPAP	2400 Hz (^{13}C) 24.2 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 40.0 ms	8800 Hz (^{13}C) 58.3 ms	32	0.600 s	512	700	3 d 18 h	0.2

Table S3 Experimental parameters used for the acquisition of ^{15}N relaxation NMR experiments

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
^{15}N R_1	1800 Hz (^{15}N) 126.8 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	8	3.000 s	1024	3 h 15 min – 4 h 15 min
^{15}N R_2	1800 Hz (^{15}N) 126.8 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	8	3.000 s	1024	3 h 40 min
Steady-state heteronuclear $^{15}N\{^{1}H\}$ NOEs	1800 Hz (^{15}N) 144.3 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	72	6.000 s	1024	2 d 15 h
For the determination of R_1, 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2, 10 experiments were acquired changing the variable delay from 30 to 375 ms.						

Table S4 Experimental parameters used for the acquisition of (CLEANEX-PM)-FHSQC and 3D HNHA NMR experiments

	Spectral widths and maximal evolution times			No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
(CLEANEX-PM)-HSQC	1700 Hz (^{15}N) 147.3 ms	8400 Hz ($^{1}H^{N}$) 61.1 ms		8	5.000 s	512	5 h 55 min
3D HNHA	1700 Hz (^{15}N) 46.0 ms	7000 Hz ($^{1}H^{\alpha}$) 9.1 ms	7000 Hz ($^{1}H^{N}$) 107.2 ms	8	1.000 s	750	2 d 7 h
For the determination of the extent of amide proton exchange with the solvent, 5 experiments were acquired changing the length of the mixing time from 8 to 40 ms. An additional FHSQC experiment was measured with the same parameters to obtain the reference peak volumes.							

Table S5 Experimental parameters used for the acquisition of the NMR experiments for temperature dependence

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
2D BEST-TROSY	2300 Hz (^{15}N) 222.6 ms	11400 Hz ($^{1}H^{N}$) 400.0 ms	2	0.001	4600	20 min
2D CON-IPAP	3500 Hz (^{15}N) 72.6 ms	9600 Hz (^{13}C) 53.6 ms	16	1.500 s	512	7 h 40 min