SUPPLEMENTARY INFORMATION

Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP

Sara Contreras-Martos1,2 & Alessandro Piai3,4, Simone Kosol1,2,5, Mihaly Varadi1,2,6, Angela Bekesi1,2, Pierre Lebrun1,2, Alex Volkov1,2, Kris Gevaert,1,2 Roberta Pierattelli1,4, Isabella Felli1,2, Peter Tompa*1,2,7

1VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie (VIB), Brussel, Belgium
2Vrije Universiteit Brussel (VUB), Brussel, Belgium
3Magnetic Resonance Center, University of Florence, Florence, Italy
4Department of Chemistry "Ugo Schiff", University of Florence, Florence, Italy
5Department of Chemistry, University of Warwick, Gibbet Hill, Coventry, CV4 7AL, United Kingdom.
6Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
7VIB-UGent Center for Medical Biotechnology, Ghent, Belgium
8Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary
9Department of Biochemistry, Ghent University, Ghent, Belgium

* Corresponding author

To whom correspondence should be addressed: Prof. Peter Tompa, Structural Biology Research Center, Vlaams Instituut voor Biotechnologie (VIB) at Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussel, Belgium, Telephone +32 2 629 19 62; E-mail: peter.tompa@vub.be
**SI MATERIALS AND METHODS**

**Sequence analysis**

PSI-BLAST\(^1\) search was performed using the human ID3 fragment sequence against the non-redundant (nr) protein dataset of NCBI. Multiple sequence alignment (MSA) of ID3 sequences was generated by MAFFT\(^2\) to characterize the conservation of the amino acid sequence. Structural disorder was analyzed using the DisCons online tool\(^3\). Within DisCons, IUPred in ‘short’ running mode\(^4\) and VSL2\(^5\) were used to predict residue-level structural disorder; other parameters were left as default. Secondary-structure propensities were predicted by Psipred\(^6\). Identification of functional motifs and potential binding sites within ID3 was based on ELM database\(^7\), MoRFpred\(^8\), ANCHOR\(^9\) and DynaMine\(^10\).

**Scrambling the sequence of human ID3**

To ensure that the scrambled sequence generated is disordered, we used IUPred\(^4\), ESPritz\(^11\) and VSL2\(^5\), and only sequences in which the disorder scores of each residue exceeded 0.5 by all three methods were retained for further analysis. Aggregation-prone sequences were filtered out by TANGO\(^12\) applying a threshold of 0.001. Finally, sequences were also screened for the presence of potential binding motifs by definitions derived from the ELM database, and only sequences without recognizable motifs were retained. Only one protein sequence was selected for the control experiments. ID3 and scrambled-ID3 (scrID3) sequences were aligned using T-Coffee (default mode)\(^13\) and EMBOSs Needle\(^14\). To graphically display the sequence alignment, BoxShade open service from the SIB ExPASy Bioinformatics Resources Portal was used (cf. Supplementary Fig. S8).

**Cloning and mutagenesis**

Full-length human CBP (UniProt id Q92793) was amplified by PCR from synthesized CBP cDNA (Origene - RC219036), cloned in pDONOR221 and transferred to pDEST10 using Gateway\(^\text®\) Cloning Technology (Life Technologies). The gene fragment encoding the disordered domain ID3 (residues 674-1080 of CBP) was sub-cloned into pet16b using the restriction sites Ndel and BamHI and into pGEX4T using the restriction sites BamHI and XhoI. The cDNA encoding for a shorter fragment, denoted as ID3-F1 (aa 674-848), was sub-cloned into pet200D via TOPO cloning. The cDNA encoding for the core domain of CBP (residues 1095-1849) was sub-cloned into pDEST10 via TOPO cloning. The fragment of ZFP106 (ZFP106-f) identified in Y2H binding assay (residues 1265-1521 of full-length ZFP106 (UniProt Q9H2Y7)) was PCR-amplified and cloned into pet16b using the restriction sites XhoI and BamHI. All the constructs were sequenced before transformation into expression strain E. coli BL21(DE3).

Cys-to-Ser mutant ID3(C845S) was obtained by site-directed mutagenesis of Cys845 of ID3 following the manufacturer’s protocol of QuikChange II Site-Directed Mutagenesis Kit (Agilent #200523).

Scrambled-ID3 (scrID3), ID3delCTR and the three single-cysteine mutants of the ZFP106-f were cloned into pET16b (GenScript, Piscataway, USA) (M1: C1472S (paramagnetic label on C1307), M2: insertion of Cys between Thr1346 – Arg1347 (paramagnetic label on insC1347*), M3: C1307S (paramagnetic label on C1472)).

**Yeast-two-hybrid screening**
The coding sequence of CBP ID3 (GenBank accession number gi: 119943103) was provided to Hybrigenics Services S.A. The sequence was PCR-amplified and cloned into pB27 as a C-terminal fusion to LexA (N-LexA-ID3-C) and into pB35 as a C-terminal fusion to Gal4 DNA binding domain (N-Gal4-ID3-C). The constructs were confirmed by sequencing and were used as baits to screen random-primed human placenta and lung cancer cDNA libraries. In case of the LexA construct, 142.7 million interactions were analyzed, whereas for the Gal4 construct the number of interactions was 72.2 million. After selection on a medium lacking leucine, tryptophan and histidine, 189 clones were finally processed. The prey fragments were amplified by PCR and sequenced for identification of the corresponding interacting proteins. Sequence analysis of the identified proteins (full length and fragments) was performed using IUPred\(^\text{13}\) for disorder prediction and UniProt\(^\text{18}\) for gene annotation (GO) mapping. In order to compare the disorder content of the Y2H assay fragments with that of their corresponding full-length sequences, we have assembled two sequence sets. The first set contained the amino acid sequences of all the identified fragments, while the second set consisted of the full-length protein sequences retrieved from UniProt/SwissProt\(^\text{18}\). We used IUPred\(^\text{13}\) to quantify the per residue disorder propensities of all the collected sequences and calculated the ratio of disordered residues within each sequences, i.e. the ratio of residues that have a disorder propensity of 0.5 or greater. The distribution of these ratios within the two sequence sets was compared (c.f. Supplementary Fig. S3b). The distributions were compared using a non-parametric paired Wilcoxon-Mann-Whitney test, since this statistical test does not rely on Gaussian distribution and only assumes continuity of the variables and independence within the two samples. According to the test the average ratio of disordered residues is significantly higher in the fragments than within the full-length sequences (p-value 0.03808). This indicates that the fragments that were identified by the Y2H assay are relatively more disordered than the rest of their protein chains, and consequently this might suggest that ID3 favors interactions with disordered binding regions.

**Recombinant protein expression, purification and labeling**

1) **Full-length CBP and its core domain**

Full-length CBP was expressed in Sf9 insect cells\(^\text{17}\) and purified by consecutive nickel (His-select, Sigma-Aldrich) and anti-FLAG (Sigma-Aldrich) immunoaffinity chromatography followed by size exclusion chromatography in the presence of protease inhibitors (2 \(\mu\)g/ml leupeptin, 1 \(\mu\)g/ml bestatin, 1 g/ml pepstatin A, 5 \(\mu\)g/ml E64, and 2 \(\mu\)M 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF)). Chelating agents were avoided to keep the Zn-finger domains intact. Purified protein was flash frozen in liquid \(N\)\(_2\) and stored at -80°C in storage buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 50 \(\mu\)M tris(2-carboxyethyl)phosphine (TCEP), and protease inhibitors: 1 \(\mu\)g/ml leupeptin, 0.5 \(\mu\)g/ml bestatin, 0.5 \(\mu\)g/ml pepstatin A, 1 \(\mu\)g/ml E64, and 0.1 \(\mu\)M PMSF).

2) **Other protein constructs**

ID3, scrID3, ID3delCTR, ID3(C845S), ID3-F1, ZFP106-f, Single-Cys mutants of ZFP106-f (M1, M2 and M3) and human calpastatin domain 1 (hCSD1) were generated as follows:

Transformed *E. coli* BL21(DE3) (invitrogen) cells were grown in Luria Broth (LB) medium at 37°C until optical density (OD) reached 0.6 and protein expression was induced with 1 mM IPTG. For \([^{15}\text{N-13}\text{C}]\) labeled protein samples, cells were grown to high cell density and centrifuged (15 minutes, 3500 g); pelleted cells were resuspended in M9 medium and the culture was grown for 2.5 h at 37°C before induction\(^\text{18}\). Cell cultures were harvested after overnight expression at 25°C for all ID3 samples or 2h at 37°C for ZFP106-f. The pelleted cells were resuspended in lysis buffer (50 mM KH\(_2\)PO\(_4\), 300 mM NaCl, 10 mM Imidazole, 1 EDTA-free protease inhibitor cocktail tablet (Roche), DNase, lysozyme, pH 8.0), sonicated and followed by heat-treatment (85°C, 10 minutes), in case of all ID3 samples\(^\text{19}\). Cell
lysates were cleared by centrifuging at 13000 g for 45 minutes. Subsequent purifications were done by IMAC (HisTrap HP column; GE Healthcare) followed by size-exclusion chromatography (Superdex200 26/600 and Superdex75 26/600; GE Healthcare) in a buffer 50 mM KH₂PO₄, 150 mM NaCl, 1 mM DTT, pH 7.5, complemented by 2 EDTA-free protease inhibitor cocktail tablets (Roche). Following buffer exchange into distilled water (dH₂O), the sample was dried at room temperature (RT) in a vacuum concentrator (ID3) or lyophilized (ZFP106-f) and stored at -20°C. Samples for PRE measurements, ID3(C845S) and single-cysteine ZFP106-f mutants (M1, M2 and M3), were flash frozen in solution after size exclusion and stored at -80°C.

3) Paramagnetic labeling of the ZFP106-f mutants and hCSD1
Purified single-cysteine ZFP106-f mutants (M1, M2 and M3, see “Protein constructs”) and the hCSD1 (for control experiment) were incubated with 10 mM DTT for 15 minutes at 30°C followed by buffer exchange into 20 mM Tris-HCl pH 8.0 in a desalting column (HiTrap 5mL desalting, GE Healthcare). The metal-containing paramagnetic probe was prepared by mixing together one molar equivalent of MnCl₂ (Sigma-Aldrich) and the thiol-reactive label, N-[S-(2-pyridylthio)cysteaminy]ethylenediamine-N,N,N',N'-tetraacetic acid, monoamide (Toronto Research Chemicals, #P996250). The desalted protein sample was mixed with a 5-fold molar excess of the metal-containing probe, and the reaction was carried out overnight at RT. The modified proteins were purified by ion-exchange chromatography (HiTrap Q FF, GE Heliathcare) using 20 mM Tris-HCl buffer pH 8.0 and, eluted with a linear salt gradient. A single defined peak was eluted at 0.5 mM NaCl concentration. The purified samples were desalted into 50 mM potassium phosphate, 0.02% NaN₃ pH6.5. All buffers were pre-treated with chelating beads (Chelex, Sigma) to remove traces of heavy metals.

Determination of the paramagnetic labeling by intact mass (mass spectroscopy) analysis
Purified proteins were desalted using C18 Spin Columns (Thermo Scientific) and diluted in 50% acetonitrile 0.1% acetic acid. Intact protein mass measurements were performed by direct infusion in a microelectrospray ionization ion trap mass spectrometer (LTQ XL, ThermoFisher Scientific, San José, CA). The parent ions were submitted to in-source dissociation (SID) using the minimal energy to promote efficient declustering of water molecules and salts adducts. The mass spectra were deconvoluted using the software ProMass Deconvolution from ThermoFisher Scientific.

Circular dichroism spectroscopy
Far-UV circular dichroism spectra of ID3 (0.1 mg/ml), ZFP10-f (0.1 mg/ml) and the complex in 1:1 ratio were recorded on a J-715 spectropolarimeter in 50 mM KH₂PO₄, 150 mM NaCl, 5 mM DTT buffer pH 6.5. Data points were collected every 0.5 nm at the controlled temperature of 24°C. Spectra are presented as the circular dichroism absorption coefficient calculated from the mean residue ellipticity (cf. Supplementary Fig. S1)²⁰.

Small-angle X-ray scattering experiments
The Aviex charge-coupled device detector was positioned at 1507.5 mm. A total volume of 60 µl of a ID3 sample (10 mg/ml) was injected into a size-exclusion column (SEC-3, 300 Å Agilent), using an Agilent HPLC system, and eluted directly into the SAXS flow-through capillary cell at a flow rate of 0.2 ml min⁻¹ and at 15°C²¹. The size-exclusion buffer used was 50 mM Tris, 150 mM NaCl, 5 mM Tcep, pH 6.5. SAXS data were collected continuously, with frame duration of 1.0 s and dead time of 0.5 s between frames. Selected frames corresponding to the main protein elution peak were averaged using FOXTROT²¹. A large number of frames were collected during the
void volume of the elution and averaged to account for buffer scattering, which was subsequently subtracted from the signals during the protein elution. Data reduction to absolute units, frame averaging and subtraction were done using FOXROT. All subsequent data processing, analysis and modelling steps were carried out with the ATSAS suite\textsuperscript{32}. The radius of gyration $R_g$ was derived by the Guinier approximation $I(q) = I(0) \exp(-q^2R_g^2/3)$ for $qR_g < 1.0$ using PRIMUS QT\textsuperscript{33}. GNOM was used to compute the pair-distance distribution functions, $P(r)^{34}$. This approach also features the maximum dimension of the macromolecule, $D_{max}$. Normalized Kratky plots (i.e. $(qR_g)^2I(q)/I(0)$ as a function of $qR_g$) were used to assess the conformational behavior of the polypeptide chain\textsuperscript{35}.

**Nuclear magnetic resonance (NMR) and paramagnetic relaxation enhancement (PRE) experiments**

See also Supporting Information Tables S1-S5

1) **Sequence-specific assignment:**
All the multidimensional NMR experiments for sequence-specific assignment were performed on a sample of 0.4 mM uniformly $^{13}$C,$^{15}$N labeled ID3 in 50 mM KH$_2$PO$_4$, 150 mM NaCl, 2.5 mM TCEP buffer at pH 6.5, with 10% D$_2$O added for the lock.

A set of multidimensional $^1$H and $^{13}$C detected NMR experiments tailored to achieve sequence-specific assignment of IDPs was acquired at 297 K. 2D CON\textsuperscript{36,27}, 2D (HCA)CON\textsuperscript{36,27}, 2D BEST-CON\textsuperscript{38}, 3D (H)CACON\textsuperscript{39}, 4D HCBACON\textsuperscript{39}, 4D (HCA)CON(CA)CON\textsuperscript{40}, 4D (H$^{deuter}$N)CON(CA)CON\textsuperscript{40}, 5D H$^{deuter}$NCACON\textsuperscript{29} and 5D H$^{deuter}$NCACON\textsuperscript{29} experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700 MHz 1H, equipped with a cryogenically cooled probehead optimized for $^{13}$C-direct detection. 2D BEST-TROSY (BT)$^{31,32}$, 3D TROSY-HNCO$^{31,35,34}$, 3D (H)N(CA)NNH\textsuperscript{25}, 4D TROSY-(H)NCO(CA)NNH\textsuperscript{26} and 4D TROSY-HN(COCA)NNH\textsuperscript{25} experiments were performed at 22.3 T on a Bruker Avance III spectrometer operating at 950 MHz, equipped with a cryogenically cooled probehead.

NMR experiments for sequence-specific assignment of ID3-F1 were acquired on a sample of 0.2 mM uniformly $^{13}$C, $^{15}$N labeled protein in 50 mM KH$_2$PO$_4$, 150 mM NaCl, 5.0 mM TCEP buffer at pH 6.5, with 10% D$_2$O added for the lock. A series of multidimensional proton detected NMR experiments (2D BEST-TROSY, 3D BT-HNCO, 3D BT-HN(CA)CO\textsuperscript{31,34}, 3D BT-HNCCAB\textsuperscript{31,34}, 3D BT-HN(CO)CAB\textsuperscript{31,34} for sequence-specific assignment were acquired at 21.1 T on a Bruker Avance spectrometer operating at 900 MHz, equipped with a cryogenically cooled probehead.

The parameters used for the acquisition of all the experiments are reported in the Supporting information S1-S4 Tables. 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\textsuperscript{37} was chosen to generate the time schedules with the RSPack program (available at http://nmr.cent3.uw.edu.pl). The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with $\sigma$=0.5.

2) **Backbone dynamics**
Backbone dynamics (at 297 K) of 0.2 mM uniformly $^{15}$N-labeled ID3 were studied by using $^{15}$N relaxation data obtained from 2D $^1$H-$^{15}$N HSQC-edited experiments\textsuperscript{38}. $^{15}$N spin-lattice relaxation rate constant ($R_1$), $^{15}$N spin-spin relaxation rate constant ($R_2$) and steady-state heteronuclear $^1$H-$^{15}$N NOEs were determined for about 85% of non-proline backbone amide $^{15}$Ns. The experiments were performed in 50 mM KH$_2$PO$_4$, 150 mM NaCl, 2.5 mM TCEP buffer at pH 6.5, with 10% D$_2$O added for the lock. Heteronuclear $^{15}$N relaxation experiments for the measurement of $R_1$, $R_2$ and $^{15}$N-$^1$H NOEs were measured at 297 K at 16.4 T on a Bruker Avance spectrometer operating at 700 MHz, equipped with a cryogenically cooled probehead.
3) Paramagnetic relaxation enhancement (PRE) experiments

For the PRE measurements, 0.25 mM [13C, 15N] ID3(C84SS) and 1 molar equivalent (eq) of single-Cys ZFP106-f mutant tagged with EDTA(Mn2+), were used for the paramagnetic samples. The diamagnetic references were done with wt ZFP106-f, as it has been previously shown that wt proteins can be considered equivalent to their EDTA(Zn2+) tagged version for PRE referencing. The control experiments were carried out on 0.25 mM [15N, 13C] ID3(C84SS) samples and 1 eq of wt hCSD1-EDTA(Mn2+) (paramagnetic sample) or wt hCSD1 (diamagnetic sample). For the reduced-label control experiments, the same sample used for the Iw/Iax plots (at 0 and 150 mM) was desalted (HiTrap Desalting 5 mL, GE Healthcare) in 50 mM KH2PO4, 0.02% NaN3, pH 6.5, concentrated and re-measured in absence and presence of reducing agents (addition of 2.5 mM DTT final concentration). All NMR samples contained 50 mM KH2PO4, 0.02% NaN3, pH 6.5 and 10 % D2O for the lock. For the high salt measurements, NaCl was added to the sample to the final concentration of 150 mM. In case of wt references and control experiments (reduced label or wt hCSD1 diamagnetic reference), 2.5 mM DTT was added to the sample.

2D [1H, 15N] HSQC spectra of the paramagnetic and diamagnetic samples were acquired with 10 seconds repetition delay in order to determine Iw/Iax intensity ratios as described elsewhere. The parameters used for the acquisition of all the experiments are reported in the Supporting information S5 Tables.

For representation purpose of the Iw/Iax plots a threshold value was calculated. To that aim the higher 10% values were left out and the following 20% were averaged. All the Iw/Iax values were divided by this average to get the normalized PRE.

PRE experiments were performed at the Jean Jeener Center, VUB (Belgium) at 297k in a Varian Direct-Drive 800 MHz or 600 MHz NMR spectrometers, the latter equipped with a cold probe for enhanced sensitivity.

4) NMR data processing

Conventionally sampled NMR data sets were processed using Bruker TopSpin 1.3 software, except those of PRE experiments, which were processed using nmrPipe. NUS data sets were converted with nmrPipe 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), implemented in ToASTO and reduced programs respectively. Both programs are available at http://nmr.cent3.uw.edu.pl.

Sparky was used to analyze the spectra for sequence-specific assignment, whereas CcpNmr Analysis was employed to analyze 15N relaxation and PRE data.

Secondary structure propensity calculation

Secondary structure propensity analysis was performed using C1, C2, C3 and N chemical shifts (95% assigned). The ncSP from the heteronuclear chemical shifts was determined by using the neighbor-corrected structural propensity calculator (ncSP) tool, available online at http://nmr.chem.rug.nl. The random-coil chemical shift library derived from IDP NMR analysis was chosen for the analysis.

Ensemble modelling

Ensemble models of ID3 were based on the combination of SAXS data and ncSP derived from experimental chemical shift values. Flexible-Meccano was used to generate a pool of 10,000 semi-random conformations; the conformational sampling was biased by ncSP. Side chains were modeled onto the backbone of the conformers by SCCOMP. Then, the theoretical SAXS scattering curve of
each single conformer was calculated by CRYSOL\textsuperscript{33}, and finally the GAJOE algorithm was employed\textsuperscript{11} to select subsets of the random pool that – when combined together – had a theoretical scattering curve that agreed with the experimental SAXS data.

**Acetylation assay and identification of the acetylation sites**

*In vitro* acetylation assays were performed in 50 μl buffer (50 mM NaH\textsubscript{2}PO\textsubscript{4}, 125 mM NaCl, 0.5 mM DTT, pH 7.4) at 30°C except for the time points of the H4 reaction by core domain- and full-length CBP that were done at 20°C (50mM Tris, 150 mM NaCl, 5% Glycerol, 50 μM Tcep, 0.02 mg/mL BSA, protease inhibitors, pH 7.5). The reactions were completed to saturation. For each reaction, 624 pmols of lyophilized ZFP106-f or histone 4 (H4, control reaction), 18.5 pmols of HAT-active full-length, and 20 μmol of acetyl-CoA were used. For competition experiments, ID3, scr-ID3 or ID3delCTR (control reactions) were added in 10-, 100- or 150-times molar excess to CBP. The samples were analyzed by WB by an anti-acetylated-lysine antibody (Bioke) and anti-His (Sigma-Aldrich).

For the identification of acetylation sites, the proteins in solution were desalted and concentrated on a C-18 spin column (Pierce, ThermoScientific) prior to an overnight pepsin or endoglu-C digest in 50 mM NH\textsubscript{4}HCO\textsubscript{3} at 30°C. In case of SDS-PAGE gel band samples, the proteins were washed, completely dry in the vacuum system and digested with trypsin ON at 37°C in 50 mM NH\textsubscript{4}HCO\textsubscript{3}, pH 8.0, 10% ACN. The reactions were stopped by adding 0.1% trifluoroacetic acid. The peptides were analyzed by LC-MS/MS in a microelectrospray ionization ion trap mass spectrometer (LTQ XL, Thermo Scientific or Orbitrap XL) as described here \textsuperscript{52,53}. For peptide identification, peak lists were generated using the application spectrum selector in the Proteome Discoverer 1.4 package. The resulting peak lists were searched using Sequest against a homemade protein database containing the protein sequences or the SwissProt *E. coli* database concatenated with full-length CBP and histone H4. Peptide matches were filtered using charge-state versus cross-correlation scores (Xcorr) and manually validated.

For the complete MS sequence coverage, two MS analyses were needed as each independent analysis covered only 62% and 90% of ZFP106-f amino acid sequence.

**Cross-linking Assay**

Cross-linking reactions were carried out for 15 or 30 min at RT in PBS containing 1 mM DTT (S: sample, ID3:ZFP106-f; C: control, GST:ZFP106-f). Proteins were mixed in approximate 1 to 1 ratio (concentrations adjusted by SDS-PAGE) and two different concentrations of the BS3 cross-linking reagent 0.25 mM and 2.5 mM were added to sample and control reaction. SDS-PAGE gels were run with samples before (12.5% SDS-PAGE gel; no cross-linker added) and after the cross-linking reaction (7.5% SDS-PAGE gel to improve the resolution of high Mw). The SDS-PAGE after the cross-linking was transferred to a blotting membrane and two western blots (anti-His and anti-GST with stripping step in between) were done in order to visualize the appearance of high Mw overlapping bands (formed complexes).
REFERENCES


Goddard, T. D. Sparky - NMR Assignment and Integration Software


SUPPLEMENTARY FIGURES AND FIGURE LEGENDS

Fig. S1. 2D HN and CON spectra of ID3 and quality-check on the heat-treatment step during ID3 purification. (a) Comparison of 2D HN and CON spectra of ID3 highlights the improvement on the signal dispersion obtained by applying $^{13}$C-direct detection methods. (b) Circular dichroism intensities of ID3 non-boiled (light green) and ID3 boiled (dark green) given as a mean residue ellipticity ($\Theta_{	ext{MRE}}$). (c) The superimposition of the 2D BEST-TROSY spectra of boiled (black) and non-boiled (red) ID3 sample.
Fig. S2. Structural and dynamic characterization of ID3 by NMR. Dynamic characterization of ID3: (a) $^{15}$N $R_1$ relaxation rates; (b) $^{15}$N $R_2$ relaxation rates; (c) $^{15}$N-$^1$H NOEs.
Fig. S3. Ensemble fit and ID3 preference for binding disordered regions of the identified proteins. (a) Experimental scattering curve of the SAXS measurements (green) and the fit of the selected ensemble (teal). (b) Density plot. Paired Wilcoxon Mann-Whitney statistical test gives a p-value of 0.03808, indicating statistically significant difference between the distributions of the ratios of disordered residues in fragments (red) versus their full-length proteins (black).
**Fig. S4. Observed and predicted sequence features of ID3.** Experimental and in silico analyses of sequence features of potential functional importance. The ncSP scores plotted on the bottom show regions with a tendency to form helices \( y > 0 \) or sheet \( y < 0 \). Missing segments correlates to unassigned residues. This (experimental) information correlates with backbone dynamics predicted with DynaMine and predicted MoRFs. Interestingly, the more pronounced helices derived from ncSP scores correspond to DynaMine scores associated with context dependence. DynaMine profile peaks observed at the C-terminal region of ID3 (underlined). There are several potential ANCHOR binding regions scattered along the sequence, and sequence conservation also suggests the functional importance of a significant portion of ID3. Some highly conserved segments also correspond to predicted MoRFpred sites. One segment (682-844) is also observed as a transient helix in the ensemble model of ID3. A high density of post-translational modification (PTM) sites points to the functional importance of the C-terminal region of ID3 (Methylation\(^{44} \) in red; Acetylation\(^{55} \) in blue; SUMOylation\(^{56} \) in orange; Phosphorylation\(^{47} \) in purple.

![Graph showing sequence features of ID3 with DynaMine and MoRF predictions.](image)
Fig. S5. ID3–ZFP106-f interaction assessed by competition binding assay. (a) Competition assays demonstrate that recombinant ID3 competes with ID3 in the full-length CBP for binding ZFP106-f, which is shown by a concentration-dependent decrease on the ZFP106-f acetylation in presence of an excess of recombinant ID3 (WB developed with anti-acetylated-lysine antibody). In contrast, addition of the same amount of scrambled ID3 (scrID3) did not cause any effect (control experiment 1). (b) Control experiment 2: The same experiments were run using histone 4 as a substrate, and no changes on H4 acetylation level could be detected after addition of an excess of ID3. Acetylation was followed by anti-Ac-Lys antibody, ID3 was added to the acetylation reaction in an excess of 10, 100 or 150 times to full-length CBP (18.5 pmol) and its effect was also compared to that of scrambled-ID3 (scrID3; control reaction) (C+: control reaction). Loading controls are shown with the SDS-PAGE and WB developed with anti-His antibody (both ID3 and ZFP106-f are histagged). (c) Competition assays demonstrate that recombinant ID3 competes with ID3 in the full-length CBP for binding ZFP106-f, which is shown by a concentration-dependent decrease on the ZFP106-f acetylation in presence of an excess of recombinant ID3 (WB developed with anti-acetylated-lysine antibody). In contrast, addition of the same amount of a C-terminal deletion mutant of ID3 (ID3delCTR) did not cause any effect (control experiment 3). Note: The apparent Mw on SDS-PAGE gel of ID3delCTR and ZFP106-f is the same. Therefore, addition of an excess of ID3delCTR is seen as an increase in thickness of ZFP106-f band (or just above it), without this perturbing the WB analysis of acetylation.
Fig. S6. ID3 and ZFP106-f can be specifically cross-linked “in vitro”

(a) In the SDS-PAGE gel before the proteins of sample (S: sample containing ID3 - at 100 kDa - and ZFP106-f - about 40 kDa-) and control (C: control containing GST - 25 kDa – and ZFP106-f- about 40 kDa -) previous to the cross-linking reaction are shown. (b) Cross-linking reactions were carried out for 15 or 30 min at RT and, two different concentrations of the BS3 cross-linker (0.25 mM and 2.5 mM) were added to sample and control reaction (Note: Excess of the cross-linker reagent results in the nonspecific cross-linking of all species present in the reaction). SDS-PAGE gels after the cross-linking reaction, together with two western blots (anti-His and anti-GST with stripping step in between) allow us to visualize the appearance of high Mw overlapping bands corresponding to ZFP106-f – ID3 complex (red box). Specificity of the reaction was assessed by the control reaction, where under similar conditions, a complex between ZFP106-f and GST was not observed. However, we could identify the formation of other complexes such as ZFP106-f and GST dimer.
Fig. S7. C-terminal region (CTR) of ID3 interacts with the middle region of ZFP106-f. (a) NMR titration of ID3 and ZFP106-f. Garrett plot of ID3 chemical shift perturbations upon adding 7.8 equivalents of ZFP106-f. (b – e) $I_{para}/I_{ax}$ plots of the HSQC cross-peaks of ID3 in the presence of the paramagnetically tagged single-Cys ZFP106-f mutants M1, M2 and M3 or hCSD1 (control) under different conditions. A non-specific interaction between the label attached to the protein and the NTR of ID3 was observed in all experiments (blue). In contrast, only ZFP106-f-M2 caused PRE effect at the CTR of ID3 (red) due to the specific interaction between the two protein fragments. This signal diminished in the presence salt as shown by $I_{para}/I_{ax}$ values in the absence (b) and presence (c) of 150 mM NaCl. Control experiments (d) reduced label and (e) unrelated protein were carried out only at 0 mM NaCl.
Fig. S8. Scrambled-ID3 sequence and alignment with human ID3. Scrambled ID3 (scrID3) sequence generated as control for the acetylation competition assay. Identical or similar residues in the multiple-alignment chart are represented by black or grey, respectively. ID3 and scrID3 sequence alignment shows a level of amino acid identity lower than 30%.

>scrID3
QPSSGATPGAGVEYPMVEEPDQAGGQSRVGQTSAMGSPPLSNGSNAPDDTAPNTKVPMQPAVPSVPAQQMRMPSQTLKQOPNMSSEQQTATIMMPSPDAIV ANGQPSRTSHQLNSARGTQANKKVQVPKKLPVQPLCQPQKNAPKMGPAEPSLSSLANSQPSVESPVQQASEKPNPVMQEGTQASVAPAPTQSPSNAL QGPPPPQAPMAPSVMGQPTAQPSTQETDPGGSSQTTTPAPAQTDDQQTSPSEQPPERPDNLSEGMTEMQHTQTPGLTPQVPPAQATAMPMQQKKQLPPLPPAPMVNPNSMGELALQGSEPOSQGQTVPSPTSHPNESTPTFEEQOAPQMAMISTNLQSGPESVTVIQVVFEMVHHSVPELKPNSLV

<table>
<thead>
<tr>
<th>ID3</th>
<th>scrID3</th>
<th>ID3</th>
<th>scrID3</th>
<th>ID3</th>
<th>scrID3</th>
<th>ID3</th>
<th>scrID3</th>
<th>ID3</th>
<th>scrID3</th>
<th>ID3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HKQGILGNQPALPAQAMPDPVPLAPFREN------GPLEPQPDQVYTRGAGFMDK</td>
<td>1</td>
<td>Q----------</td>
<td>FSAAT----SEPAPLQVEYEDQQAGQGQGPGDSHARFL</td>
<td>56</td>
<td>GNYYLQAPMGPAADAQEGTGSMGVSFVMAA ESFRPPPENMGAETNNQA</td>
<td>45</td>
<td>SNEPES-----HISTASPPFPPPASVQQRRGSTQGQPNSSEQATAMPS</td>
<td>115</td>
<td>-------------</td>
</tr>
</tbody>
</table>
## SUPPLEMENTARY TABLES

### Table S1: Experimental parameters used for the acquisition of the NMR experiments on ID3 for sequence-specific assignment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Spectral widths and maximal evolution times</th>
<th>No. of scans</th>
<th>Inter-scan delay (s)</th>
<th>No. of complex points (a)</th>
<th>No. of hypercomplex points</th>
<th>Duration of the experiment</th>
<th>Relative data points density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D BEST-TROSY</td>
<td>2200 Hz ((^1)N) 231.2 ms 11/400 Hz ((^1)H) 358.3 ms</td>
<td>4</td>
<td>0.005</td>
<td>4066</td>
<td>512</td>
<td>40 min</td>
<td>100.0</td>
</tr>
<tr>
<td>2D CON-IFAP</td>
<td>2700 Hz (^1)N) 34.8 ms 8800 Hz (^1)H) 58.3 ms</td>
<td>16</td>
<td>1.2</td>
<td>512</td>
<td>256</td>
<td>12 h 30 min</td>
<td>100.0</td>
</tr>
<tr>
<td>2D [HAC]CON-IFAP</td>
<td>2700 Hz (^1)N) 34.8 ms 8800 Hz (^1)H) 58.3 ms</td>
<td>16</td>
<td>1.2</td>
<td>512</td>
<td>256</td>
<td>6 h 50 min</td>
<td>100.0</td>
</tr>
<tr>
<td>2D BEST-CON-IFAP</td>
<td>1600 Hz (^1)N) 31.3 ms 5300 Hz (^1)H) 96.8 ms</td>
<td>64</td>
<td>0.4</td>
<td>512</td>
<td>50</td>
<td>3 h</td>
<td>100.0</td>
</tr>
<tr>
<td>3D [H]CON-IFAP</td>
<td>4800 Hz (^1)C) 2700 Hz (^1)H) 8800 Hz (^1)H)</td>
<td>8</td>
<td>1.2</td>
<td>512</td>
<td>1250</td>
<td>1 d 7 h 25 min</td>
<td>9.3</td>
</tr>
<tr>
<td>3D HRECA/NNH</td>
<td>2000 Hz (^1)N) 23.5 ms 2000 Hz (^1)H) 21.5 ms 9800 Hz (^1)H) 105.0 ms</td>
<td>32</td>
<td>1.0</td>
<td>1024</td>
<td>770</td>
<td>1 d 11 h</td>
<td>36.5</td>
</tr>
<tr>
<td>3D TROSY HNCD</td>
<td>2300 Hz (^1)C) 10.4 ms 2300 Hz (^1)H) 21.8 ms 14300 Hz (^1)H) 71.9 ms</td>
<td>8</td>
<td>1.2</td>
<td>1024</td>
<td>685</td>
<td>8 h 30 min</td>
<td>20.0</td>
</tr>
<tr>
<td>4D [H]C=BACON-IFAP</td>
<td>11200 Hz (^1)H) 6.8 ms 2700 Hz (^1)N) 30.0 ms 8800 Hz (^1)H) 58.3 ms</td>
<td>8</td>
<td>1.1</td>
<td>512</td>
<td>1085</td>
<td>2 d 3 h 40 min</td>
<td>48.8</td>
</tr>
<tr>
<td>4D J[C]BACON-IPAP</td>
<td>11200 Hz (^1)H) 6.8 ms 2700 Hz (^1)N) 31.9 ms 8800 Hz (^1)H) 58.3 ms</td>
<td>16</td>
<td>1.2</td>
<td>512</td>
<td>775</td>
<td>3 d 12 h</td>
<td>32.8</td>
</tr>
<tr>
<td>4D [H]ACON[C]CON-IPAP</td>
<td>2000 Hz (^1)C) 3000 Hz (^1)N) 3000 Hz (^1)H) 8800 Hz (^1)H)</td>
<td>16</td>
<td>1.2</td>
<td>512</td>
<td>1050</td>
<td>4 d 16 h 20 min</td>
<td>0.33</td>
</tr>
<tr>
<td>4D [H]3N[C]CON-ACON-IPAP</td>
<td>2000 Hz (^1)C) 3000 Hz (^1)N) 3000 Hz (^1)H) 8800 Hz (^1)H)</td>
<td>12</td>
<td>0.7</td>
<td>512</td>
<td>950</td>
<td>5 d 16 h 10 min</td>
<td>0.36</td>
</tr>
<tr>
<td>4D TROSY [V]R[4]C[5]C[6]A/NNH</td>
<td>2300 Hz (^1)N) 26.1 ms 2300 Hz (^1)C) 26.3 ms 2500 Hz (^1)H) 26.7 ms 11300 Hz (^1)H) 77.0 ms</td>
<td>8</td>
<td>1.2</td>
<td>1024</td>
<td>2900</td>
<td>3 d 15 h</td>
<td>1.4</td>
</tr>
<tr>
<td>4D TROSY HN[C]C[5]C[6]A/NNH</td>
<td>1100 Hz (^1)N) 24.5 ms 2100 Hz (^1)C) 26.1 ms 2500 Hz (^1)H) 26.7 ms 11300 Hz (^1)H) 77.0 ms</td>
<td>8</td>
<td>1.2</td>
<td>1024</td>
<td>1360</td>
<td>1 d 11 h</td>
<td>1.4</td>
</tr>
<tr>
<td>5D [H]3[N]ACON-IFAP</td>
<td>1000 Hz (^1)N) 1700 Hz (^1)N) 4800 Hz (^1)H) 2700 Hz (^1)H) 8800 Hz (^1)H)</td>
<td>16</td>
<td>0.5</td>
<td>512</td>
<td>620</td>
<td>2 d 23 h</td>
<td>0.016</td>
</tr>
<tr>
<td>5D [H]3[N]ACON-IFAP</td>
<td>1000 Hz (^1)N) 1700 Hz (^1)N) 4800 Hz (^1)H) 2700 Hz (^1)H) 8800 Hz (^1)H)</td>
<td>12</td>
<td>0.5</td>
<td>1024</td>
<td>575</td>
<td>6 d 1 h</td>
<td>0.009</td>
</tr>
</tbody>
</table>
Table S2 Experimental parameters used for the acquisition of $^{15}$N relaxation NMR experiments.

<table>
<thead>
<tr>
<th>$^{15}$N R$_1$</th>
<th>Spectral widths and maximal evolution times</th>
<th>No. of scans</th>
<th>Inter-scan delay (s)</th>
<th>No. of complex points</th>
<th>Duration of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{15}$N R$_2$</td>
<td>1700 Hz ($^{15}$N) 150.3 ms</td>
<td>10 500 Hz ($^{13}$C) 97.6 ms</td>
<td>8</td>
<td>3.0</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>1700 Hz ($^{15}$N) 150.3 ms</td>
<td>10 500 Hz ($^{13}$C) 97.6 ms</td>
<td>8</td>
<td>3.0</td>
<td>256</td>
</tr>
<tr>
<td>Steady-state heteronuclear $^{15}$N($^{13}$C) NOEs</td>
<td>1700 Hz ($^{15}$N) 150.3 ms</td>
<td>10 500 Hz ($^{13}$C) 97.6 ms</td>
<td>80</td>
<td>6.0</td>
<td>256</td>
</tr>
</tbody>
</table>

For the determination of R$_1$, 10 experiments were acquired changing the variable delay from 15.0 ms to 1.0 s. For the determination of R$_2$, 10 experiments were acquired changing the variable delay from 10.0 to 410.0 ms.

Table S3 Experimental parameters used for the NMR titration between ID3 and ZFP106.

<table>
<thead>
<tr>
<th>2D BEST-TROSY</th>
<th>Spectral widths and maximal evolution times</th>
<th>No. of scans</th>
<th>Inter-scan delay (s)</th>
<th>No. of complex points</th>
<th>Duration of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2200 Hz ($^{15}$N) 231.2 ms</td>
<td>11 400 Hz ($^{13}$C) 358.3 ms</td>
<td>16</td>
<td>0.005</td>
<td>4096</td>
<td>2 h 30 min</td>
</tr>
</tbody>
</table>

Table S4 Experimental parameters used for the acquisition of the NMR experiments on ID3-F1 for sequence-specific assignment.

<table>
<thead>
<tr>
<th>2D BEST-TROSY</th>
<th>Spectral widths and maximal evolution times</th>
<th>No. of scans</th>
<th>Inter-scan delay (s)</th>
<th>No. of complex points</th>
<th>Duration of the experiment</th>
<th>Relative data points density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000 Hz ($^{15}$N) 256.0 ms</td>
<td>10 800 Hz ($^{13}$C) 95.1 ms</td>
<td>4</td>
<td>0.2</td>
<td>1024</td>
<td>512</td>
<td>35 min</td>
</tr>
<tr>
<td>2500 Hz ($^{15}$C) 59.6 ms</td>
<td>2000 Hz ($^{15}$N) 63.5 ms</td>
<td>4</td>
<td>0.2</td>
<td>1024</td>
<td>2900</td>
<td>5 h 35 min</td>
</tr>
<tr>
<td>2500 Hz ($^{15}$C) 59.6 ms</td>
<td>2000 Hz ($^{15}$N) 63.5 ms</td>
<td>4</td>
<td>0.2</td>
<td>1024</td>
<td>2900</td>
<td>5 h 40 min</td>
</tr>
<tr>
<td>16 000 Hz ($^{13}$C) 21.8 ms</td>
<td>2000 Hz ($^{15}$N) 63.5 ms</td>
<td>4</td>
<td>0.2</td>
<td>1024</td>
<td>3700</td>
<td>7 h 20 min</td>
</tr>
<tr>
<td>16 000 Hz ($^{13}$C) 21.8 ms</td>
<td>2000 Hz ($^{15}$N) 63.5 ms</td>
<td>32</td>
<td>0.2</td>
<td>1024</td>
<td>3300</td>
<td>2 d 5 h 45 min</td>
</tr>
</tbody>
</table>

Table S5 Experimental parameters used for the PRE NMR experiments between ID3 and ZFP106-f-M1/2/3 or hCSD1.

<table>
<thead>
<tr>
<th>($^{15}$N, $^{13}$N) HSQC</th>
<th>Spectral widths and maximal evolution times</th>
<th>No. of scans</th>
<th>Inter-scan delay (s)</th>
<th>No. of complex points</th>
<th>Duration of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1300 Hz ($^{15}$N) 100.3 ms</td>
<td>9800 Hz ($^{13}$N) 104.8 ms</td>
<td>16</td>
<td>10.0</td>
<td>1024</td>
<td>11 h 40 min</td>
</tr>
</tbody>
</table>