LETTER

doi:10.1038/nature13999

Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch

Alaji Bah[1,2], Robert M. Vernon[1,2], Zeba Siddiqui[1], Mickaël Krzeminski[1,2], Ranjith Muhandiram[2,3], Charlie Zhao[1], Nahum Sonenberg[4], Lewis E. Kay[1,2,3,5] & Julie D. Forman-Kay[1,2]

Intrinsically disordered proteins play important roles in cell signalling, transcription, translation and cell cycle regulation[1,2]. Although they lack stable tertiary structure, many intrinsically disordered proteins undergo disorder-to-order transitions upon binding to partners[3,4]. Similarly, several folded proteins use regulated order-to-disorder transitions to mediate biological function[5,6]. In principle, the function of intrinsically disordered proteins may be controlled by post-translational modifications that lead to structural changes such as folding, although this has not been observed. Here we show that multisite phosphorylation induces folding of the intrinsically disordered 4E-BP2, the major neural isoform of the family of three mammalian proteins that bind eIF4E and suppress cap-dependent translation initiation. In its non-phosphorylated state, 4E-BP2 interacts tightly with eIF4E using both a canonical YXXXXLΦ motif (starting at Y54) that undergoes a disorder-to-helix transition upon binding and a dynamic secondary binding site[7–11]. We demonstrate that phosphorylation at T37 and T46 induces folding of residues P18–R62 of 4E-BP2 into a four-stranded β-domain that sequesters the helical YXXXXLΦ motif into a partly buried β-strand, blocking its accessibility to eIF4E. The folded state of pT37pT46 4E-BP2 is weakly stable, decreasing affinity by 100-fold and leading to an order-to-disorder transition upon binding to eIF4E, whereas fully phosphorylated 4E-BP2 is more stable, decreasing affinity by a factor of approximately 4,000. These results highlight stabilization of a phosphorylation-induced fold as the essential mechanism for phospho-regulation of the 4E-BP:eIF4E interaction and exemplify a new mode of biological regulation mediated by intrinsically disordered proteins.

Proteins exist in a continuum of states, from ensembles of interconverting conformers for intrinsically disordered proteins (IDPs) to stable structures for folded proteins[2,4,12,13]. Although many undergo subtle conformational rearrangements in response to biological signals, some undergo larger changes: that is, disorder-to-order or order-to-disorder transitions[4] that can be mediated by interactions with ligands. Post-translational modifications (PTMs) such as phosphorylation can also potentially induce such transitions. While phosphorylation can convert the folded oligomerization domain of nucleophosmin into a disordered monomer[14,15], PTMs have been generally known only to stabilize or destabilize individual secondary structural elements in IDPs[16–18]. Thus, PTM-mediated folding of a protein domain, as shown here for the phosphorylation-induced folding of 4E-BP2, is a novel regulatory mechanism for IDPs.

Here we study the phospho-regulation of the 4E-BP:eIF4E interaction[19]. The 4E-BPs compete with eIF4G for eIF4E binding to prevent cap-dependent translation initiation by using similar canonical eIF4E-binding YXXXXLΦ motifs[20,21]. Non-phosphorylated or minimally phosphorylated 4E-BPs interact tightly with eIF4E, while the binding of highly phosphorylated 4E-BPs is much weaker and can be outcompeted by eIF4G. T37 and T46 are known to be phosphorylated first, followed by T70 and S65 (ref. 22). Interestingly, in their free states, highly phosphorylated 4E-BPs are very stable while non- or minimally phosphorylated 4E-BPs are targeted for degradation with ubiquitination at K57 (within the YXXXXLΦ motif) by the KLHL25-CUL3 E3 ligase[23]. However, there is no consensus on how phosphorylation regulates binding to eIF4E or affects the stability of 4E-BPs. Phosphorylation may result in electrostatic repulsion with the negative surface of eIF4E[21] or S65 phosphorylation could inhibit binding by destabilizing the YXXXXLΦ motif helical structure[24]. The inability of glutamic acid phospho-mimetics to weaken 4E-BP:eIF4E binding has also suggested that additional PTMs may be required[25].

To address the phospho-regulatory mechanism, we have used NMR, isothermal titration calorimetry (ITC) and mutagenesis. Figure 1 and Extended Data Fig. 1 show structural and dynamic properties of wild-type (WT) 4E-BP2 uniformly phosphorylated at T37, T46, S65, T70 and S83 (Fig. 1a). IDPs such as non-phosphorylated 4E-BP2 have intense peaks with narrow ^{1}HN chemical shift dispersion[26]. Unlike other IDPs for which phosphorylation only causes downfield chemical shifts of phosphorylated residues[27], 4E-BP2 phosphorylation induces widespread downfield and upfield chemical shifts for residues spanning T19–R62, suggesting folding upon phosphorylation (Fig. 1b). Notably, G39 and G48, the first glycines in the six-residue repeat sequences (TTPGGT) containing two of the five phosphorylation sites, show dramatic downfield ^{1}HN chemical shift changes (Fig. 1b, inset). Peaks for the rest of the protein are intense with narrow ^{1}HN chemical shift dispersion, indicating that these residues remain disordered. Like its non-phosphorylated state[11], phosphorylated 4E-BP2 exchanges between major and minor conformers, probably from *cis–trans* isomerization of the multiple prolines and between unfolded and partly folded states for the folded region (see below). Lower [^{1}H]–^{15}N nuclear Overhauser effect (NOE) values are expected for IDPs because of their rapid motions[26,28]. Elevated [^{1}H]–^{15}N NOEs for most residues from T19–R62 (Fig. 1c), along with the large increase in chemical shift dispersion and the high amide proton temperature coefficients indicative of intramolecular hydrogen bonding[29] (Extended Data Fig. 1d), provide strong evidence that this region folds upon phosphorylation, while the rest of the protein remains disordered.

We made alanine mutations to mimic *in vivo* phosphorylation states of 4E-BP2 (ref. 22) (Extended Data Fig. 2). No significant change in global dispersion was observed for 4E-BP2 phosphorylated only at S65/T70/S83 (Extended Data Fig. 2a), demonstrating that it remains disordered, while phosphorylating T37 and T46 (pT37pT46) induces a 4E-BP2 fold identical to phosphorylated wild type (pWT) (Extended Data Fig. 2b). Interestingly, when phosphorylated individually, pT37 or pT46 result in a partly folded state, with some chemical shift changes indicative of ordered structure (pT37) and the presence of one β-turn leading to one characteristic downfield shifted glycine peak (G39 for pT37, G48 for pT46; Extended Data Fig. 2c, d). Thus, phosphorylation of both T37 and T46 is necessary and sufficient for phosphorylation-induced folding of 4E-BP2.

We determined the structure of P18–R62 using CS-Rosetta[30] including initially only chemical shifts, and subsequently adding NOEs (see

[1]Molecular Structure and Function Program, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. [2]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]Department of Biochemistry and Goodman Cancer Research Centre, McGill University, Montréal, Quebec H3G 1Y6, Canada. [5]Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada.

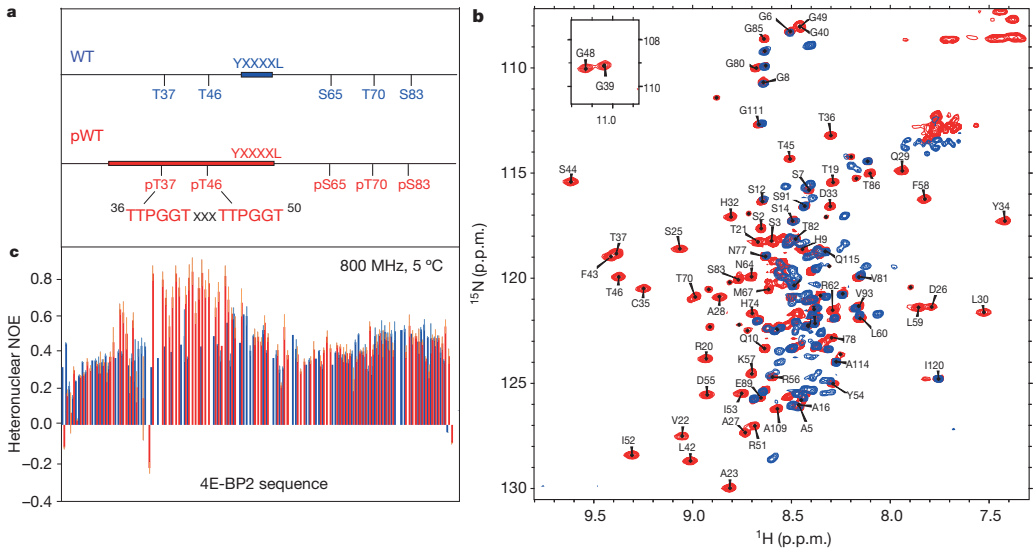


Figure 1 | Effects of phosphorylation on the structural and dynamic properties of 4E-BP2. **a**, Schematic representation of 4E-BP2 showing the relative positions of the phosphorylation sites, the canonical eIF4E binding site (thick blue bar) and the region which undergoes phosphorylation-induced folding (thick red bar). **b**, **c**, Overlay of (**b**) ^{1}H–^{15}N heteronuclear single quantum correlation (HSQC) spectra and (**c**) ^{1}H–^{15}N NOE values at 800 MHz, 5 °C for non-phosphorylated WT (blue) and phosphorylated WT (red), respectively. Missing data represent prolines and residues that are too overlapped/weak to be accurately quantified. Errors (standard deviations) around the average values are based on multiple repeats ($n = 3$) of the experiment.

Methods, Extended Data Figs 3–6 and Extended Data Table 1a, b). The domain is a four-stranded β-fold (Fig. 2), with β1(T19–I24), β2(C35–pT37), β3(L42–T45), β4(R51–R56) and a 3_{10} helix (A27–Q29). pT37 and pT46 are central to a network of hydrogen bonds that stabilize tight β-turns composed of the identical pTPGGT motifs connecting strands β2 with β3 (pT37–T41) and β3 with β4 (pT46–T50). The hydrogen bonding correlates well with ^{1}HN temperature coefficients (Extended Data Fig. 1d). G39 and G48 amide protons hydrogen bond to T37 and T46 phosphate groups, respectively, explaining their unexpectedly large downfield chemical shifts (Fig. 1a, inset, and Extended Data Fig. 2b). The YXXXXLϕ eIF4E-binding motif forms part of strand β4 with Y54 largely buried within a hydrophobic cluster also involving V22, I24, L30, P31, Y34 and I52 (Fig. 2a). The phosphorylation-induced structure provides a mechanism by which phosphorylation reduces eIF4E binding by sequestering the helical YXXXXLϕ motif into a β-strand. Its burial stabilizes the β-fold; phosphorylated Y54A/L59A [p(Y54A/L59A)] retains the β-turns, evident from the characteristic downfield ^{1}HN shifts of G39 and G48, but no longer folds (Fig. 3a). A stable fold sequestering YXXXXLϕ could also block ubiquitination of K57, thereby preventing degradation *in vivo*[23].

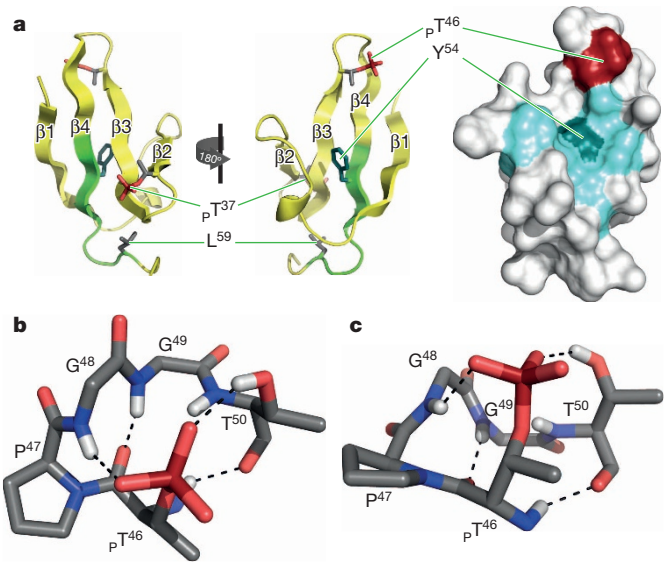


Figure 2 | Phosphorylation-induced structure of the major state of residues R18–R62 of 4E-BP2. **a**, Cartoon (left) and surface (right) representations of the solution NMR structure. Phosphorylated residues, pT37 and pT46 (red), the surface formed by residues of the hydrophobic cluster (right, cyan) and Y54 (cyan stick representation (left) or dark cyan surface (right)) are shown. The residues in the canonical eIF4E binding site (YXXXXLΦ) are shown in green (left), demonstrating the binding-incompatible extended structure. **b**, **c**, Hydrogen-bonding networks formed by the pTPGGT motifs. Note that pT46–T50 is shown in two different views and the same hydrogen bonds are formed for pT37–T41.

ITC (Extended Data Table 2 and Extended Data Fig. 7) and NMR data were obtained on 4E-BP2 variants to probe the role of electrostatics in folding and reducing eIF4E affinity. Unlike pT37pT46, neither T37D/T46D nor T37E/T46E (phospho-mimics) showed evidence of folding (Extended Data Fig. 2e, f). In contrast to pT37pT46, which has reduced affinity (dissociation constant (K_d) = 267 ± 32 nM), T37D/T46D and T37E/T46E bind tightly to eIF4E (K_d = 3.89 ± 1.1 nM and 4.37 ± 0.8 nM, respectively), similar to non-phosphorylated 4E-BP2 (K_d = 3.20 ± 0.6 nM). Thus acidic residues do not mimic phosphorylation by inducing folding of 4E-BP2 or reducing eIF4E affinity. The binding affinity of pS65pT70pS83 4E-BP2 for eIF4E was very tight (K_d = 11.3 ± 2.9 nM), as expected since a folded structure is not formed (Extended Data Fig. 2a). Strikingly, five-phospho protein (pWT) decreases affinity by about three orders of magnitude (K_d = 12,320 ± 600 nM). Valines were substituted for the first glycines in the TPGGT motifs (G39V/G48V) of 4E-BP2 to disrupt hairpin formation through steric contacts and prevent folding. As predicted from ΔΔG calculations (Extended Data Table 1c), folding was not induced when G39V/G48V was fully phosphorylated at all five sites [p(G39V/G48V)] (Fig. 3b). Notably, the affinity of the p(G39V/G48V) for eIF4E is high (K_d = 36.1 ± 3.5 nM), only an order of magnitude weaker than that of non-phosphorylated 4E-BP2 (K_d = 3.20 ± 0.6 nM) and approximately 2.5 orders of magnitude stronger than pWT (K_d = 12,320 ± 600 nM). These data, together with unfolding energies of pT37pT46 (~2.6 kcal mol^{-1}) and pWT (~4.8 kcal mol^{-1}) estimated from K_d values (Extended Data Table 2), suggest that, although phosphorylation of only T37 and T46 is required to induce folding, phosphorylation of S65, T70 and S83 stabilizes the fold. This is possibly through long-range transient electrostatic

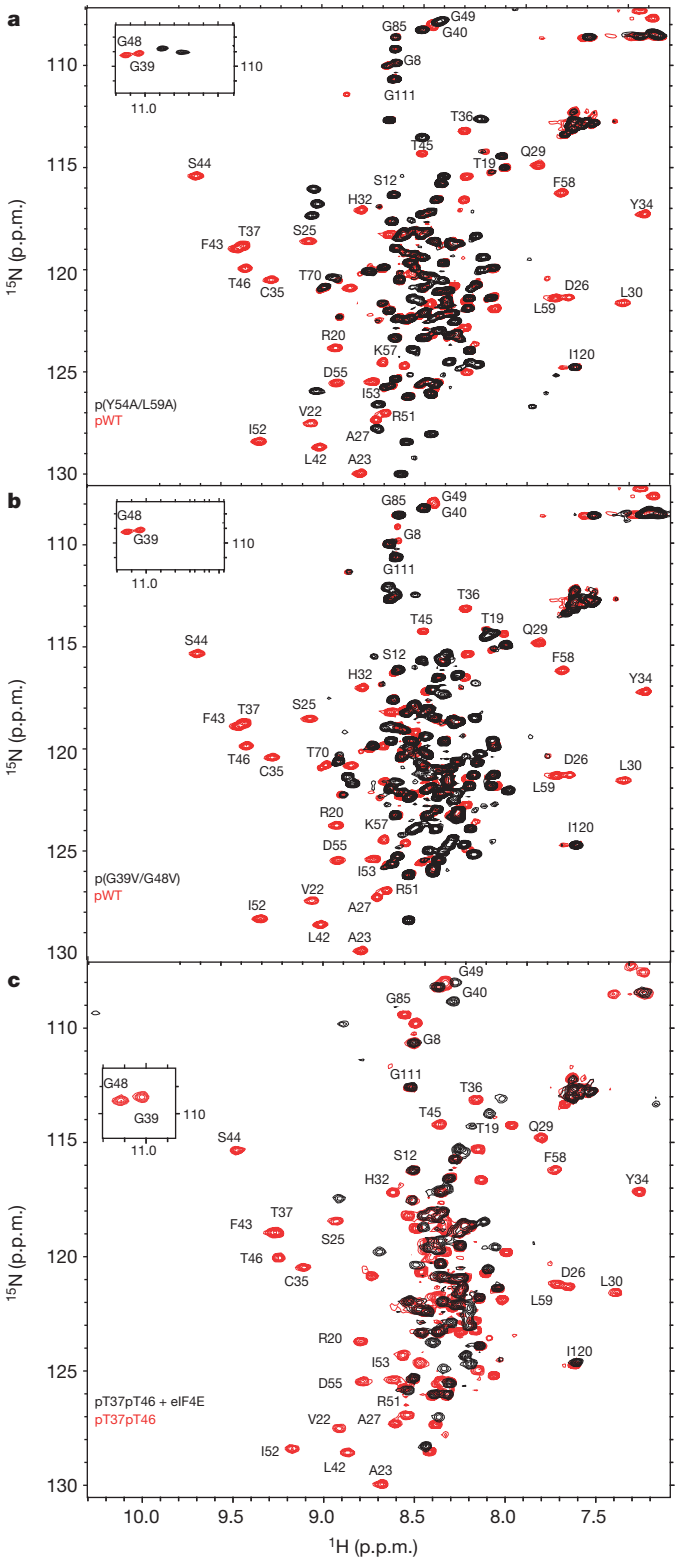


Figure 3 | Probing the structural and binding properties of phosphorylated 4E-BP2. **a**, **b**, ^{1}H–^{15}N HSQC spectra of pWT (red) overlaid with (**a**) p(Y54A/L59A) and (**b**) p(G39V/G48V). **c**, Spectrum of pT37pT46 in isolated (red) and eIF4E-bound (black) states, demonstrating an order-to-disorder transition upon eIF4E binding.

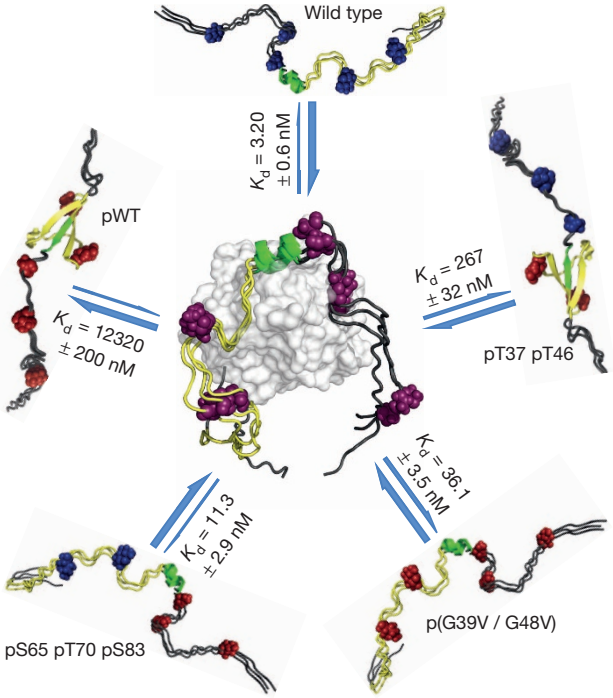


Figure 4 | Phospho-regulation of the eIF4E:4E-BP interaction. Schematic representations of different phosphorylation states of 4E-BP2 and K_d values for eIF4E interaction (grey surface). The 4E-BP2 residues spanning Q10–D90 are shown with the canonical eIF4E binding site (YXXXXLΦ, green), other residues involved in the phosphorylation-induced fold (P18–R62, yellow) and non-phosphorylated (blue) and phosphorylated (red) Ser/Thr residues. Upon phosphorylation, 4E-BP2 undergoes a disorder-to-order transition, which significantly weakens eIF4E interaction, allowing eIF4E to bind eIF4G to initiate translation. Three representations, without spanning the likely broad conformational space sampled, are shown to indicate the disorder present in 4E-BP2 regions that are not folded.

interactions between the phosphorylated acidic carboxy (C) terminus with the basic folded domain. Sequestration of the canonical eIF4E motif by pT37pT46 reduces affinity by about two orders of magnitude, and not the three orders observed for the fully phosphorylated form, possibly because of a binding-competent minor disordered state that is visible in the spectrum, providing additional evidence for lower stability of the pT37pT46 folded state. These affinities and calculations support the view that stabilization of a phosphorylation-induced folded structure plays a dominant role in weakening the eIF4E:4E-BP2 interaction, with a probable small contribution from electrostatic repulsion from eIF4E as reflected in the ten-fold reduction in affinity for p(G39V/G48V) from non-phosphorylated WT 4E-BP2.

Because the 4E-BP2 β-fold sequesters the eIF4E-binding interface, binding to eIF4E must be coupled to unfolding. Although it is folded in the absence of target, pT37pT46 undergoes an order-to-disorder transition upon binding to eIF4E (Fig. 3c), as established by the resulting poorly resolved spectrum confined to a narrow amide proton chemical shift range and by the disappearance of well dispersed folded peaks of the *apo*-state, except for a single very weak ^{1}HN glycine peak at 10.25 p.p.m. that may reflect a low population of a β-turn (Fig. 3c and Extended Data Fig. 2g). The spectrum is similar to that of non-phosphorylated 4E-BP2 (ref. 11); both complexes are disordered with significant chemical shift differences (Extended Data Fig. 2g), probably reflecting changes in interactions due to different binding affinities (3.20 ± 0.6 nM versus 267 ± 32 nM) and effects of phosphates, including low population of a β-turn.

Our study provides key insights into how the structural polymorphism of 4E-BP2 allows it to regulate translation initiation through PTM-mediated folding (Fig. 4). This mechanism establishes a new means of IDP-mediated control of biological function. Large structural changes such as folding could sequester or enhance the accessibilities of protein binding and other PTM sites or provide new interaction surfaces, thereby expanding signalling output. Importantly, PTM-induced folding has important potential impact for targeting IDP interactions for therapeutics.

Since the stability of the phosphorylation-induced 4E-BP fold is critical in controlling binding to eIF4E, small molecules that stabilize or destabilize folding of 4E-BPs are likely to be potent modulators of translation initiation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 24 October 2014.

Published online 22 December 2014.

1. Dunker, A. K. & Uversky, V. N. Signal transduction via unstructured protein conduits. *Nature Chem. Biol.* **4,** 229–230 (2008).
2. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6,** 197–208 (2005).
3. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19,** 31–38 (2009).
4. Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **1834,** 932–951 (2013).
5. Mitrea, D. M. & Kriwacki, R. W. Regulated unfolding of proteins in signaling. *FEBS Lett.* **587,** 1081–1088 (2013).
6. Schultz, J. E. & Natarajan, J. Regulated unfolding: a basic principle of intraprotein signaling in modular proteins. *Trends Biochem. Sci.* **38,** 538–545 (2013).
7. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136,** 731–745 (2009).
8. Tomoo, K., Abiko, F., Miyagawa, H., Kitamura, K. & Ishida, T. Effect of N-terminal region of eIF4E and Ser65-phosphorylation of 4E-BP1 on interaction between eIF4E and 4E-BP1 fragment peptide. *J. Biochem.* **140,** 237–246 (2006).
9. Fletcher, C. M. *et al.* 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry* **37,** 9–15 (1998).
10. Gosselin, P. *et al.* The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS. *Nucleic Acids Res.* **39,** 3496–3503 (2011).
11. Lukhele, S., Bah, A., Lin, H., Sonenberg, N. & Forman-Kay, J. D. Interaction of the eukaryotic initiation factor 4E with 4E-BP2 at a dynamic bipartite interface. *Structure* **21,** 1–11 (2013).
12. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23,** 105–116 (2010).
13. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313,** 1638–1642 (2006).
14. Mitrea, D. M. & Kriwacki, R. W. Cryptic disorder: an order-disorder transformation regulates the function of nucleophosmin. *Pac. Symp. Biocomput.* **2012,** 152–163 (2012).
15. Mitrea, D. M. *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc. Natl Acad. Sci. USA* **111,** 4466–4471 (2014).
16. Pufall, M. A. *et al.* Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309,** 142–145 (2005).
17. Theillet, F. X. *et al.* Cell signaling, post-translational protein modifications and NMR spectroscopy. *J. Biomol. NMR* **54,** 217–236 (2012).
18. Espinoza-Fonseca, L. M., Kast, D. & Thomas, D. D. Thermodynamic and structural basis of phosphorylation-induced disorder-to-order transition in the regulatory light chain of smooth muscle myosin. *J. Am. Chem. Soc.* **130,** 12208–12209 (2008).
19. Lin, T.-A. *et al.* PHAS-I as a link between mitogen-activated protein kinase and translation initiation. *Science* **266,** 653–656 (1994).
20. Mader, S., Lee, H., Pause, A. & Sonenberg, N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol. Cell. Biol.* **15,** 4990–4997 (1995).
21. Marcotrigiano, J., Gingras, A.-C., Sonenberg, N. & Burley, S. K. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3,** 707–716 (1999).
22. Gingras, A.-C. *et al.* Regulation of 4E-BP1 phosphorylation: a novel two step mechanism. *Genes Dev.* **13,** 1422–1437 (1999).
23. Yanagiya, A. *et al.* Translational homeostasis via the mRNA cap-binding protein, eIF4E. *Mol. Cell* **46,** 847–858 (2012).
24. Tait, S. *et al.* Local control of a disorder-order transition in 4E-BP1 underpins regulation of translation via eIF4E. *Proc. Natl Acad. Sci. USA* **107,** 17627–17632 (2010).
25. Oulhen, N. *et al.* A variant mimicking hyperphosphorylated 4E-BP inhibits protein synthesis in a sea urchin cell-free, cap-dependent translation system. *PLoS ONE* **4,** e5070 (2009).
26. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104,** 3607–3622 (2004).
27. Selenko, P. *et al. In situ* observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nature Struct. Mol. Biol.* **15,** 321–329 (2008).
28. Klein-Seetharaman, J. *et al.* Long-range interactions within a nonnative protein. *Science* **295,** 1719–1722 (2002).
29. Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* **21,** 249–261 (2001).
30. Vernon, R., Shen, Y., Baker, D. & Lange, O. F. Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J. Biomol. NMR* **57,** 117–127 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Augustyniak, Z. Bozoky, V. Csizmok, J. Dawson and P. Farber for discussions, and A. Hansen for help in NMR data processing. A. Chong, R. Hudson and H. Lin are acknowledged for their technical expertise. This work was funded by grants from the Canadian Institutes of Health Research (MOP-114985, MOP-119579) and the Canadian Cancer Society to J.D.F.-K. A.B. was partly supported by a Restracomp award from the Hospital for Sick Children and a post-doctoral fellowship from the Canadian Institutes of Health Research (CIHR). R.M.V. was partly supported by a post-doctoral fellowship from the CIHR Strategic Training Program in Protein Folding and Interaction Dynamics. Z.S. was partly supported by the Summer Research Program at the Hospital for Sick Children.

Author Contributions A.B. and J.D.F.-K. designed experiments. A.B., Z.S. and N.S. contributed reagents. A.B., L.E.K. and R.M. performed NMR experiments. R.M.V. and M.K. performed structure calculations. A.B., Z.S. and C.Z. performed biochemical experiments. A.B., R.M.V., L.E.K. and J.D.F.-K. analysed data. A.B., R.M.V., M.K., N.S., L.E.K. and J.D.F.-K. wrote and edited the paper.

Author Information Chemical shifts of non- and fully phosphorylated 4E-BP2 have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession numbers 19114 (ref. 11) and 19905, respectively. The coordinates for the folded state of phosphorylated 4E-BP2 has been deposited in the Protein Data Bank (accession number 2MX4). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.F.-K. (forman@sickkids.ca).

METHODS

Protein expression and purification. Small ubiquitin-like modifier (Sumo) fusion constructs of human eIF4E or wild type/mutant 4E-BP2 were expressed and purified to homogeneity as previously described[11]. Expression and purification of activated His-tagged Erk2 used a protocol and plasmid co-expressing Erk2 and MEK1 obtained from Attila Remenyi at Eötvös Loránd University.

Phosphorylation of 4E-BP2. All 4E-BP2 constructs were phosphorylated to homogeneity with Erk2 using a dialysis technique. Briefly, each phosphorylation reaction was made up of ~50 ml of phosphorylation buffer (50 mM Tris pH 7.5 at 25 °C, 1 mM EGTA, 2 mM DTT, 20 mM $MgCl_2$ and 10 mM ATP) containing ~20 μM 4E-BP2 and ~5 μM Erk2 in a dialysis bag placed in 1 l of phosphorylation buffer. Phosphorylation was allowed to proceed overnight with stirring before a 20 μl aliquot of the reaction was removed for mass spectrometric analysis. Once the expected number of sites was uniformly phosphorylated, the reaction was stopped by running over a nickel-nitrilotriacetic acid (Ni-NTA) column to remove the kinase. Flow-through and wash fractions were collected, concentrated and then purified via reverse-phase high-performance liquid chromatography (HPLC). HPLC fractions containing phosphorylated protein were pooled, concentrated and dialysed in 4 l of buffer for about 16 h. Site-directed mutagenesis and the above-described method of phosphorylation by dialysis allowed the generation of reproducible large quantities of samples phosphorylated at any chosen combination of phosphorylation sites for biophysical studies (Extended Data Table 1). Mass spectrometry and NMR analysis of all samples confirmed the phosphorylation state (Extended Data Fig. 8).

NMR spectroscopy and binding studies. NMR samples comprised approximately 0.1–1.0 mM $^{1}H^{15}N^{13}C$-labelled protein in a buffer containing 30 mM Na_2HPO_4, 100 mM NaCl, 2 mM DTT, 1 mM EDTA, 10% D_2O v/v, pH 6.0. All NMR experiments were performed on Varian INOVA 500, 600 and 800 MHz spectrometers equipped with pulsed-field gradient units and triple resonance probes with a 600 MHz spectrometer equipped with a cryogenically cooled probe. NMR data sets were processed with the NMRPipe software package[31] and analysed using SPARKY[32] and FuDA (http://pound.med.utoronto.ca/~flemming/fuda/).

ITC studies and most NMR experiments, including those for chemical shift assignment, were performed at pH 6.0 and 20 °C. Temperature- and pH-dependent NMR experiments were recorded with temperature and pH values that varied from 5 to 35 °C and 4.0 to 10.3, respectively. The $[^{1}H]$–^{15}N-NOE relaxation measurements were performed at pH 6.0, 5 °C and 800 MHz. Other experimental details for the ITC and NMR were as previously described for the non-phosphorylated 4E-BP2 (ref. 11). To obtain distance restraints for CS-ROSETTA structural calculations (see below), we also recorded combined ^{15}N- and ^{13}C-edited nuclear Overhauser effect spectroscopy (CN-NOESY) and ^{15}N-edited NOESY[33] data sets, with amide proton temperature coefficients measured to identify amides involved in intramolecular hydrogen bonding[29].

Calculation of structures of the folded state of phosphorylated 4E-BP2 (pWT). The structure of the major state of pWT was calculated with the CS-Rosetta program[34] using chemical shifts with ^{1}H–^{1}H NOEs as distance restraints. First, TALOS+[35] was used to determine the secondary structure propensity of pWT on the basis of the measured backbone ^{1}H, ^{15}N and $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts; a prediction of four strands and a small helix was obtained for residues P18–R62 (Extended Data Fig. 5). Interestingly, this region showed both well-dispersed ^{1}HN chemical shifts and intramolecular amide proton hydrogen-bond formation as well as high positive ^{1}H–^{15}N heteronuclear NOEs for most residues (see text, Fig. 1 and Extended Data Fig. 1d). To test whether this secondary structure was compatible with an independently folded domain, CS-Rosetta2 was used to generate approximately 20,000 models using the standard protocol and the best five structures by the Rosetta energy function converged to a single topology within 2.1 Å in Cα root mean square deviation (r.m.s.d.) for residues P18–K57, confirming that the chemical shift data were compatible with the formation of a folded domain (Extended Data Fig. 3). Note that ^{1}H–^{1}H NOEs involving residues C-terminal to K57 could not be confidently assigned, so structure calculations were performed only for residues P18–K57 and, thus, no structural data are available for the five motionally restricted residues F58–R62.

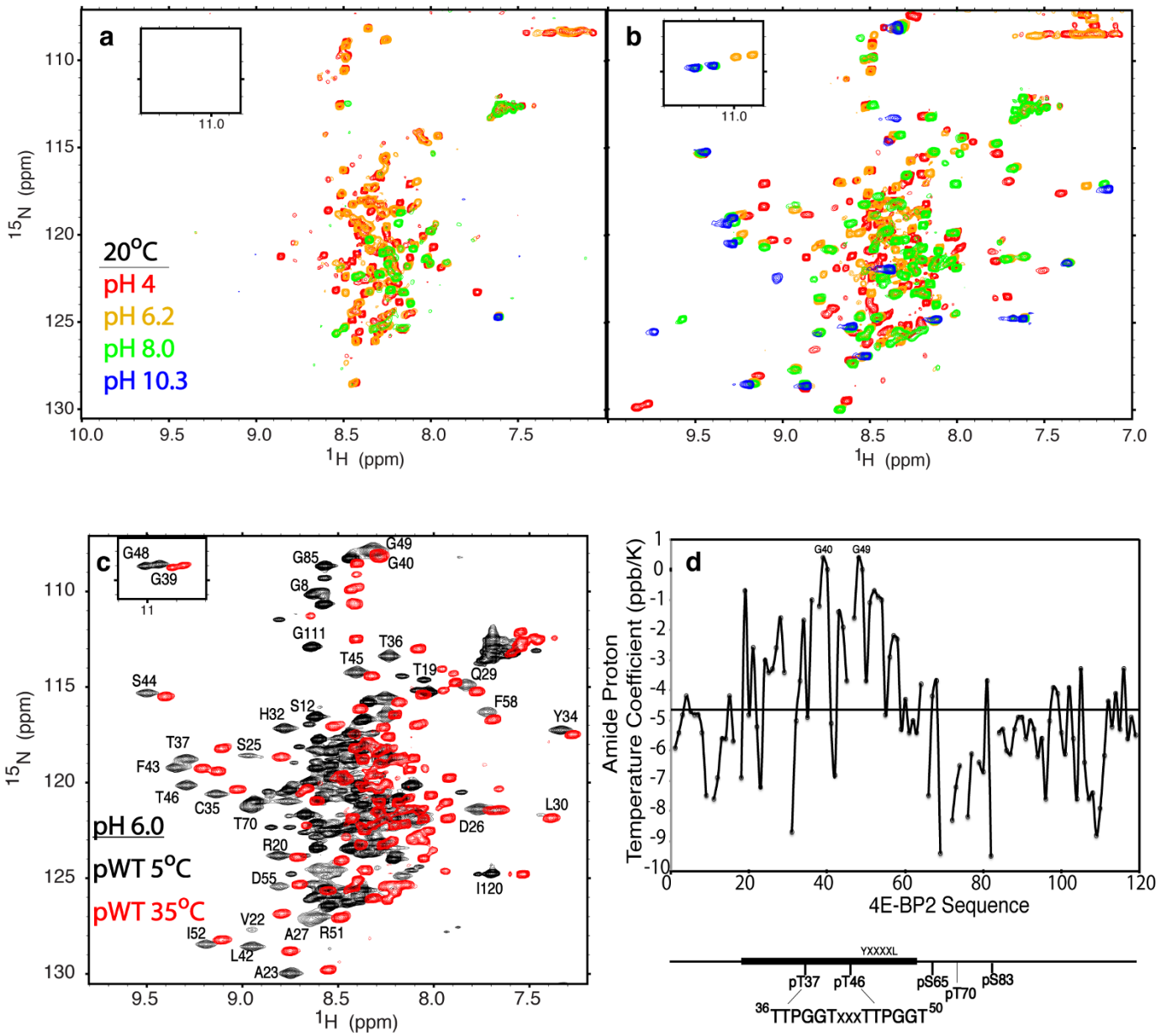
To provide further structural details for the folded state, ^{1}H–^{1}H NOE intensities were then used to identify distance restraints between atoms. Complications arose during the assignment of the NOESY data, as spectra showed significant overlap due to large regions of disorder in much of the protein as well as line broadening and evidence for minor states for the residues composing the fold (see text). Consequently preliminary attempts to assign the NOE data automatically failed, and manual assignment was complicated by ambiguity. We thus selected our most confident assignments for a final set of 494 discrete atom–atom pairs and six ambiguous pairs, which we then split into strong, medium and weak distance restraints by taking the highest 25%, middle 50% and lowest 25% by peak intensity, with separate boundaries defined for NOEs involving HN atoms and those that did not. Importantly, the CS-Rosetta2 ensemble using both chemical shifts and NOEs demonstrated the same overall fold as that calculated by CS-Rosetta2 using only chemical shifts. Comparing our distance restraints with the CS-Rosetta2 ensemble calculated with NOEs showed that the majority of the restraints were satisfied, with an average of 79.7% of the full set and 61.0% of the long-range contacts (atom pairs separated by ten or more residues) satisfied across the ensemble (Extended Data Table 1). Notably, the only clusters of long-range distance restraints observed in the NOEs corresponded to the strand pairs in the ensemble.

On the basis of our expectation that a significant number of the violated restraints come from transiently populated minor states, and to focus our calculations on the structure of the major state, we produced models using a protocol that allowed individual restraints to be violated while selecting for models that had the highest overall fit to the data. For this, we first converted the NOESY data into ambiguous distance restraints for use in CS-Rosetta2, then identified the largest violations by generating 2,000 models and removing 13 constraints from atom pairs that were not observed within 6/7/8 Å (for strong/medium/weak NOEs) of each other in the structure with the best constraint score. We then ran a second round of CS-Rosetta2 for a fixed amount of time, generating 20,359 models using the remaining restraints filtered down to 6,490 models by taking structures where the amide protons of G39 and G48 were both hydrogen bonded (as indicated by the temperature coefficients and large downfield ^{1}HN chemical shifts), then scored each model on the basis of their percentage of long-range NOEs satisfied (closest atom pair distances within 4/5/6 Å for strong/medium/weak). We then used those percentages to filter the pool further to 325 structures representing the best 5% by NOEs satisfied, and finally out of those selected the best 20 by their Rosetta energies (Extended Data Table 1 and Extended Data Figs 3 and 4).

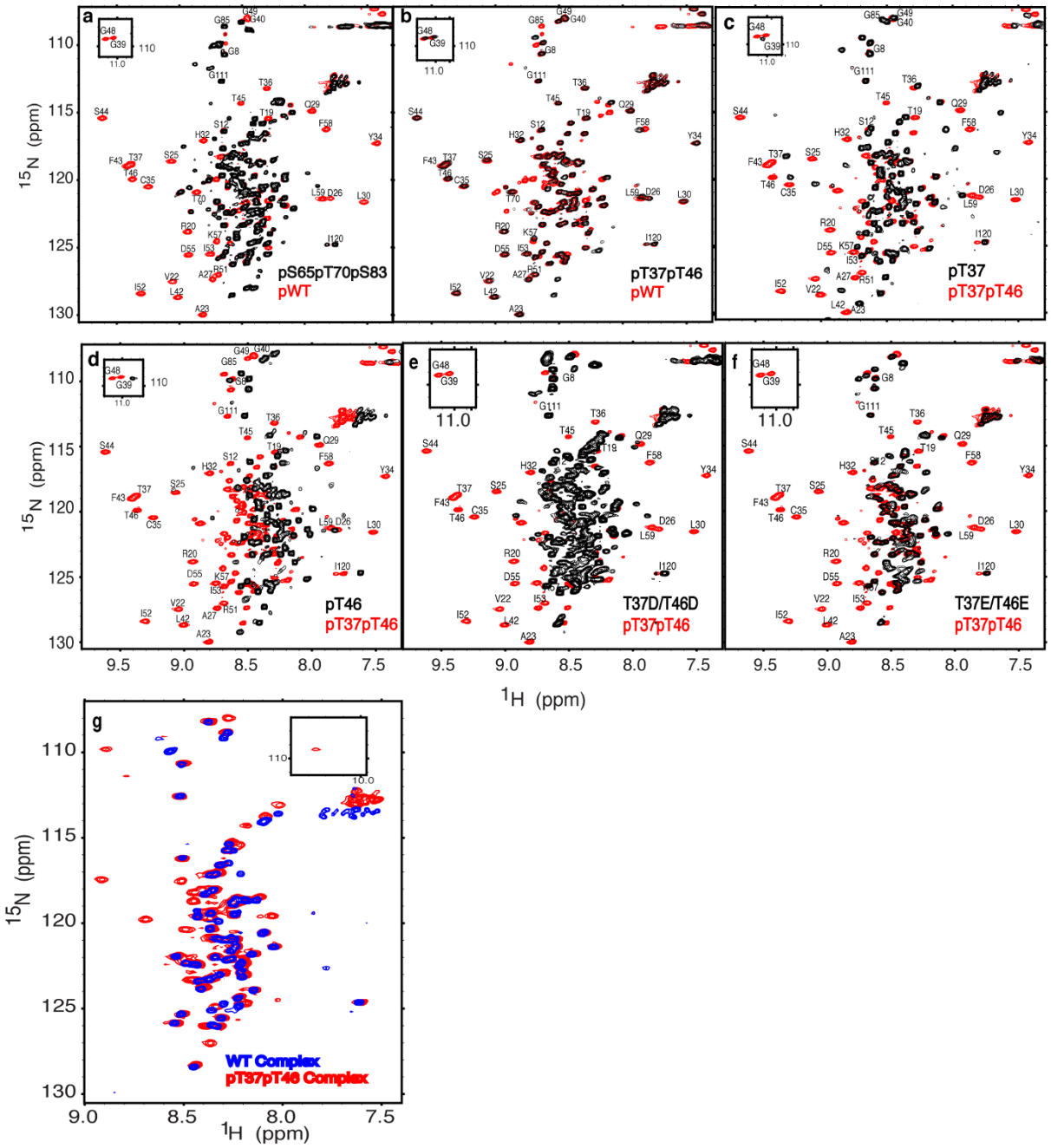
A contact map plotting NOE and NOE violations by their residue pairs (Extended Data Fig. 6a) shows how the folded topology emerges from the majority of the data, with violations clustering in specific locations. Consistent with our interpretation, violations appear to arise from conformational exchange with minor states that retain significant folded structure, representing folding intermediates or partial structures that have lost one or two contacts, as demonstrated by analysis of specific NOEs (Extended Data Fig. 6b–d). Note that these could give rise to chemical shifts that overlap the major folded state for most residues with additional resonances for the disordered portions, consistent with the observation of minor peaks in the disordered region of the spectrum.

Validating the structural model of the phosphorylation-induced folded domain. To test the four β-stranded structural model, we performed mutagenesis of strategic residues in the fold using both *in silico* ΔΔG predictions[36] and experimental site-directed mutagenesis (see text and Extended Data Table 1c). According to the ΔΔG predictions made on the basis of the lowest scoring CS-Rosetta structure before the incorporation of NOE-based restraints, perturbing the TpTPGGT motifs, which form the tight turns of the hairpin, or the canonical binding motif (YXXXXLΦ) significantly destabilizes the fold. Some of these mutations were tested experimentally and shown to destabilize the fold (Fig. 3, Extended Data Figs 2 and 7 and Extended Data Table 2).

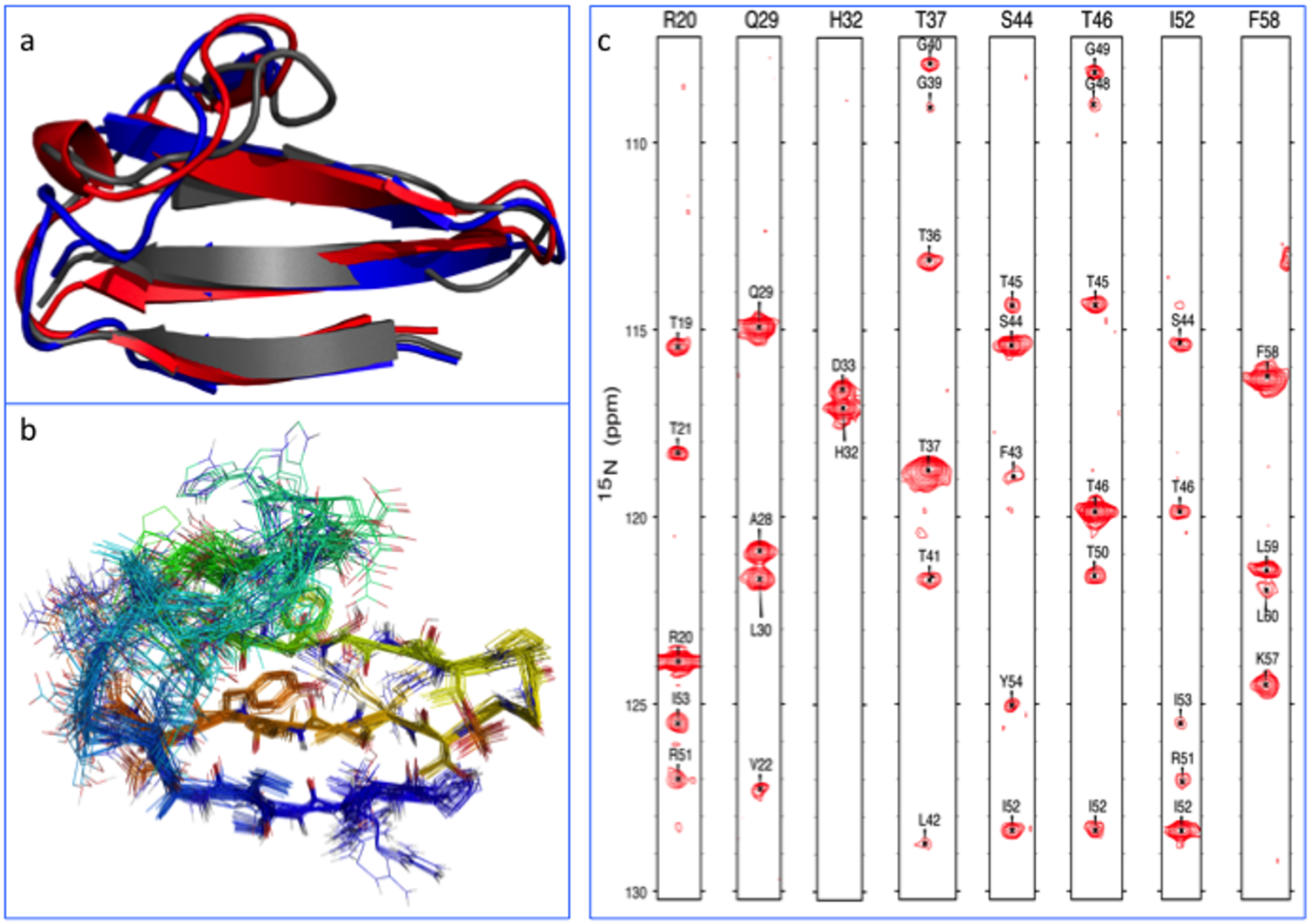
31. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
32. Goddard, T. D. & Kneller, D. G. SPARKY 3 (University of California, San Francisco, 2006).
33. Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Formankay, J. D. & Kay, L. E. Simultaneous acquisition of 15N- and 13C-edited NOE spectra of proteins dissolved in H_2O. *J. Magn. Reson. B* **103,** 197–201 (1994).
34. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105,** 4685–4690 (2008).
35. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
36. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79,** 830–838 (2011).



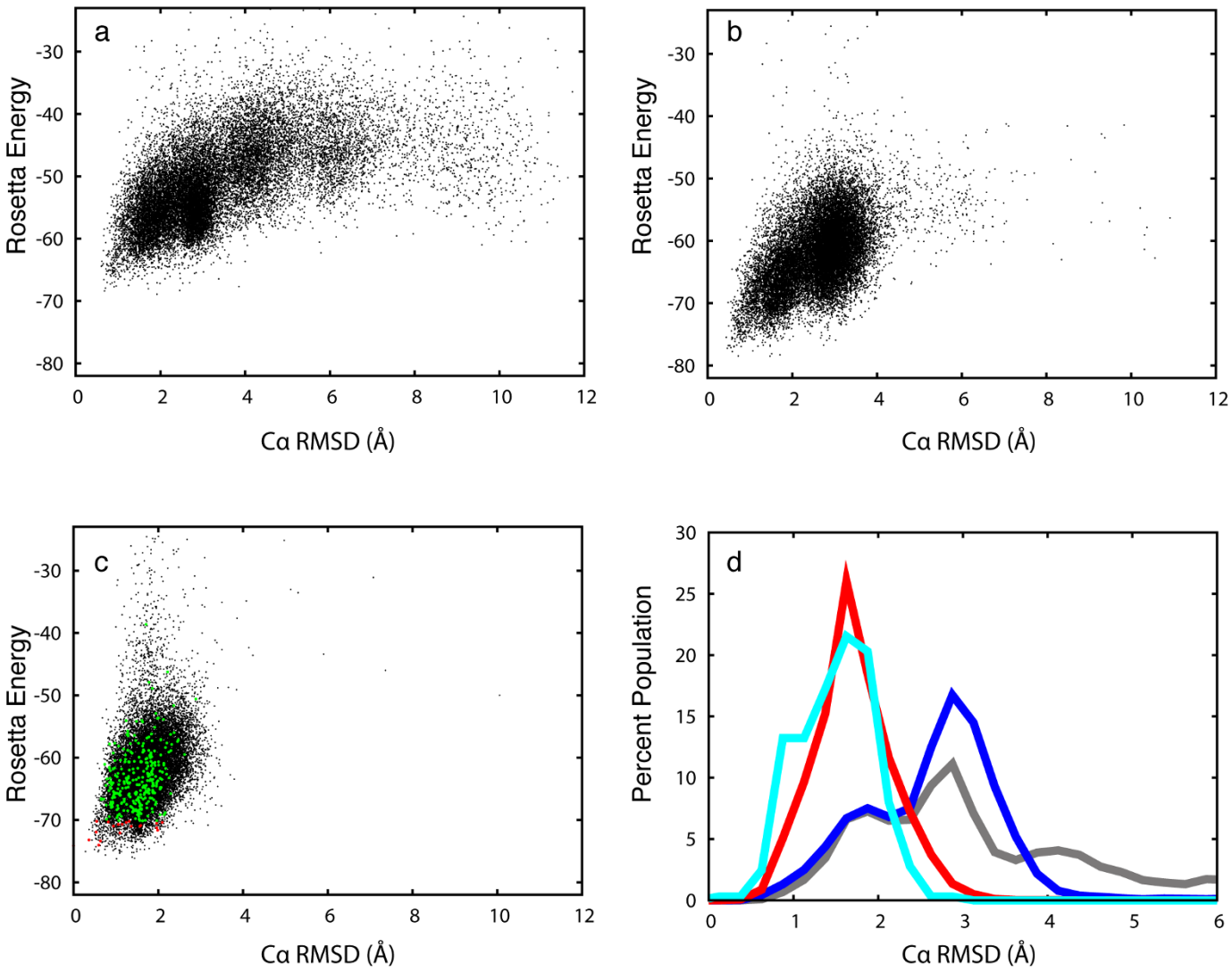
Extended Data Figure 1 | Effects of solution conditions on the structural and dynamic properties of phosphorylated 4E-BP2. **a**, **b**, Overlay of $^{1}H–^{15}N$ HSQC spectra of 4E-BP2 in 100 mM NaCl, 2 mM DTT, 1 mM EDTA, 1 mM benzamidine, 30 mM acetate pH 4 (red), 30 mM phosphate pH 6.2 (orange), 30 mM Tris pH 8.0 (green) or 30 mM CAPS pH 10.3 (blue) for (**a**) non-phosphorylated and (**b**) phosphorylated 4E-BP2. Because of the rapid ^{1}HN-solvent exchange of IDPs, ^{1}H-detected NMR experiments are usually performed at acidic pH. For non-phosphorylated 4E-BP2, all the peaks except one disappear by pH 10.3, while for phosphorylated 4E-BP2, most of the resonances for residues involved in hydrogen bonds upon folding remain visible. **c**, Effects of temperature on the structural and dynamic effects of phosphorylated 4E-BP2. Conformational exchange including *cis–trans* isomerization results in major and minor states of phosphorylated 4E-BP2. In addition to the assigned major states of both the folded and disordered regions of phosphorylated 4E-BP2, there remain many unassigned low-intensity peaks with ^{1}HN chemical shifts between 7.8 and 8.8 p.p.m., indicating that minor states contain significant disorder. The number of minor peaks decreases with increasing temperature, as shown in the $^{1}H–^{15}N$ HSQC spectra at 5 °C (black) and 35 °C (red), respectively. **d**, ^{1}HN temperature coefficients indicate intramolecular hydrogen bonding for many residues between P18 and R62. Chemical shift changes of $^{1}H–^{15}N$ HSQC spectra of pWT from 5 to 15 °C at pH 6.8 were used to calculate the ^{1}HN temperature coefficients. A horizontal line is plotted at -4.6 p.p.b. K^{-1}, with values above this line indicative of intramolecular hydrogen bonding[29] and missing data points representing proline residues.



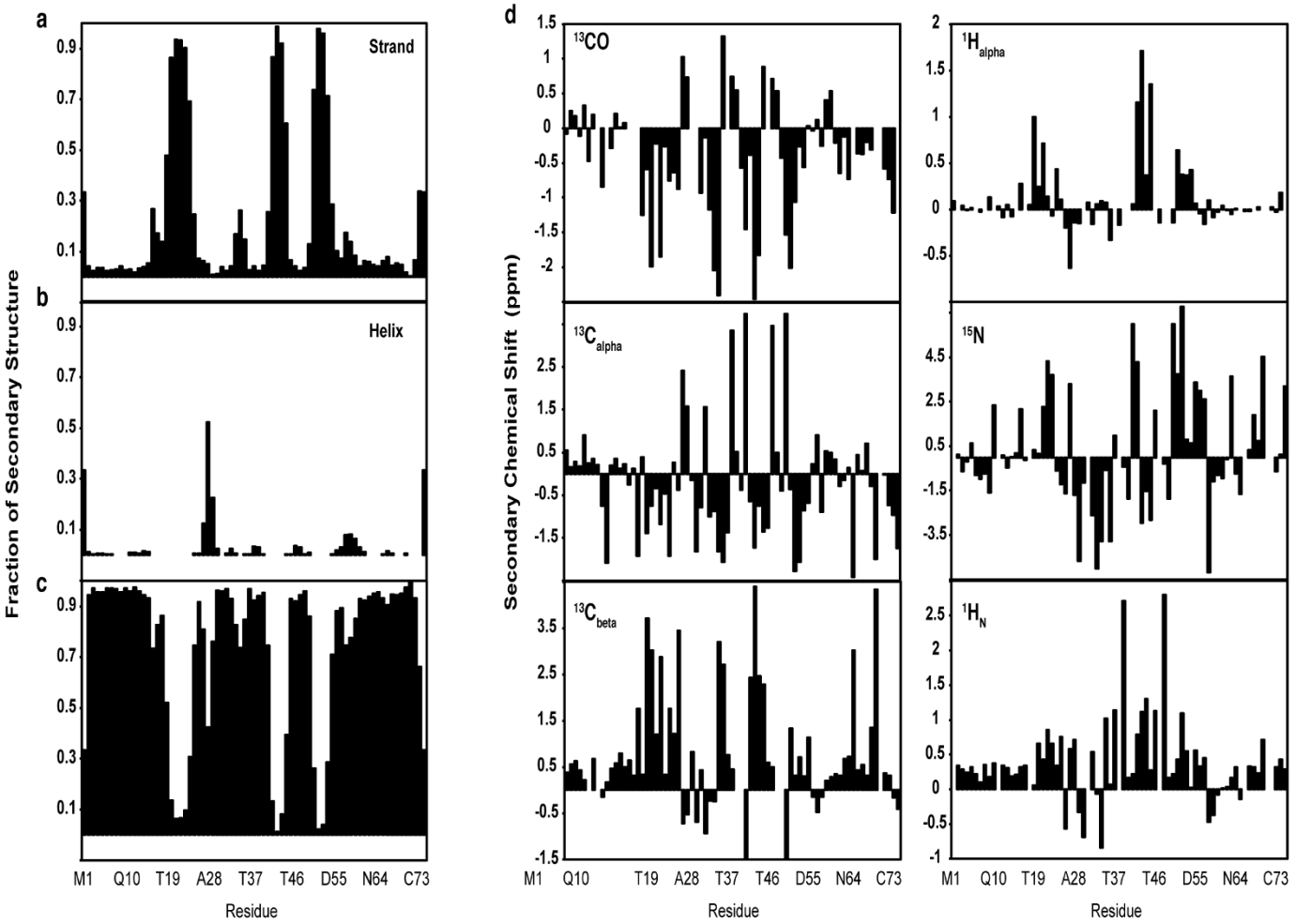
Extended Data Figure 2 | Phosphorylation, but not phosphomimetics, of both T37 and T46 are required to induce folding of 4E-BP2. **a**, **b**, Overlay of ^{1}H–^{15}N HSQC spectra of pWT with (**a**) pS65pT70pS83 and (**b**) pT37pT46. pS65pT70pS83 remains disordered as indicated by the lack of large ^{1}HN chemical shift dispersion of the folded state and absence of the downfield shifted G39 or G48 resonances. In contrast, pT37pT46 is nearly identical to pWT for residues P18–R62. **c**–**f**, Overlay of pT37pT46 with (**c**) pT37, (**d**) pT46, (**e**) T37D/T46D and (**f**) T37E/T46E. pT37 and pT46 are partly folded as indicated by the presence of only one of the downfield shifted G39 or G48 resonances in the insets, while aspartic acid or glutamic acid substitutions at T37 and T46 did not induce any folding, demonstrating that these are not good phosphomimetics. **g**, Binding of non-phosphorylated WT and pT37pT46 to unlabelled eIF4E. Overlay of two-dimensional ^{1}H–^{15}N HSQC spectra of non-phosphorylated WT (blue) and pT37pT46 (red) 4E-BP2 in complex with unlabelled eIF4E. Although both proteins are mostly disordered when bound to eIF4E, as evident from the poor ^{1}HN chemical shift dispersion, the pT37pT46 complex retains some structure, resulting in more chemical shift dispersion than WT.



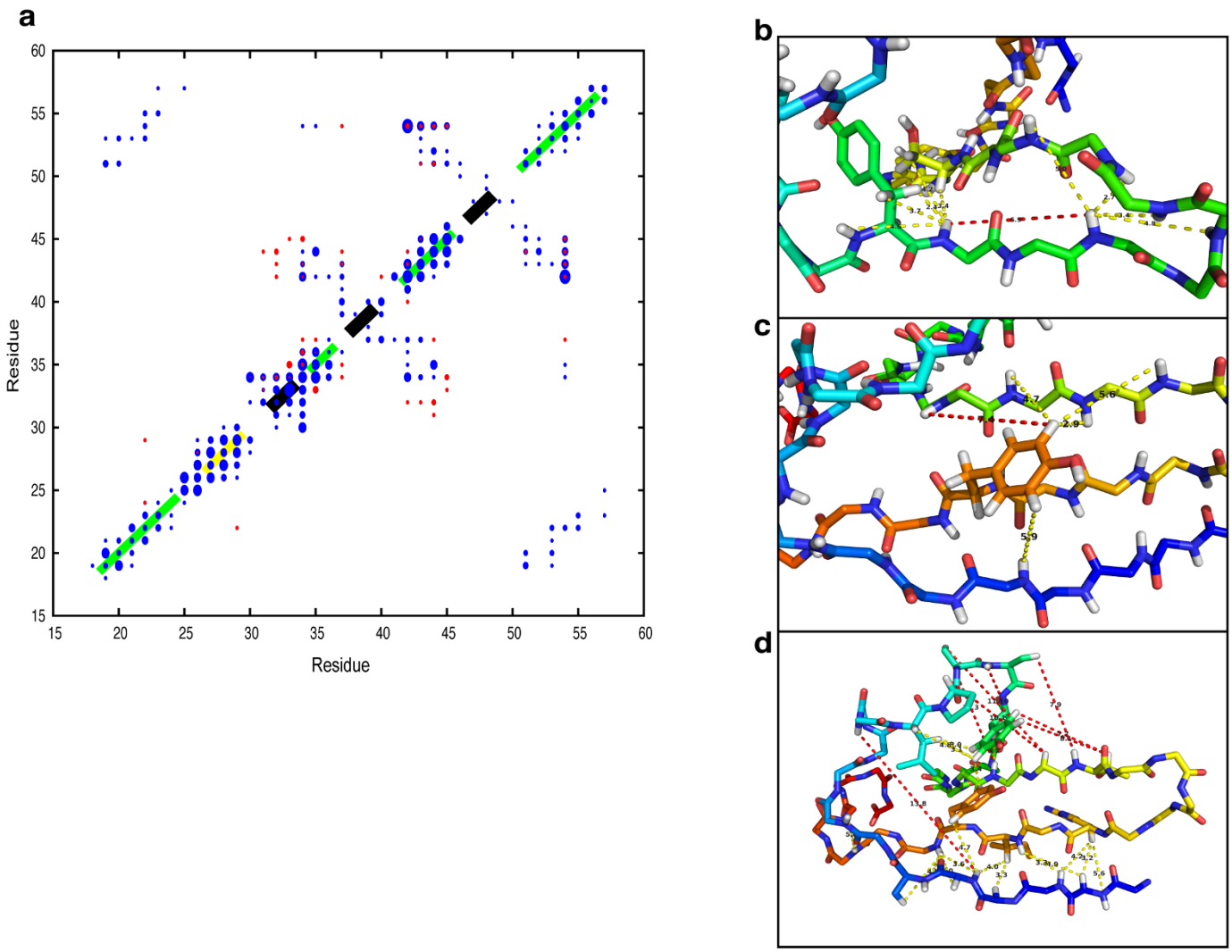
Extended Data Figure 3 | Structural models of phosphorylated 4E-BP2 calculated with CS-Rosetta and NMR data. **a**, Alignment of ribbon diagrams of the lowest energy structures from ~ 20,000 structural models using different inputs: ^{1}HN, ^{15}N, Cα and Cβ chemical shifts (grey), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts (blue), and ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts as well as NOEs (red). **b**, Superposition of the final 20 lowest energy structures calculated using all chemical shifts and NOEs. Residues P18–K57 are shown using a rainbow colour spectrum from amino (N) to C termini. **c**, Examples of ^{1}HN–^{1}HN NOEs within the folded 4E-BP2. Shown are strips from ^{15}N-edited ^{15}N-NOESY demonstrating both short- and long-range interactions. Note that residues within the long loop connecting strands β1 and β2 contain no short- or long-range NOEs, indicating that it is very dynamic, consistent with low [^{1}H]–^{15}N NOEs. Not surprisingly, this loop shows the largest variation in the models (see **a**, **b**).



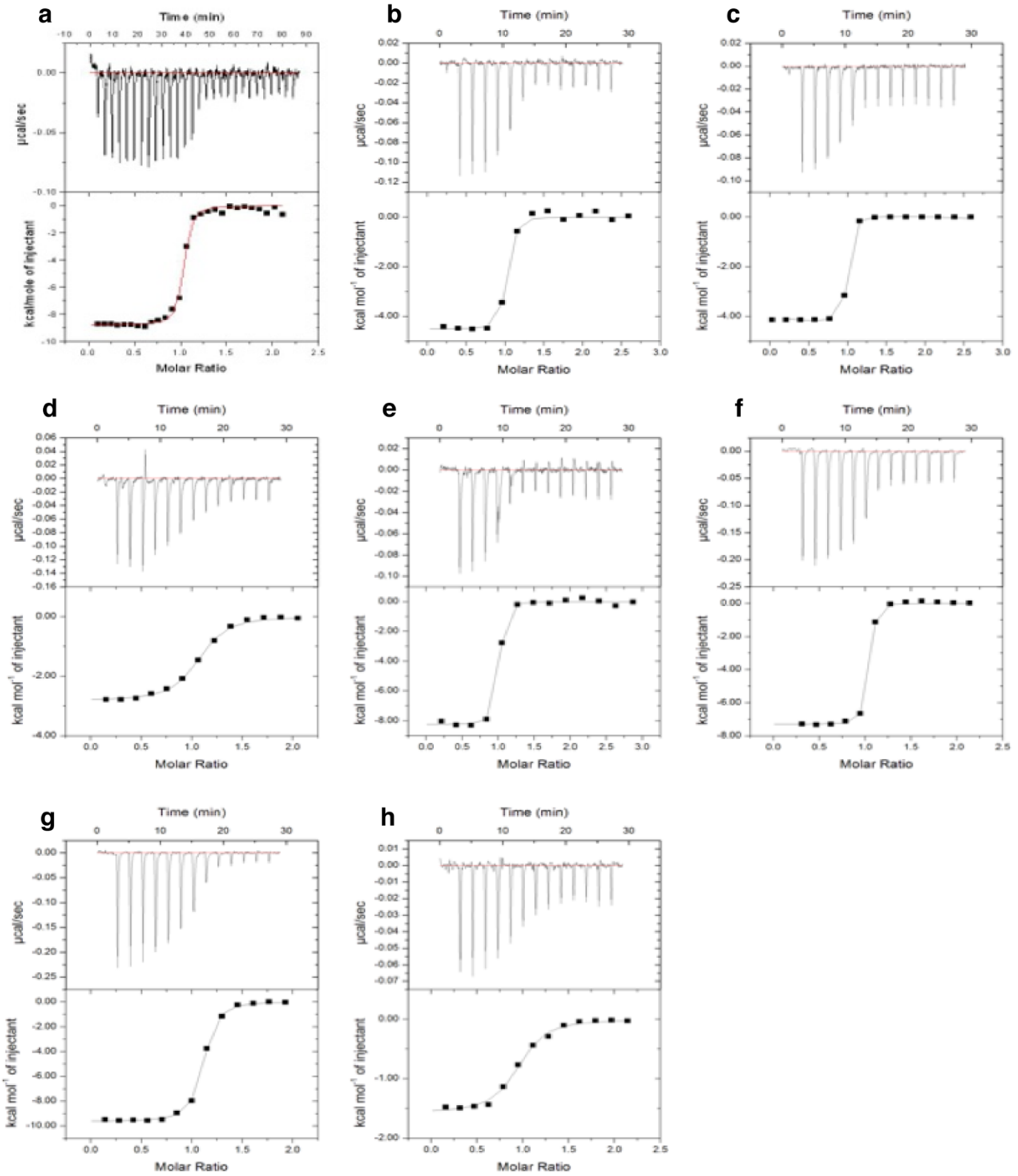
Extended Data Figure 4 | CS-Rosetta scores for calculations of the folded state of phosphorylated 4E-BP2 using different input data. CS-Rosetta2 was used to create models three separate times as new input data were acquired, each time with the observation that low Rosetta energy models converge to the same topology. Over these three sequential runs each addition of new data served primarily to drive sampling towards a previously observed energy minimum. **a**–**c**, CS-Rosetta energy for ~20,000 structural models as a function of Cα r.m.s.d. to the structure with the lowest energy point in the final ensemble using (**a**) ^{1}HN, ^{15}N, ^{13}Cα and ^{13}Cβ chemical shifts, (**b**) ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts, and (**c**) all chemical shifts and NOEs (final ensemble). In green are the best 5% of structures on the basis of agreement with NOEs, whereas in red are the 20 structures with the lowest NOE violations used to generate Extended Data Fig. 3b. Note that the CS-Rosetta energy plotted here is the empirical Rosetta energy value without chemical shift or NOE terms, reflecting the intrinsic energy rather than the fit to experimental data. **d**, Histograms showing the percentage distribution of structures with Cα r.m.s.d. as a function of Cα r.m.s.d. (going out to 6 Å) for the different CS-Rosetta input data: ^{1}HN, ^{15}N, ^{13}Cα and ^{13}Cβ chemical shifts (grey), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts (blue), ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ, and ^{13}CO chemical shifts and NOEs without filtering (red) and with the final set of filters (cyan) for (1) the hydrogen bonding observed from the temperature coefficient measurements and (2) the best 5% by number of NOE satisfied.



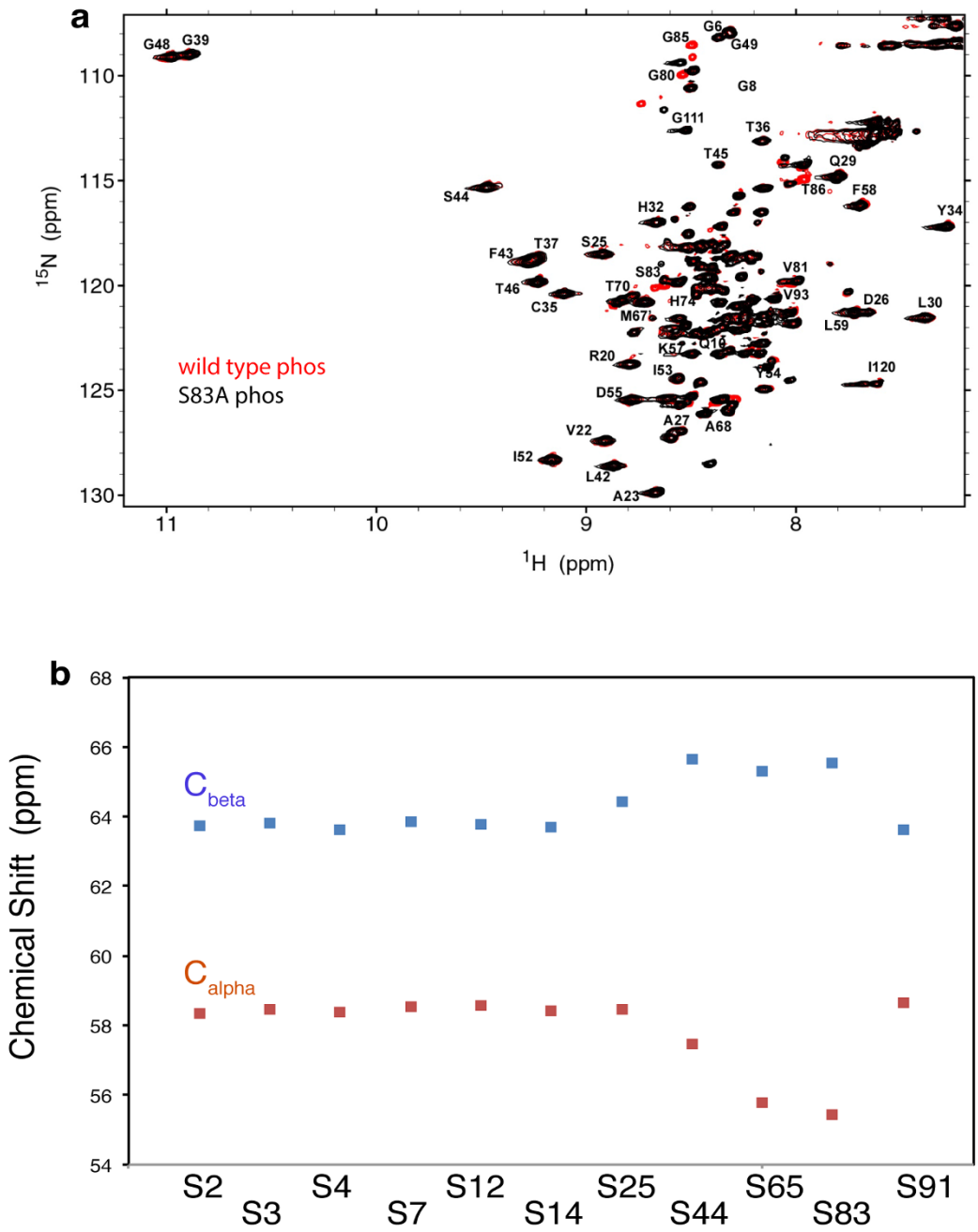
Extended Data Figure 5 | Chemical shifts and calculated secondary structure of phosphorylated 4E-BP2. Secondary chemical shifts define the topology of phosphorylated 4E-BP2 and validate the CS-Rosetta approach for structure determination. **a**–**c**, Fractional secondary structure as calculated by Talos+ as a function of residues number for (**a**) strand, (**b**) helix and (**c**) loop for residues 1–75 of phosphorylated 4E-BP2. **d**, Secondary chemical shifts for the folded region of phosphorylated 4E-BP2 from Talos+ as a function of residue number for ^{13}CO, ^{1}Hα, ^{13}Cα, ^{15}N, ^{13}Cβ and ^{1}HN shifts.



Extended Data Figure 6 | NOE violations for calculated structure of phosphorylated 4E-BP2. **a**, Contact map showing observed NOEs for each pair of residues for satisfied NOE restraints in blue and unsatisfied NOE restraints in red, with the areas of the circles proportional to the total number of NOEs in each case. Violations were calculated using distance boundaries of 5/6/7Å for strong, medium and weak NOEs, and the number of violations for each residue pair was either averaged across the ensemble (above the $x = y$ line) or by only counting restraints that were never satisfied in any of the models in the ensemble (below the $x = y$ line). The secondary structure of the protein is represented on the diagonal in green (β-strand), yellow (3_{10} helix) and black (turn) bars. **b**–**d**, Examples of NOE violations consistent with dynamic conformational exchange. A detailed look at individual NOE pairs (satisfied shown in yellow dashed lines, unsatisfied in red dashed lines) supports the conclusion that minor conformations contribute to the high number of violations, as consistently violated restraints conflict with the majority of the data that define the major conformation. For more information about the NOE violations and conformational exchange within phosphorylated 4E-BP2, see Supplementary Information.



Extended Data Figure 7 | ITC binding profiles of several 4E-BP2 constructs to eIF4E at 20 °C. **a**, WT; **b**, pT37; **c**, pT46; **d**, pT37pT46; **e**, T37D/T46D; **f**, T37E/T46E; **g**, pG39VG48V); **h**, pWT using competition with pS65pT70pS83.



Extended Data Figure 8 | Resonance assignments of phosphorylated residues. **a**, Overlay of ^{1}H–^{15}N HSQC spectra of pWT in red with phosphorylated S83A (pT37pT46pS65pT70) in black, showing the absence of the pS83 peak in pWT and other local changes. The blue arrow indicates the position of A83. **b**, Serine ^{13}Cα (red) and ^{13}Cβ (blue) chemical shifts in phosphorylated 4E-BP2. Although they did not show significant downfield shifts in the ^{1}HN–^{15}N HSQC spectrum (Fig. 1), S65 and S83 showed significant deviations from the random coil values compared with the other serines, consistent with phosphorylation. S25 and S44 also showed deviations as a result of the interactions within the folded domain (close in space to other phosphates), but not to the degree expected for a phosphorylated serine.

Extended Data Table 1 | Structural and energetic properties of phosphorylated 4E-BP2

(a) Percentage of satisfied NOE restraints for various distance boundaries in CS-Rosetta2 calculations.

Distance Boundaries (Å) (Strong/Medium/Weak)	Initial CSRosetta2 Ensemble (% restraints below boundaries)		Constrained CSRosetta2 Ensemble (% restraints below boundaries)	
	All NOEs	Long Range*	All NOEs	Long Range*
3.0/4.0/5.0	58.9	41.6	62.2	52.3
3.5/4.5/5.5	72.3	51.6	78.3	68.1
4.0/5.0/6.0	79.8	61.0	86.3	77.7
4.5/5.5/6.5	86.1	68.7	90.6	81.2
5.0/6.0/7.0	88.8	72.4	92.8	82.7
5.5/6.5/7.5	91.9	78.2	94.7	86.9
6.0/7.0/8.0	93.4	81.7	96.0	88.8

(b) Summary of NOE restraints and structural statistics

Number of unambiguous NOEs	450
Intraresidue (|i − j| = 0)	116
Short range (|i − j| = 1)	172
Medium-range (2 ⩽ |i − j| ⩽ 9)	95
Long-range (|i − j| > 9)	67
Number of ambiguous NOEs	6
Ramachandran plot statistics (%)	
Residues in most favored regions	93.5
Residues in additionally allowed regions	6.5
Residues in generously allowed regions	0
Residues in disallowed regions	0
Average RMSD to mean (Å)	
Backbone (residues P18-K57)	0.92 ± 0.24
Heavy atoms (residues P18-K57)	1.47 ± 0.22

(c) ΔΔG predictions of energetic effects of mutation of the TTPGGT motifs and canonical binding site using the initial CS-Rosetta structural model.

Amino Acid	Residue #	Mutate to / Predicted ΔΔG							
T	36 / 45	A	+1.1 / +1.3	D	+1.8 / +5.5	E	+0.66 / +4.8	S	+1.3 / +1.6
T	37 / 46	A	+1.5 /+0.62	D	+0.61/+0.78	E	+1.2 / +0.94	S	+1.4 / +0.77
P	38 / 47	A	+1.8 / +2.1	G	+3.1 / +3.6				
G	39 / 48	A	+6.8 / +3.4	V	+8.3 / +5.6	P	+15 / +120		
G	40 / 49	A	+6.4 /+0.29	V	+9.6 / -0.10	P	+22 / +13		
T	41 / 50	A	-0.42/+2.11	D	+0.64 / +3.5	E	+0.05 / +1.4	S	+0.39 / +2.7
Y	54	A	+5.0	F	+0.75	E	+1.5		

Extended Data Table 2 | ITC binding parameters for 4E-BP2 constructs to eIF4E

	ΔH (kcal/mol)	TΔS (kcal/mol)	K_d (nM)	ΔG (Binding & Unfolding)[a] (kcal/mol)	Folded?	Estimated Unfolding Free Energy[a] (kcal/mol)
WT	-8.81 ± 0.08	2.61	3.20 ± 0.6	-11.42	No	-
pT37	-4.51 ± 0.08	6.42	7.25 ± 1.7	-10.93	Partially	0.49
pT46	-4.14 ± 0.02	7.42	2.43 ± 0.5	-11.56	Partially	~ 0
pT37pT46	-2.83 ± 0.03	5.98	267 ± 32	-8.81	Yes	2.61[b]
T37D/T46D	-8.25 ± 0.09	3.08	3.89 ± 1.1	-11.33	No	-
T37E/T46E	-7.29 ± 0.04	3.93	4.37 ± 0.8	-11.22	No	-
pS65pT70pS83	-4.27 ± 0.04	6.39	11.3 ± 2.9	-10.66	No	-
p(G39V/G48V)	-9.62 ± 0.05	0.361	36.1 ± 3.5	- 9.98	No	-
pWT	-	-	12,320 ± 200	-6.59	Yes	4.83[b]

Enthalpy changes (ΔH), entropy changes (ΔS) and dissociation constants (K_d) from ITC measurements as well as combined binding and unfolding free energies derived from K_d values using $\Delta G = -RT\ln(1/K_d)$, with R the ideal gas constant and T the temperature, and estimated unfolding free energies using WT as a reference for pT37, pT46, pT37pT46 and pWT.

[a]ΔG values, which may include contributions from binding, unfolding of the β-sheet structure encompassing residues P18–R62 and folding of the helix encompassing the YXXXXLΦ sequence, were calculated from $\Delta G = \Delta H - T\Delta S = -RT\ln(1/K_d)$. The estimated unfolding free energy was obtained using $\Delta G - \Delta G_{WT}$, as for WT there was no contribution from unfolding.

[b]The enhanced stability of the fivefold phosphorylated (pWT) over pT37pT46, 4.8 versus 2.6 kcal mol^{-1} estimated unfolding free energy, is possibly due to long-range transient electrostatic interactions between the phosphorylated acidic C terminus with the basic folded domain. The predicted isoelectric point (pI) of non-phosphorylated full-length 4E-BP2 is 6.16; however, the region involved in the phosphorylation-induced folding is very basic (pI = 9.77), while the C terminus (S65–I120) is acidic (pI = 4.68). Phosphorylation is expected to drop the pI even further, potentially causing stabilizing transient electrostatic interactions between the folded region and the phosphorylated C terminus.