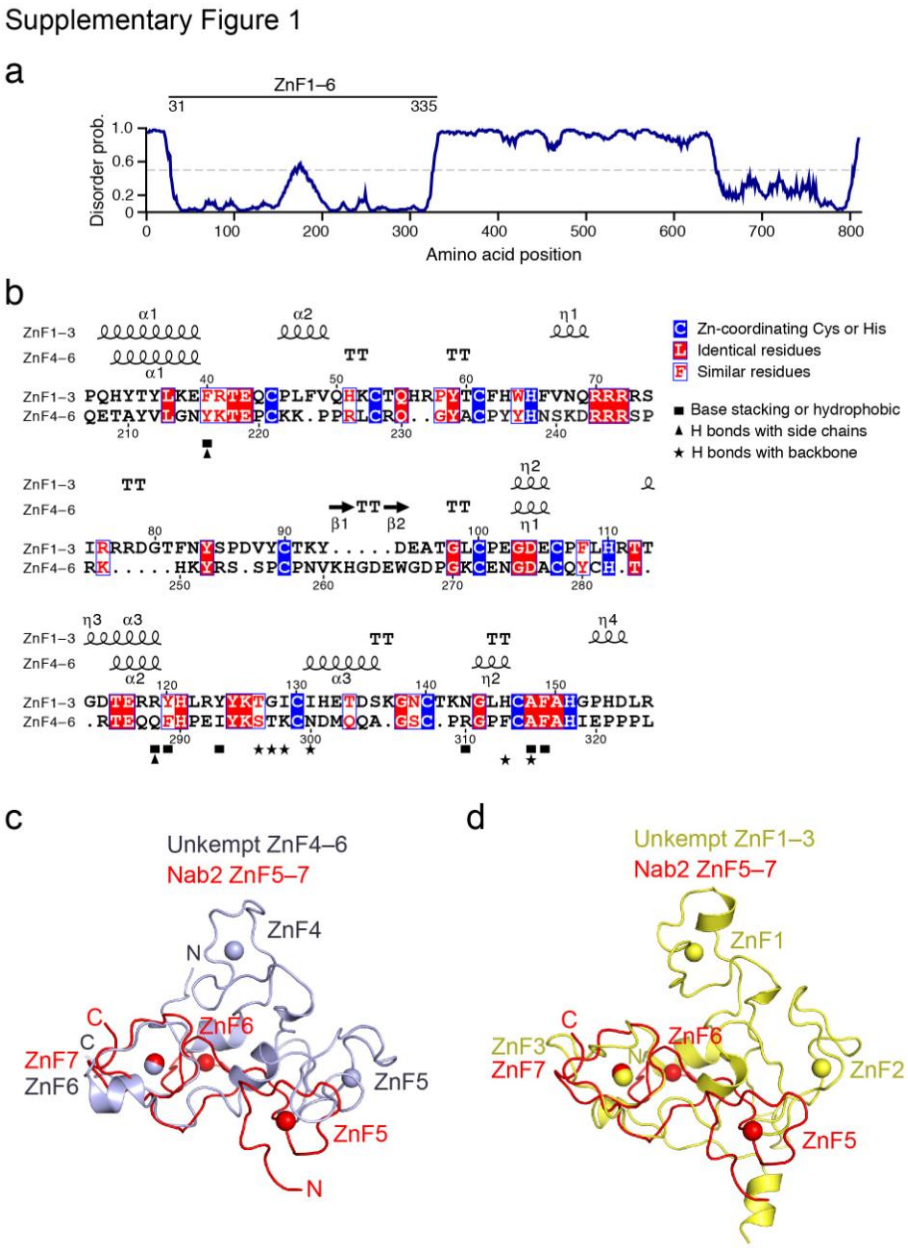
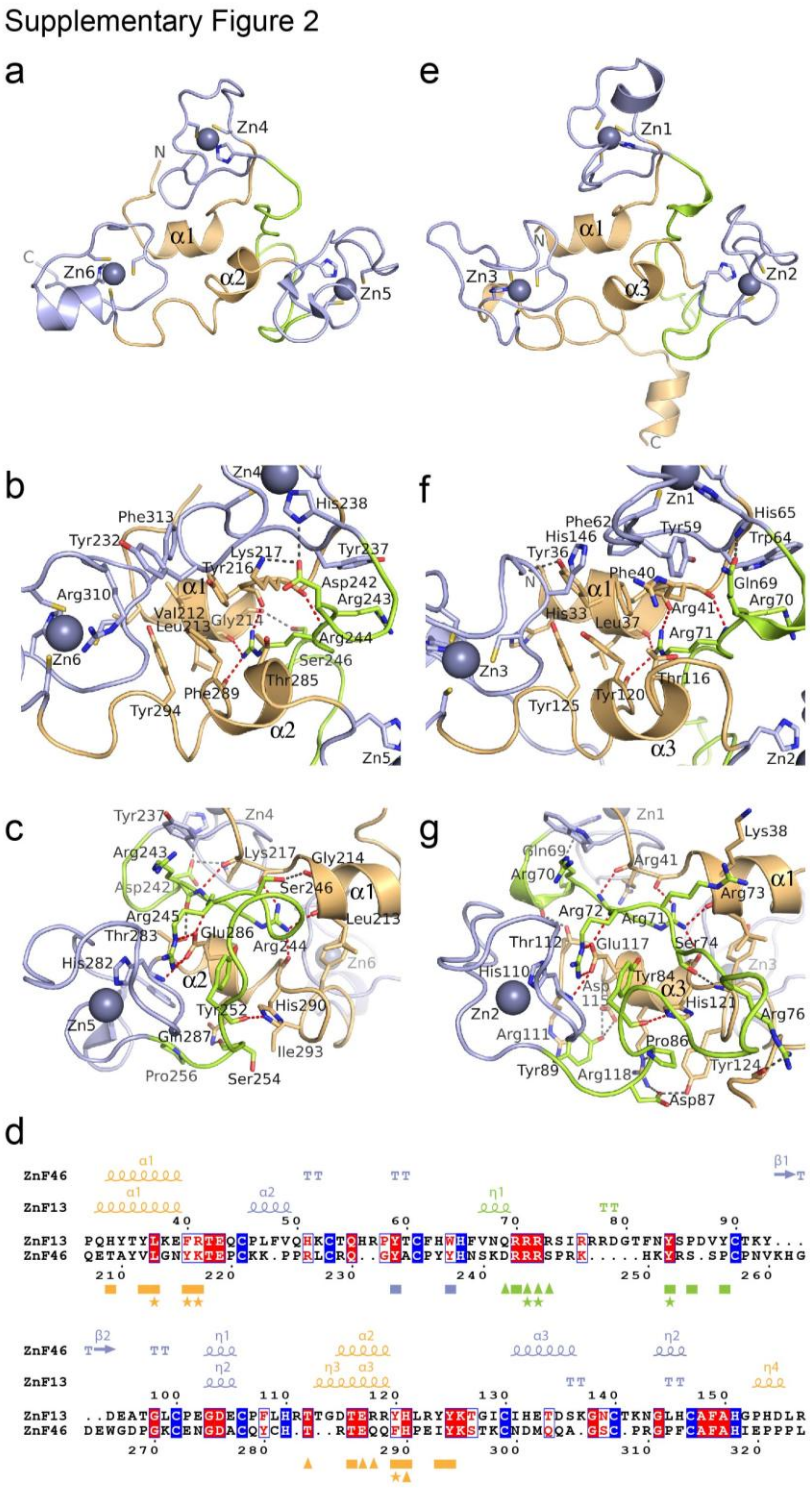


Supplementary Figure 1

The unique topologies of the CCCH ZnF clusters of Unkempt.

(**a**) Disorder probability profile of the full-length mouse Unkempt protein determined by the DISOPRED prediction method within the PSIPRED protein sequence analysis workbench (http://bioinf.cs.ucl.ac.uk/psipred/). (**b**) Structure-based sequence alignment of ZnF1–3 and ZnF4–6 tandem ZnF cassettes. Zn-coordinating cysteine and histidine residues are highlighted in blue and the secondary structure elements are shown above the sequences. Residues involved in base stacking and hydrophobic interactions with RNA nucleotides are indicated by rectangles, while asterisks and triangles indicate residues that form hydrogen bonds (H bonds) with RNA bases via their backbone functional groups and side chains, respectively. The α, βand ηsymbols refer to αhelix, βstrands and 3_{10}-helix, respectively. Helices are displayed as squiggles, βstrands are rendered as arrows, strict βturns as **TT** letters. (**c**) Superimposition of the ZnF4–6 cluster of Unkempt (light blue) and the CCCH ZnFs 5–7 of yeast Nab2 protein (red) revealing substantial differences in their topologies. (**d**) Comparison of ZnF1–3 of Unkempt (yellow) with the CCCH ZnFs 5–7 of yeast Nab2 protein (red) indicates structural diversity of both folds. Both comparisons were done by superimposing the most C-terminal ZnFs in each pair of clusters, i.e. ZnF3 and ZnF6 of Unkempt and ZnF7 of Nab2 protein.



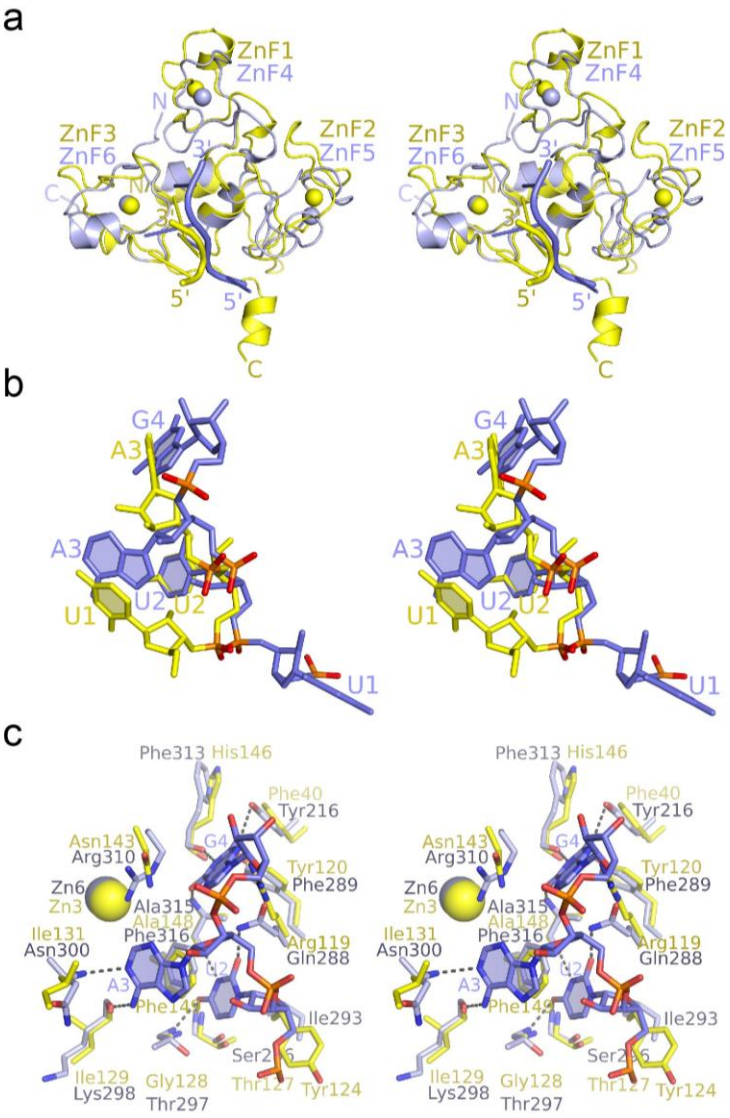
Supplementary Figure 2

Intramolecular protein-protein interactions within the ZnF1–3 and ZnF4–6 clusters of Unkempt in their RNA-bound conformation.

(**a**) Structure of the ZnF4–6 cluster shown in a ribbon representation. ZnFs 4, 5, and 6 are colored in light blue, the N-terminal loop with αhelix 1 and the linker separating ZnFs 5 and 6 with αhelix 2 are colored in light orange, and the linker separating ZnFs 4 and 5 is colored in green. (**b**) Hydrophobic and hydrogen-bonding interactions of the N-terminal αhelix 1 and the loop residues (Thr209, Val212, Leu213, Tyr216 and Lys217)

with the linker separating ZnFs 5 and 6 (Thr285, Phe289 and Tyr294), and with the linker separating ZnFs 4 and 5 (Asp242, Arg244 and Ser246). (**c**) Hydrogen-bonding interactions and van der Waals contacts of the linker separating ZnFs 4 and 5 (Arg244, Arg245, Ser246, Tyr252, Ser254 and Pro256) with the N-terminal loop and αhelix 1 (Leu213, Gly214, Tyr216 and Lys217), and with the linker between ZnFs 5 and 6 (Thr283, Glu286, Gln287, His290 and Ile293). (**d**) Structure-based sequence alignment of the ZnF1–3 and ZnF4–6 clusters. Cysteine and histidine residues coordinated to Zn are highlighted in blue. Secondary structure elements of ZnF1–3 and ZnF4–6 are shown above the sequences and are colored as in **a** and **b**. Residues involved in intramolecular hydrophobic interactions (rectangles) and hydrogen-bonding via their backbone functional groups (asterisks) or side chains (triangles) in both domains are color-coded for the ZnF regions, the N- and C-termini, and for the intervening linkers. (**e**) Structure of the ZnF1–3 cluster shown in ribbon representations. ZnFs 1, 2, and 3 are colored in light blue, the N- and C-termini and the linker separating ZnFs 2 and 3 with αhelix 3 are colored in light orange, and the linker separating ZnFs 1 and 2 is colored in green. (**f**) Hydrophobic and hydrogen-bonding interactions of the N-terminal αhelix 1 and loop residues (His33, Tyr36, Leu37, Phe40 and Arg41) with the linker between ZnFs 2 and 3 (Thr116, Tyr120 Tyr125), and with the linker separating ZnFs 1 and 2 (Arg71). (**g**) Hydrogen-bonding interactions and van der Waals contacts of the linker separating ZnFs 1 and 2 (Arg71, Arg72, Arg73, Ser74, Arg76, Tyr84, Pro86, Asp87 and Tyr89) with the N-terminal loop and αhelix 1 (Leu37, Lys38, Phe40 and Arg41), and with the linker between ZnFs 2 and 3 (Thr112, Glu117, Arg118, His121 and Tyr124). Hydrogen bonds that are equivalent between ZnF1–3 and ZnF4–6 are colored red. Cysteine and histidine side chains coordinated to Zn atoms (light blue balls) are shown in stick representation.

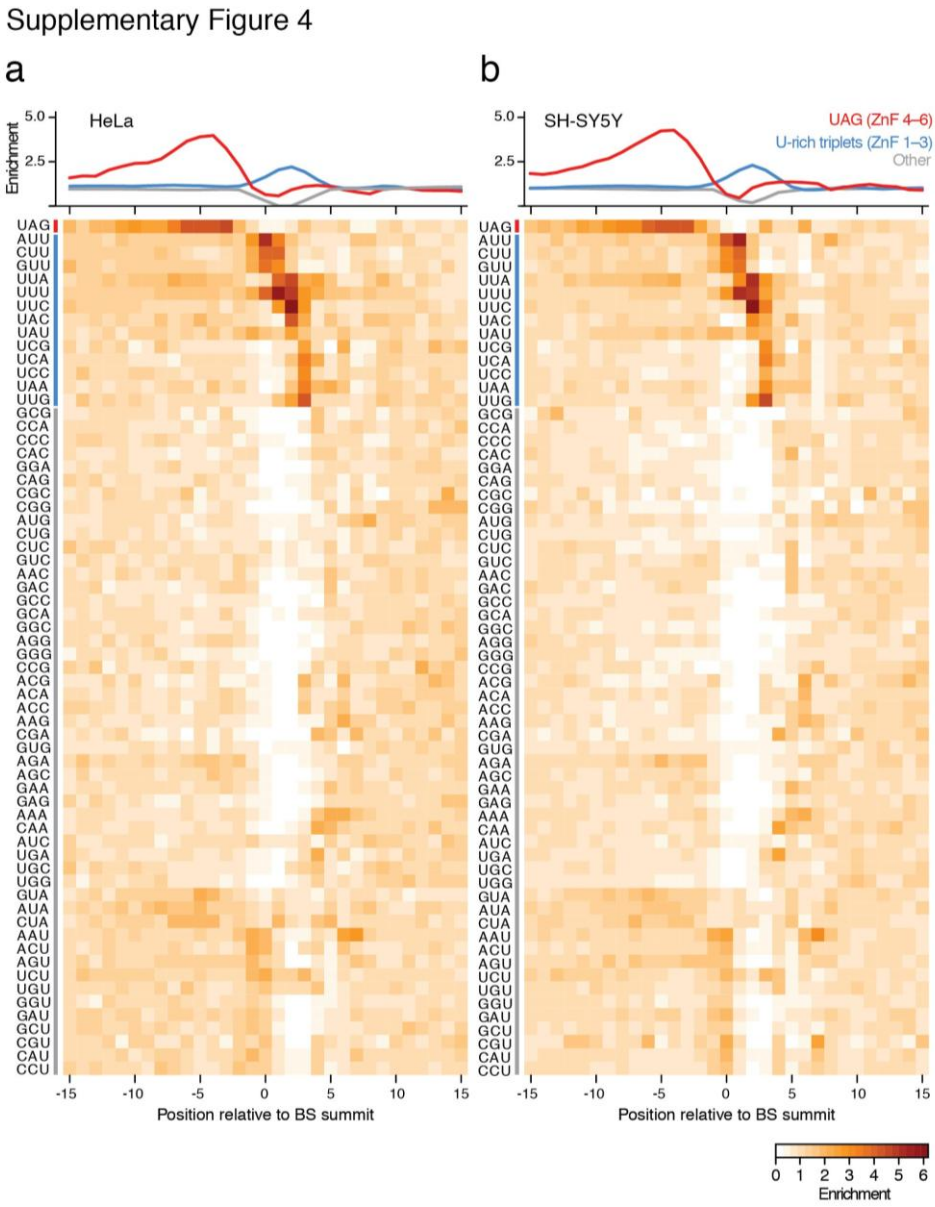
Supplementary Figure 3



Supplementary Figure 3

Stereo views comparing the structures of ZnF4–6 and ZnF1–3 bound to their RNA substrates.

(**a**) Superimposition of structures of Unkempt's ZnF1–3 bound to a U1-U2-A3 RNA element (yellow) and ZnF4–6 bound to a U1-U2-A3-G4 RNA element (light blue) showing similarities in spatial arrangement of both clusters and in the positions of the bound RNA substrates. (**b**) Conformations of the bound RNA substrates in the ZnF1–3 and ZnF4–6 complexes as shown in panel (a). **(c)** Superimposition of the RNA-contacting amino acid residues in their orientations as shown in the complexes in panel (a). The UAG motif (dark blue) bound on the surface of the ZnF4–6 cluster (light blue) highlighting the key residues contacting the RNA. Positions of corresponding residues in the ZnF1–3 cluster are shown in yellow.

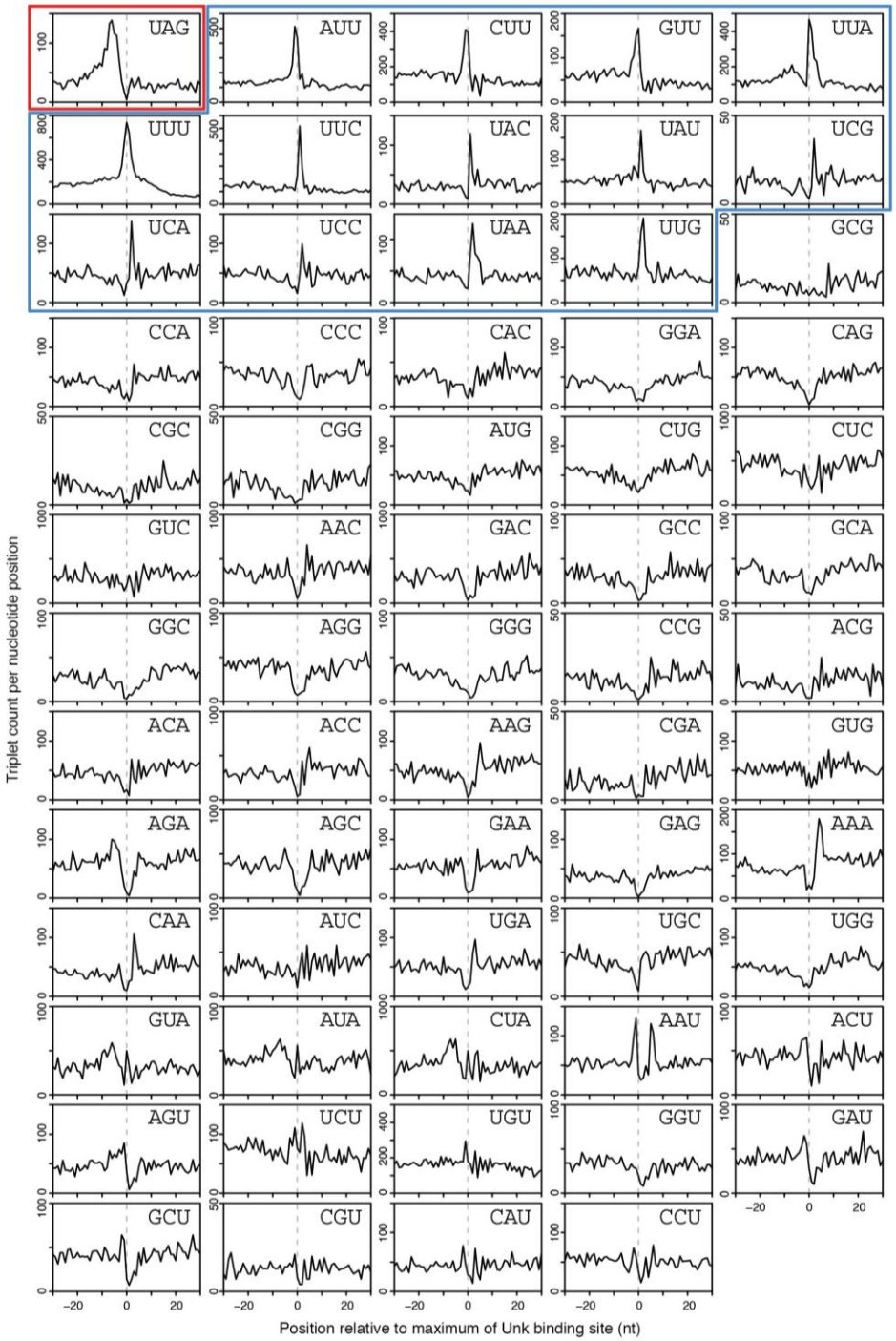


Supplementary Figure 4

Base composition of the U-rich motif in HeLa and SH-SY5Y cells.

Heatmaps illustrate the positional frequency of each of the 64 possible trimers within Unkempt binding sites between 15 nts upstream and downstream of the binding site maxima in HeLa (**a**) and SH-SY5Y cells (**b**). Traces of individual trimers are displayed in the same order as in Fig. 3a. Plots above each heatmap profile the mean enrichment of different clusters of triplets color-coded as shown in the heatmap (UAG in red, the 13 enriched U-rich triplets in blue, and all other triplets in gray). The scale below the heatmap indicates fold enrichment over the median triplet frequency in a 103-nt window around the binding site maxima.

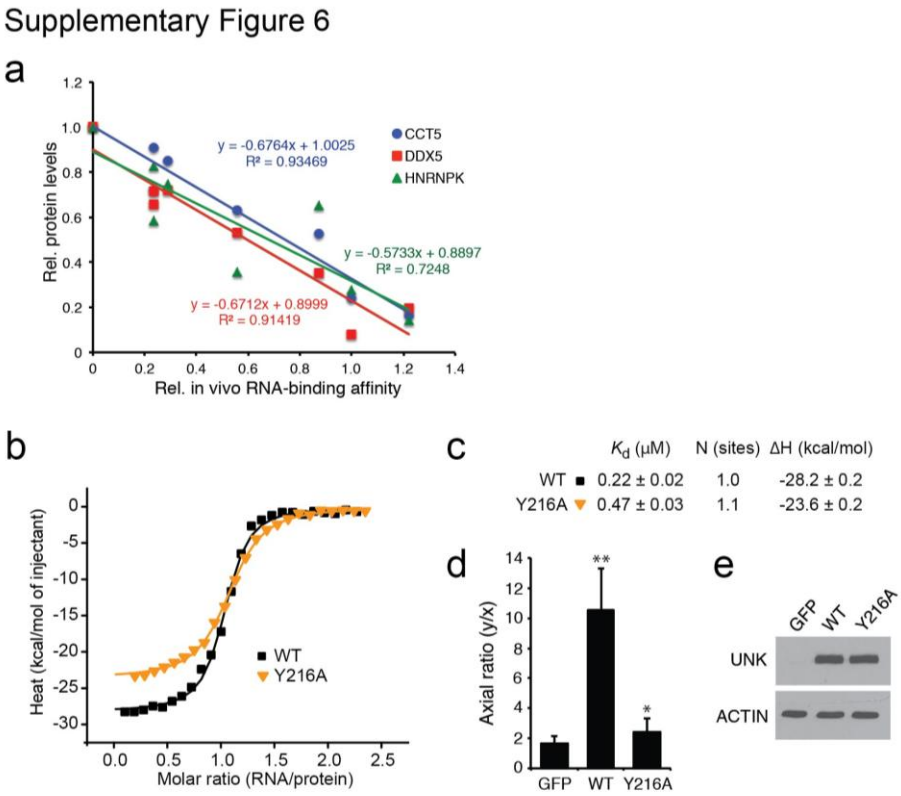
Supplementary Figure 5



Supplementary Figure 5

Profiles of individual RNA trimers within Unkempt-binding sites in the mouse embryonic brain.

Each of the 64 profiles indicates the positional frequency of a trimer in a 101-nt window around the binding site maxima. The different clusters of triplets are color-coded: UAG in red and the 13 enriched U-rich triplets in blue.

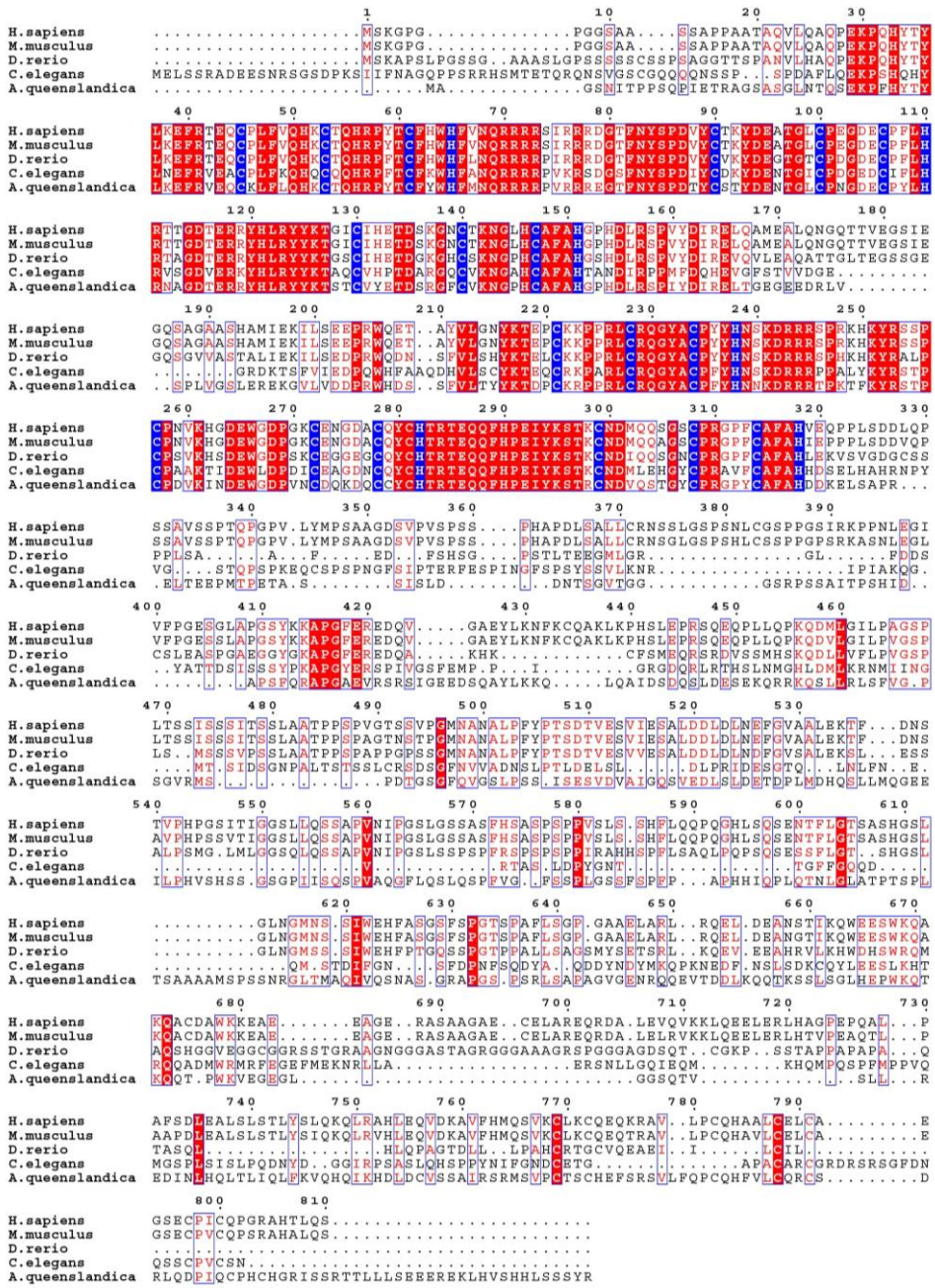


Supplementary Figure 6

Effects of structure-guided mutations on RNA-binding affinity and protein translation.

(**a**) Anti-correlation between the in vivo RNA-binding affinity of Unkempt (Fig. 5d) and steady-state protein levels shown in Fig. 6a. Relative protein levels are quantified band intensities normalized to the corresponding actin control. All values are compared with no Dox control. Also shown are color-coded equations of linear regression trendlines with corresponding correlation coefficients for each target protein. (**b-e**) Impact of Y216A mutation on RNA-binding affinity of Unkempt. (**b**) Binding of the wild-type or Y216A mutant recombinant ZnF1–6 of Unkempt to *HSPA8* 18-mer RNA. Binding affinities were measured by ITC. (**c**) Quantification of the results shown in **b**. (**d**) The morphologies of HeLa cells inducibly expressing GFP alone or GFP and either wild-type or Y216A Unkempt mutant were quantified at 48 hours of Dox treatment by calculating their axial ratios (see also Fig. 6c). The results are compared with GFP control. Error bars, s.d. (n = 30 GFP-positive cells). **P* < 0.0005; ***P* < 0.00001 by two-tailed Student's t-test. (**e**) Immunoblot analysis showing similar expression levels of the induced Unkempt protein in cells used for the morphologic analysis in **d**.

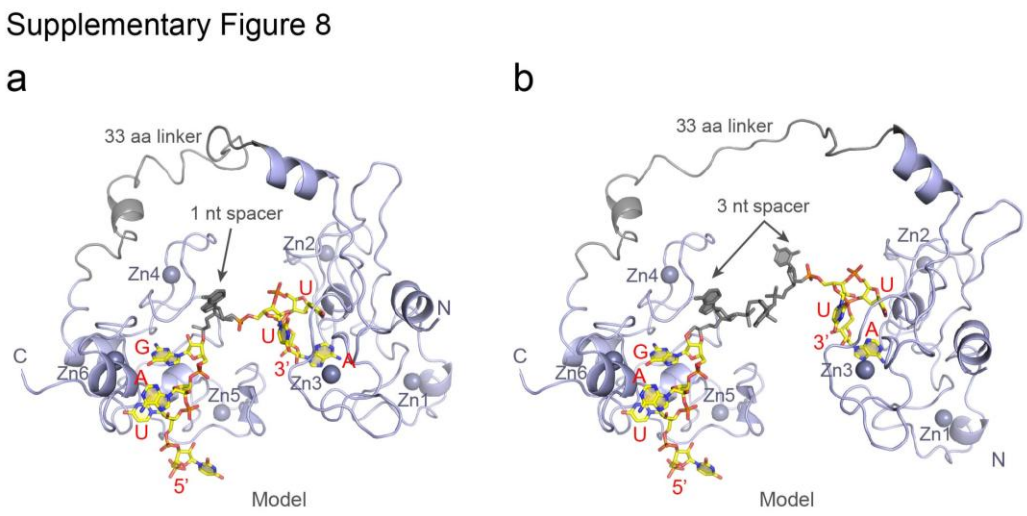
Supplementary Figure 7



Supplementary Figure 7

Sequence conservation of the Unkempt protein across species.

Full-length sequence of each Unkempt ortholog is shown. Sequence numbering corresponds to consecutive residue position in human (*H. sapiens*) and mouse (*M. musculus*) Unkempt orthologs. White letters on blue background highlight Zn-coordinating cysteine and histidine residues, white letters on red background indicate identical residues, red letters on white background indicate residues with similar physicochemical properties, and blue frames indicate global similarity. Multiple sequence alignment was performed using the Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and re-drawn using the ESPript 3.0 software (http://espript.ibcp.fr/ESPript/ESPript/index.php).



Supplementary Figure 8

Molecular models of RNA recognition by the six CCCH ZnFs of Unkempt.

Molecular models of RNA recognition by the six CCCH ZnFs of Unkempt. Two models illustrate our proposed binding mode of the ZnF1–6 domain to different RNA substrates in which the UAG and the UUA motifs are separated by spacers of different lengths, one nucleotide in the model shown in panel **a**, and three nucleotides in the model shown in panel **b**. The bound RNA nucleotides are colored yellow with the nitrogens in blue, oxygens in red, and phosphorous in orange. The modeled 1- and 3-nt spacers between the bound UAG and UUA motifs are colored gray. The protein and RNA molecules are drawn to scale.

Supplementary Notes

Structure of the ZnF4–6 cluster of Unkempt.

All three ZnFs within the ZnF4–6 cluster adopt a similar arrangement of the zinc (Zn)-coordinating residues, but with different spacing between the cysteine residues in direct contact with the Zn (Supplementary Fig. 2a,d). Each of the three ZnFs exhibits a different peptide structure between the first two Zn-coordinating cysteine residues; ZnF5 entails a short two-stranded β-sheet, ZnF6 forms a short α-helix, and ZnF4 lacks any particular secondary structure. The relative orientations of these ZnFs are determined by extensive interactions between the regions connecting the ZnFs: the N-terminal α-helix 1, the loop connecting ZnFs 4 and 5, and the linker between ZnFs 5 and 6 containing the α-helix 2 (Supplementary Fig. 2a-c). The hydrophobic core, located on or near α-helixes 1 and 2 (10 conserved residues, Supplementary Fig. 2b), is formed by a combination of aromatic and hydrophobic residues, and the overall fold is additionally stabilized by multiple intramolecular hydrogen-bonding and van der Waals interactions.

Structure of the ZnF1–3 cluster and comparison with ZnF4–6.

The ZnF1–3 cluster assumes a compact fold similar to ZnF4–6, but with a structured C-terminus, which is largely disordered in the ZnF4–6 cluster (Fig. 2e and Supplementary Fig. 2a,e). Although the individual ZnFs 1, 2, and 3 can be superimposed with ZnFs 4, 5, and 6, respectively, the two clusters share only 23% sequence identity, vary in spacing between Zn-coordinating residues and linker lengths (Supplementary Fig. 1b), and display substantial structural differences (root-mean-square deviation of 3.2 Å for 83 matching Cα atoms of ZnF1–3, amino acids 32–153; Supplementary Figs. 2d and 3a). Nevertheless, the spatial arrangement of ZnF1–3 within a single structural unit is conferred by essentially the same core interactions as in ZnF4–6, engaging the N-terminal α-helix 1, the loop connecting ZnFs 1 and 2, and the linker between ZnFs 2 and 3 comprising the α-helix 3 (Supplementary Fig. 2e-g).

RNA recognition by the ZnF4–6 and ZnF1–3 clusters.

The hydrophobic clefts of ZnF4–6 and ZnF1–3 clusters each accommodates three RNA bases, whereas the RNA sugar–phosphate backbones appears largely exposed to the

solvent and surrounded by positively charged surface residues (Fig. 3a,d). The key residues of ZnF4–6 that directly contact UAG RNA bases through either stacking, hydrophobic or hydrogen bonding interactions (Phe316, Ala315, Gln288, Phe289, Arg310, and Tyr216; Fig. 2b) are strictly conserved among species, suggesting that a similar sequence specificity may be used by Unkempt in lower metazoans (Supplementary Fig. 7). Structurally equivalent key residues in ZnF1–3 (Phe149, Ala148, Arg119, Tyr120, Asn143 and Phe40) contact the UUA motif (Fig. 2d). Bases U1 and U2 in the ZnF1–3 complex are stacked with the conserved Phe149 ring, analogous to the Phe316 ring in ZnF6, and are recognized through hydrogen-bonding interactions with the functional groups of the ZnF3 backbone, equivalent to the interactions of bases U2 and A3 with the ZnF6 backbone (Fig. 3b,e). Base A3 in a *syn* conformation is inserted into the shallow groove between the ZnF3 and the α-helix 3, stacked between the side chains of Asn143 and Arg119 while also forming two hydrogen bonds with the Ala148 backbone amide and the Tyr120 side chain, similar to the hydrogen bonding interactions of G4 with the ZnF6 (Fig. 3c,f).

Contrast between specific recognition by ZnF4–6 and promiscuous recognition by ZnF1–3.

The structure of ZnF4–6 bound to the UAG element shows that ZnF4–6 structurally distinguishes UAG from other triplets through binding of uracil and guanine bases in specific pockets and forming multiple base-specific hydrogen-bonding interactions predominantly with its peptide backbone moieties. Replacement of each of the three bases with any other base would result in a loss of hydrogen bonds and unfavorable contacts and/or steric hindrance. ZnF1–3 binds to the UUA element using basically the same surface and a similar set of contact residues, but the bound RNA adopts a different backbone conformation with flipped-over first uracil and the adenine in *syn* conformation (Supplementary Fig. 3b,c). The possible variability in the RNA backbone conformation might explain how ZnF1-3 could accommodate diverse U-rich trinucleotides. To understand the relative promiscuity of ZnF1–3, we modeled the ZnF4–6 – bound UAG trinucleotide conformation on the surface of the ZnF1–3 cluster (Fig. 2e and Supplementary Fig. 3). We did not see any significant differences between the contacts of

the first two nucleotides of either UUA or UAG triplets with the ZnF1–3, apart from one lacking hydrogen bond with the Gln288 side chain (Supplementary Fig. 3c). We conclude that both UU and UA sequences can be accommodated on the surface of the ZnF1–3 cluster, assuming different backbone structures. The guanine base could in principle also be accommodated in the third position upon minor adaptation of the narrower ZnF1–3 pocket (note the clash with Arg119 and Tyr120 and the loss of one hydrogen bond due to Tyr216 replacement with Phe40, Supplementary Fig. 3c). These minor differences between the contacts of either UUA or UAG triplets with the ZnF1–3 or ZnF4–6 clusters could explain the experimentally determined difference in the RNA binding affinity and point to the propensity of CCCH ZnF proteins to interact with RNA via adaptation of the protein-binding interface and of the RNA conformation. Similar adaptations were observed in some single stranded nucleic acid-binding proteins that can specifically recognize various sequences using almost nearly identical networks of hydrogen bonds (Clery, A., Boudet, J. & Allain, F.H. Single-stranded nucleic acid recognition: is there a code after all? *Structure* **21**, 4-6 (2013).

Cleavage events during the crystallization of the ZnF1–6 domain.

We have confirmed by SDS-PAGE gel analysis that the protein in the crystallization solution was cleaved at the site of the flexible linker peptide, but have not checked if the RNA was also trimmed during the crystallization process. However, we clearly see 4 out of the 18 nts in the electron density map with the remaining part of the sequence either cleaved or disordered in the crystal. In addition, crystal packing excludes the possibility of the presence of ZnF1–3 bound to the RNA in the crystal. Given the small size and the low level of reproducibility of these crystals – because of the spontaneous proteolytic digestion – we do not believe that it would be possible to analyze the precise state of the RNA in these crystals.

Use of the entire ZnF1–6 domain and the 18-nt RNA substrate for K_d determination (Fig. 5).

We chose not to use the individual ZnF1–3 and ZnF4–6 clusters to measure the effects of mutations on RNA binding because of their weak binding to the 18-nt target (8-

to 25-fold lower affinity than the entire ZnF1–6 domain; Fig. 1c). Furthermore, we measured an even lower binding affinity for the ZnF4–6 cluster bound to a 7-nt *HSPA8* mRNA sequence CUUAGAA, which contains only one binding motif (K_d = 3.6 ± 0.4 μM). We thus reasoned that the effect of individual mutations on separate ZnF clusters would be much less accurately assessed, as they would further decrease the binding affinity, likely into an inestimable range.

Putative roles of ZnFs 1, 2, 4 and 5.

Our crystal structures of the ZnF clusters (Fig. 2) suggest that the overall topology of each cluster, including the shape of each RNA-binding hydrophobic cleft, is at least to some extent determined by the presence of the first and the second ZnF in each cluster (i.e. ZnFs 1 and 2 in ZnF1–3, and ZnFs 4 and 5 in ZnF4–6). Although ZnF3 and ZnF6 might be able to recognize RNA on their own, we do not believe that the specificities or affinities of these interactions would be identical to those of the entire clusters. Apart from these structure-stabilizing roles within each cluster, the functional importance of ZnFs 1, 2, 4 and 5 may include their participation in posttranscriptional control, including, for example, co-recruitment of other proteins and organization of RNA – Unkempt complexes into higher-order structures.