Kinetic stabilization of translation-repression condensates by a neuron-specific microexon

Carla Garcia-Cabau[1,#], Anna Bartomeu[1,#], Andreea Balaceanu[1], Giulio Tesei[2], Berta Duran-Arqué[1], Marcos Fernández-Alfara[1], Judit Martín[1], Cesare De Pace[3], Lorena Ruiz-Pérez[3], Jesús García[1], Giuseppe Battaglia[3,4], Kresten Lindorff-Larsen[2], Raúl Méndez[1,4,*], Xavier Salvatella[1,4,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.
[2]Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200, Copenhagen, Denmark.
[3]IBEC, The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.
[4]ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Spain.

[#]equally contributed

*to whom correspondence should be addressed

raul.mendez@irbbarcelona.org
xavier.salvatella@irbbarcelona.org

The inclusion of microexons by alternative splicing is frequent in neuronal proteins. The roles of these sequences are in most cases unknown, but changes in their degree of inclusion are associated with neurodevelopmental diseases. We recently found that the decreased inclusion of a 24-nucleotide neuron-specific microexon in CPEB4, an RNA-binding protein that regulates translation through cytoplasmic changes in poly(A) tail length, is linked to idiopathic autism spectrum disorder (ASD). Why this microexon is required and how small changes in its degree of inclusion generate a dominant-negative effect on the expression of ASD-linked genes is not clear. Here we show that neuronal CPEB4 forms condensates that dissolve upon depolarization, a phase transition associated with a switch from translational repression to activation. Heterotypic intermolecular interactions between the microexon and a cluster of histidine residues kinetically stabilize the condensates by competing with homotypic interactions between clusters, that otherwise lead to the irreversible aggregation of CPEB4. We conclude that microexon 4 in neuronal CPEB4 is required to preserve the reversible regulation of CPEB4-mediated gene expression in response to neuronal stimulation.

Introduction

Evolutionarily conserved neuron-specific microexons insert one to several amino acids in proteins associated with neurogenesis, axon guidance, and synaptic function. The alternative splicing processes that coordinate their inclusion are frequently misregulated in individuals with autism spectrum disorder (ASD)[1–4]. Recently, we reported that individuals with idiopathic ASD[5] or schizophrenia[6] show decreased inclusion of a 24-nucleotide neuron-specific microexon (exon 4) in CPEB4. This RNA-binding protein is a member of the cytoplasmic polyadenylation element binding (CPEB) family of proteins, which recognizes cytoplasmic polyadenylation elements (CPEs) in 3′-untranslated regions of transcripts and mediates their poly- and deadenylation, thus regulating their translation through the assembly of translation-repression or translation-activation ribonucleoprotein complexes in turn regulated by CPEB-specific post-translational modifications[7]. CPEBs participate in mRNA transport and localization to different subcellular compartments[8,9] and play important roles in cell division, neural development, learning, and memory.

Neuronal CPEB4 (nCPEB4) binds to the transcripts encoded by ASD risk genes. A mouse model with low inclusion of microexon 4 recapitulates the defects in mRNA polyadenylation and protein expression observed in ASD patients and also induces ASD-like neuroanatomical, electrophysiological, and behavioral phenotypes[5]. The CPEB4 variant lacking microexon 4 (nCPEB4Δ4) is dominant-negative in neurons[5] but not in non-neuronal tissues, where it is functional. CPEBs can be classified into two subfamilies—one containing CPEB1 and the other containing CPEBs 2 to 4—that differ in the motifs they recognize in mRNAs, their interactions with other proteins, and their regulation. The C-terminal RNA-binding domains of CPEBs are globular whereas the N-terminal domains (NTDs) are intrinsically disordered[10,7,11]. The phosphorylation of multiple Ser and Thr residues in the NTDs of CPEBs 2 to 4 by ERK2 and Cdk1 activates these proteins by dissolving the translation-repression condensates that they form[7,11] and, interestingly, the neuron-specific microexon of CPEB4 mis-spliced in ASD is also found in CPEB2 and 3[12].

The mechanism by which the decreased inclusion of microexon 4 leads to the misregulation of gene expression in ASD is not known[13]. Here we studied how the absence of these eight amino acids affects the properties of the translation-repression condensates formed by nCPEB4. We found that their dissolution upon neuron depolarization occurs via the deprotonation of His residues in the NTD on nCPEB4, caused by intracellular pH increases upon neuron depolarization, that alter the balance of intermolecular interactions stabilizing the condensates. In addition, we observed that decreasing the degree of inclusion of microexon 4 increases the propensity of the translation-repression condensates to convert irreversibly

into aggregates that cannot be dissolved upon depolarization. The microexon enhances the kinetic stability of the condensates by a mechanism, akin to heterotypic buffering[14], where heterotypic interactions between the microexon and a cluster of aromatic residues compete with the homotypic interactions between aromatic clusters that drive aggregation. In summary, we show how microexon 4 ensures the proper regulation of nCPEB4 in neurons and explain why a reduction in its degree of inclusion in ASD results in a dominant-negative decrease in CPEB4 activity.

Results

Endogenous nCPEB4 condensates dissolve upon neuron depolarization

CPEB4 is expressed in most tissues but is more abundant in the nervous system[15] (Supplementary Fig. 1a). The inclusion of microexon 4 in nCPEB4 is limited to neuronal tissues, such as the brain and the retina (Supplementary Fig. 1b), where SRRM4, a neuron-specific alternative splicing regulator required for microexon inclusion programs[3], is expressed (Supplementary Fig. 1c). To study the condensation properties of the protein at endogenous levels, we generated an mEGFP-nCPEB4 knock-in (ki) mouse model (Fig. 1a), from which we obtained primary striatal neurons (Fig. 1b). Imaging of the endogenous-tagged CPEB4 in these *ex vivo* neurons revealed a granular distribution of CPEB4 in the cytoplasm (Fig. 1c and Supplementary Fig. 1d), reminiscent of the condensates formed by overexpressed CPEB4 in cycling non-neuronal cells[7,11]. Depolarization of these neurons by exposure to 50 mM KCl led to the rapid ($t < 3$ min) dissolution of these assemblies (Fig. 1d and Supplementary Fig. 1e), resembling the dissolution of translation repression CPEB4 condensates upon phosphorylation in mitotic cells[7,11]. We concluded that nCPEB4 localizes in cytoplasmic condensates that dissolve upon depolarization.

To further study the properties of microexon 4 splice variants, we used murine neural crest-derived neuroblastoma Neuro2a (N2a) cells, where neuronal differentiation can be induced by retinoic acid[16]. N2a cells, like neurons, express CPEB4 variants including and excluding microexon 4 (nCPEB4 and nCPEB4Δ4, Supplementary Fig. 1f), independently of their differentiation status. In contrast, cell lines of non-neural origin only express the variant excluding the microexon (Supplementary Fig. 1f). Inclusion of microexon 4 in N2a cells correlated with the expression of the splicing factor SRRM4 but not with that of Rbfox1 (Supplementary Fig. 1g), and depletion of SRRM4 in N2a cells decreased the inclusion of microexon 4 whereas overexpression of SRRM4 in the non-neuronal cell line 239T increased it[17] (Supplementary Fig. 1h). N2a cells, therefore, recapitulate the neuron-specific regulation

of CPEB4 alternative splicing. Overexpression of both full-length and NTD GFP-tagged nCPEB4 in these cells led to the formation of assemblies equivalent to those formed by endogenous CPEB4 in neurons (Fig. 1e). Both showed a similar number of cytoplasmic condensates, of similar volume (Fig. 1f), that appeared highly dynamic in fluorescence recovery after photobleaching (FRAP) experiments (Fig. 1g). Most importantly, the condensates formed by both full-length and NTD nCPEB4 also dissolved upon N2a depolarization (Fig. 1h and Supplementary Fig. 1i), as monitored by the induction of the early depolarization markers *junB* and *cFos* (Supplementary Fig. 1j).

The NTD of non-neuronal CPEB4 recapitulates the phase separation properties of the full-length protein in cells and undergoes phase separation *in vitro* to form droplets, which dissolve upon phosphorylation[11]. We observed that the NTD of nCPEB4 (nCPEB4-NTD) also formed droplets, that had the capacity to fuse (Supplementary Fig. 1k) and efficiently recover fluorescence in FRAP experiments (Supplementary Fig. 1l). To gain further insight into the molecular basis of nCPEB4 condensation, we studied how changes in solution conditions altered the phase separation properties of this neuronal form *in vitro*. nCPEB4-NTD formed droplets upon heating in the lower critical solution temperature (LCST) regime (Fig. 1i), indicating that the phase transition is entropy-driven, and at high ionic strength (Fig. 1j), suggesting that the droplets are in part stabilized by hydrophobic interactions[18].

pH fluctuations upon neuron depolarization regulate nCPEB4 condensation

To study how depolarization leads to the dissolution of the translation repression nCPEB4 condensates, we expressed nCPEB4 in N2a cells and obtained samples before and after this process. Analysis by mass spectrometry detected eleven phosphorylations in the NTD, including five previously identified in non-neuronal cells upon mitotic entry[7,11], and nine methylations of arginine and lysine residues located mainly in the RNA recognition motifs (RRMs) (Fig. 2a and Supplementary Fig. 2a). Although the phosphorylated/non-phosphorylated ratio was low and invariable upon depolarization (Fig. 2a), we nevertheless analyzed whether this modification contributes to the dissolution of the nCPEB4 condensates, as is the case for CPEB4 during the cell cycle[7,11]. We generated phosphomimetic (S/T to D) and non-phosphorylatable (S/T to A) variants of nCPEB4 and studied their phase separation properties in N2a cells. nCPEB4-11D failed to form condensates (Supplementary Fig. 2b) and nCPEB4-11A dissolved upon depolarization (Fig. 2b and Supplementary Fig. 2b). Depolarization involves calcium mobilization, which can be mimicked with ionophores such as ionomycin. However, the treatment of differentiated N2a cells with this compound did not affect the number or the size of the nCPEB4 condensates (Supplementary Fig. 2c). Taken together,

these results confirmed that the dissolution of the condensates formed by nCPEB4 in neurons and, thus, the switch of CPEB4 from translational repressor to activator, is regulated by a mechanism other than phosphorylation and calcium mobilization.

Neuronal depolarization causes a fast decrease in intracellular pH followed by a slower, longer-lasting increase[19,20]. This pH fluctuation mirrors the changes in the size of the nCPEB4 condensates observed after depolarization in N2a cells, where condensates appear to increase in size before dissolving (Fig. 2c). nCPEB4-NTD is enriched in His residues (Supplementary Fig. 2d) with side chains that have both aromatic and basic character and thus become positively charged when the pH approaches their pK_a, which has a nominal value of 6 but can be modulated by sequence context. Protonated His residues remain aromatic but their positive charge can influence how they interact with other residues. Thus, to study whether pH influences the phase equilibrium of nCPEB4-NTD, we compared the cloud point (Tc) of a nCPEB4-NTD solution at pH 7.5 with that of the same sample at lower (7.0) and higher (8.0) pH values. We observed that the Tc decreased by at least 10 ºC at pH 7.0 and increased by 13.5 ºC at pH 8.0 (Fig. 2d), indicating that the changes in the size of the nCPEB4 condensates observed upon neuron depolarization can indeed be due to pH fluctuations.

Next, to determine whether the aromatic character of His residues when they are neutral (at pH 8.0) also contributes to driving phase separation, we studied constructs in which 25% (H25S) or 50% (H50S) of His amino acids were substituted by Ser (Fig. 2e), a residue with similar hydrogen bonding properties that lacks aromatic character. We found that the Tc of H25S was 11 ºC higher than that of nCPEB4-NTD and that H50S did not phase separate (Fig. 2f). Substitution of the same number of residues in a cluster of His residues (HClust) centered around position 240 (Fig. 2e) led to smaller Tc increases (Fig. 2f). These results indicate that, as observed for other intrinsically disordered proteins rich in aromatic residues[21–23], the number and patterning of His residues determine the phase separation propensity of nCPEB4-NTD. In addition, they suggest that a decrease in pH leads to an increase in phase separation propensity by allowing protonated His side chains to engage in cation-π or electrostatic interactions as the cationic moiety.

To understand better the effect of pH changes on the balance of intermolecular interactions stabilizing the condensates, we carried out molecular simulations using the CALVADOS residue-level model[24,25]. First, we tested the ability of the model to recapitulate the effect of pH and amino acid substitutions on the behavior of nCPEB4-NTD. We found a high correlation between experimental saturation concentrations and model predictions for nCPEB4-NTD and His to Ser constructs (Supplementary Fig. 2e,f). Moreover, the model predicts an increase in phase separation propensity with decreasing pH, in agreement with cloud point

measurements (Supplementary Fig. 2f). Next, we interrogated the model on the effect of decreasing pH on the intermolecular contacts within condensates. The ratio between the contact maps predicted from simulations at pH 6.0 and 8.0 (Fig. 2g) shows enhanced interactions of the HClust with regions rich in negatively charged residues (positions 72–147 and 424–448) (Fig. 2h). The simulations thus reveal a pH-dependent rearrangement of the network of intermolecular interactions within condensates that favors contacts between His and residues located in Asp- and Glu-rich regions whereas contacts between like-charged His and Arg residues in the HClust and microexon 4 (positions 403–410) are disfavored (Fig. 2 i,j and Supplementary Fig. 2g).

nCPEB4-NTD populates multimers on the phase separation pathway

To investigate which residues of nCPEB4-NTD drive its phase separation, we used solution nuclear magnetic resonance (NMR) spectroscopy, a technique that can provide residue-specific information on the conformation and interactions of intrinsically disordered proteins[26]. Indeed, when conditions are insufficient for phase separation, the same interactions driving the phase transition can cause the monomer to collapse into a state[27] that can be characterized at residue resolution, thus providing this key information. Under such conditions (Supplementary Fig. 3a), nCPEB4-NTD has a spectrum characteristic of a collapsed intrinsically disordered protein, where only 13% of the NMR signals are visible (Supplementary Fig. 3b, in lime green). We analyzed the sample by dynamic light scattering (DLS) to confirm its monomeric nature. However, we detected only nCPEB4-NTD multimers with a hydrodynamic diameter of *ca* 55 nm, which is much larger than that predicted for a monomer[28] (*ca* 11 nm, Fig. 3a and Supplementary Fig. 3c).

To characterize the multimers, which have also been observed for other phase-separating RNA binding proteins with intrinsically disordered domains[29,30], we first determined whether their formation is reversible. To this end, we performed DLS analysis of increasingly diluted samples of nCPEB4-NTD under the same solution conditions. We observed that at 0.5 μM the multimers appear to be in equilibrium with a species with a hydrodynamic diameter close to that predicted for the monomer (Fig. 3a, in green). Next, at this concentration, we modified the solution conditions to favor phase separation by increasing the temperature and ionic strength. We observed a decrease in the signal corresponding to the monomer and an increase in that corresponding to multimers, thus indicating a shift of the multimerization equilibrium (Fig. 3a). This result suggests that the nCPEB4-NTD multimers are stabilized by the same types of interactions as the droplets (Fig. 1i,j).

These findings prompted us to study whether these multimers grow over time to become droplets observable by optical and fluorescence microscopy, that is, whether they represent intermediates in the phase separation pathway. To this end we first analyzed, over 15 h, samples of freshly prepared nCPEB4-NTD multimers by DLS. We observed a progressive increase in size and polydispersity up to an average hydrodynamic diameter of *ca* 90 nm and a polydispersity index of *ca* 0.10 (Supplementary Fig. 3d). We also determined the morphology of the multimers at the nanoscale by liquid-phase transmission electron microscopy (LPEM), obtaining that they are spherical and have a diameter of 30 to 50 nm (Fig. 3b and Supplementary Fig. 3e). Finally, super-resolution microscopy confirmed that the multimers grow over time to give rise to larger spherical particles resembling droplets (Fig. 3c).

nCPEB4-NTD multimerization is driven by interactions involving His and Arg residues

We next carried out experiments to determine the nature of the species giving rise to the residue-specific NMR signals observed in the presence of multimers. To this end, we analyzed a sample of multimers by size exclusion chromatography coupled to multi angle light scattering (SEC-MALS). This approach showed that the multimers are assemblies of *ca* 350 nCPEB4-NTD molecules (Supplementary Fig. 3f), corresponding to a molecular weight of *ca* $1.7 \cdot 10^7$ Da. Next, we performed ^{15}N-edited X-STE experiments to measure the diffusion coefficient of the species giving rise to the signals detected by NMR, from which we derived that they diffuse much more slowly than a monomer, with a value in semi-quantitative agreement with that obtained by DLS and LPEM (Supplementary Fig. 3g), thereby indicating that they correspond mainly to multimers. We concluded that solution NMR can be used to identify the most flexible regions of the nCPEB4-NTD sequence under conditions where it forms multimers, presumably corresponding to those least involved in the interactions that stabilize them[31–33].

To assign the NMR signals to specific residues, we progressively added urea and observed that, at 1.5–2 M (the urea concentration necessary to dissociate the multimers, Fig. 3d), the signals of most residues became detectable (Fig. 3e and Supplementary Fig. 3b) and stemmed from a species with the diffusion coefficient expected for a nCPEB4-NTD monomer (Supplementary Fig. 3g). Further additions of urea led to increases in intensity that allowed their assignment and revealed, by comparison of the spectra obtained in the absence and presence of denaturant, that the signals observed for the multimer, presumably not involved in stabilizing them, corresponded to residues between positions *ca* 50 and 150 in the sequence of nCPEB4-NTD. This region of sequence is devoid of aromatic and positively charged residues and instead rich in negatively charged ones (Fig. 3e,f). Despite the presence of 4 M urea, the spectrum of monomeric nCPEB4-NTD had a wide range of intensities. An

analysis of signal intensity as a function of residue type revealed especially low values for His and Arg (Fig. 3g), suggesting that these residue types are involved in transient interactions and explaining the very low signal intensity of the His-rich HClust (residues 229–252) and the Arg-rich microexon 4 (residues 403–410).

To facilitate the interpretation of these experiments, we monitored the formation of nCPEB4-NTD multimers in molecular simulations performed using the CALVADOS model[24,25]. Contacts calculated between a chain in the center of the multimer and the surrounding chains in assemblies of 130–170 chains show that the C-terminal region, including HClust and microexon 4, is involved in more favorable intermolecular interactions than the N-terminal Asp- and Glu-rich region (positions 72–147) (Fig. 3h). The simulations thus support the conclusion that the decreased NMR signal intensity in the C-terminal region reflects an increased number of contacts of these residues within the condensates. Finally, to characterize the interactions stabilizing the multimers further, we measured ^{1}H,^{13}C-HSQC experiments to analyze the intensity of side chain signals without urea, when nCPEB4-NTD is multimeric, and in 4 M urea, when nCPEB4-NTD is instead monomeric (Fig. 3i and Supplementary Fig. 3h,i,j). We observed that the intensities, especially those of the signals corresponding to His, Trp, and Arg residues, were lower for the former than for the latter. We also recorded ^{1}H,^{15}N-CP-HISQC experiments to study the Arg side chain NH signals in the same samples and found that they were undetectable in the absence of the denaturant but could be observed in its presence (Supplementary Fig. 3k). Taken together, our results indicate that the multimers formed by nCPEB4-NTD on the phase separation pathway are stabilized mainly by interactions involving His and Arg residues that are predominantly located in the C-terminal region (Supplementary Fig. 3l).

Microexon 4 kinetically stabilizes nCPEB4 condensates against aggregation

HClust and microexon 4 have a particularly high density of aromatic and positively charged residues, respectively (Fig. 3f), and the signals of the residues that they harbor were barely detectable in the NMR spectra of nCPEB4-NTD (Fig. 3e). We thus hypothesized that these regions, especially microexon 4 (Fig. 3h), play a particularly important role in the collapse of the nCPEB4-NTD monomer, as well as in its multimerization and phase separation. To study this, we measured the stability of the multimers and droplets formed by deletion variants of nCPEB4-NTD in which we removed HClust (ΔHClust) or microexon 4 (nCPEB4Δ4), or both regions (Δ4_ΔHClust) (Fig. 4a). We found that they all formed multimers but that their stability against disassembly by urea was lower than that of nCPEB4. Whereas the nCPEB4 multimers disassembled in 1.9 M urea, ΔHClust, nCPEB4Δ4, and Δ4_ΔHClust required 1.4, 0.9, or 0.2

M urea, respectively (Fig. 4b and Supplementary Fig. 4a). Next, we studied the phase separation properties of the variants and found that the cloud points of ΔHClust, nCPEB4Δ4, and Δ4_ΔHClust were *ca* 10, 20 and 40 ºC higher than that of nCPEB4, respectively (Fig. 4b). We thus concluded that both regions of the sequence drive the multimerization and phase separation of nCPEB4-NTD (Fig. 3i).

To confirm that microexon 4 is also relevant in neurons, we measured the number and size of the condensates formed by nCPEB4-NTD and nCPEB4Δ4-NTD in N2a cells. In agreement with the results obtained *in vitro,* the former formed more condensates than the latter (Fig. 4c,d); this was neither due to differences in interactions with other proteins (Supplementary Fig. 4b) nor with RNA (Supplementary Fig. 4c). Next, we addressed whether microexon 4 affected the dynamics of condensation by analyzing the fate of the condensates formed by both variants upon depolarization (Fig. 4e). Strikingly, we observed that the condensates formed in neurons overexpressing full-length nCPEB4Δ4 dissolved in a lower fraction of cells than those formed in neurons expressing nCPEB4 (Fig. 4f), thus indicating a decrease in the reversibility of condensation in the absence of the microexon. This different behavior was not due to post-translational modifications (PTMs), which did not change in either variant upon depolarization (Supplementary Fig. 4d), or to an intrinsically different response to the effect of pH on phase separation propensity (Supplementary Fig. 4e).

To determine the molecular basis of this observation, we studied the reversibility of nCPEB4-NTD phase separation *in vitro* in the presence and absence of microexon 4. When we decreased the temperature of a solution containing nCPEB4-NTD droplets, complete dissolution occurred. However, for a solution containing nCPEB4Δ4-NTD droplets under the same conditions only partial dissolution was observed, accompanied by the formation of aggregates (Fig. 4g). We then carried out repeated temperature cycles consisting of heating, cooling and centrifugation steps (Supplementary Fig. 4f), and plotted the cloud point of the sample as a function of cycle number. For nCPEB4-NTD, the cloud point measured for each cycle (Tc_i) was constant ($Tc_i \approx Tc_0$), indicating that phase separation was fully reversible, whereas for nCPEB4Δ4-NTD it progressively increased ($Tc_{i+1} > Tc_i$), indicating that the dissolution of the condensates was incomplete, presumably as a result of aggregation (Fig. 4h and Supplementary Fig. 4g,h). On the basis of these observations, we hypothesized that the condensation of nCPEB4Δ4-NTD leads to the formation of irreversible aggregates whereas in nCPEB4-NTD this process is slowed down due to the presence of microexon 4.

To test this hypothesis, we analyzed the fate of the droplets formed by the NTD of nCPEB4, nCPEB4Δ4, ΔHClust, and Δ4_ΔHClust by fluorescence microscopy after 24 h of incubation *in vitro*. The droplets formed by nCPEB4 did not show an altered morphology whereas those

formed by nCPEB4Δ4 transitioned into aggregates (Fig. 4i). Most importantly, we observed that deletion of the His cluster from nCPEB4Δ4-NTD (Δ4_ΔHClust) prevented aggregation, thereby indicating that this region of sequence is responsible for this process and that presence of microexon 4 in the sequence inhibits it quantitatively, even after one week of incubation (Fig. 4i,j and Supplementary Fig. 4i). Finally, we assessed the effect of HClust removal from the nCPEB4Δ4 isoform in N2a cells and observed a decrease in the fraction of cells with condensates after depolarization from 49 to 22 %, recovering the nCPEB4 phenotype (Fig. 4k and Supplementary Fig. 4j). Taken together, these observations indicate that microexon 4 kinetically stabilizes the nCPEB4 condensates against HClust-driven aggregation, allowing their dissolution after depolarization.

A mechanistic model for ASD

The nCPEB4 and nCPEB4Δ4 isoforms are co-expressed in neurons, and relative increases in the mole fraction of nCPEB4Δ4 appear to have a dominant-negative effect on the translation of mRNAs regulated by nCPEB4 in ASD[5]. To understand how a partial skipping of microexon 4 inhibits the function of this protein, we first studied whether the two isoforms are found in the same condensates. Co-expression of nCPEB4 and nCPEB4Δ4 in N2a cells, as well as *in vitro* experiments carried out with purified variants, showed co-localization, with similar partition coefficients (Fig. 5a,b). Next, to study the properties of droplets containing both isoforms, we prepared samples of nCPEB4Δ4-NTD at mole fractions ($\chi\Delta4$) ranging between 0 (corresponding to pure nCPEB4-NTD) and 1 (corresponding to pure nCPEB4Δ4-NTD) and measured their cloud points (Tc). The Tc values obtained were intermediate between those of nCPEB4-NTD and nCPEB4Δ4-NTD, with a linear dependence on $\chi\Delta4$, and the values corresponding to the mole fractions of control individuals (Ctrl, $\chi\Delta4 = 0.30$) and ASD patients (ASD, $\chi\Delta4 = 0.45$) were similar (29.2 and 31.4 ºC, respectively, Fig. 5c).

We then measured the degree of aggregation (Agg) of the same mixtures after 24 h of incubation *in vitro*. This parameter ranged between that of pure nCPEB4-NTD ($\chi\Delta4 = 0$, Agg ≈ 0%) and that of pure nCPEB4Δ4-NTD ($\chi\Delta4 = 1$, Agg ≈ 20%). However, contrary to the linear dependence of Tc on $\chi\Delta4$, the dependence of Agg on $\chi\Delta4$ was sigmoidal, with an abrupt increase of Agg at $\chi\Delta4 \approx 0.4$ (Fig. 5d). This finding indicates that samples with $\chi\Delta4$ lower than 0.4 behave, as far as aggregation is concerned, similarly to samples of pure nCPEB4. In contrast, samples with $\chi\Delta4$ higher than 0.4 behave in a similar way to those of pure nCPEB4Δ4 (Fig. 5e and Supplementary Fig. 5a). Strikingly, the $\chi\Delta4$ value defining the boundary between these two regimes ($\chi\Delta4^{crit}$) is intermediate between that of control individuals and that of ASD patients, thereby providing a plausible rationale of why small changes in $\chi\Delta4$ caused by mis-

splicing can lead to large changes in the functional properties of this protein and, potentially, to the onset of ASD.

Our results indicate that the Arg residues of microexon 4 inhibit aggregation by kinetically stabilizing nCPEB4 condensates relative to aggregates by shifting the balance between homo- and heterotypic interactions[14]. The boundary between the liquid and aggregated regimes ($\chi\Delta4^{crit}$) of nCPEB4 condensates containing both isoforms (Fig. 5d) corresponds to the value of $\chi\Delta4$ associated with the minimal concentration of free HClust (C_{Agg}^{crit}), not interacting with microexon 4, that must be reached for aggregation in our experimental timescale (Fig. 5f). In this scenario, it would be possible to reduce the rate of nCPEB4-NTD aggregation by decreasing the concentration of free HClust in the condensates. To this end, we designed a synthetic peptide containing two repetitions of the sequence of microexon 4, connected by a $(Gly\text{-}Ser)_3$ linker (Fig. 5g), and studied its effect on the droplets formed by a sample with the composition associated with ASD ($\chi\Delta4 = 0.45$). We observed that the peptide partitioned in the droplets (Fig. 5h and Supplementary Fig. 5b), increased their thermodynamic stability (Supplementary Fig. 5c), and, most importantly, increased their kinetic stability against aggregation in a concentration-dependent manner (Fig. 5i and Supplementary Fig. 5d).

Discussion

We found that microexon 4, which is rich in Arg residues, increases the kinetic stability of condensed CPEB4 by interacting with a cluster of His residues in the center of the NTD (HClust, 229–252). This heterotypic interaction is essential for the expression of ASD risk genes because it preserves the reversibility of CPEB4 condensation, which regulates their translation. Indeed, in the absence of the microexon, homotypic interactions between aromatic clusters led to the irreversible aggregation of CPEB4, presumably as a result of the high concentration of this protein in the condensates[34–36]. This finding illustrates how a failure of heterotypic buffering, a generic mechanism to preserve in cells the dynamic character of biomolecular condensates proposed by Taylor and co-workers[14], can lead to protein aggregation and disease (Fig. 6).

CPEB4 and nCPEB4Δ4 coexist in neurons, and the changes in the polyadenylation of ASD risk genes associated with this disease are caused by a subtle decrease in the degree of inclusion of microexon 4, which generates a dominant-negative effect[5]. We found that the propensity of mixtures of CPEB4 and nCPEB4Δ4 to aggregate depends on the concentration of His clusters available for homotypic interactions. Given that microexon 4 can also engage in interactions with such clusters, thus competing with homotypic interactions, this

concentration decreases with the degree of inclusion of microexon 4 (Fig. 5 and 6). However, the dependence is sigmoidal, thereby indicating that a degree of inclusion of *ca* 40% is necessary and sufficient to preserve the liquid character of the condensates and the reversibility of the phase transition. The finding that a critical concentration of aromatic clusters not involved in interactions with microexon 4 must be attained before aggregation suggests that the aggregation process observed is a liquid-to-solid phase transition, as proposed for other protein aggregation diseases[34,35] (Fig. 5).

We found that the minimal degree of microexon 4 inclusion necessary to preserve the reversibility of the phase transition is intermediate between the values observed in the neurons of ASD patients and control individuals. This finding suggests that a low inclusion of this microexon leads to the onset of ASD because it triggers the aggregation of condensed CPEB4, low CPEB4 activities upon neuron depolarization and, as a consequence, low expression of ASD genes[5]. Importantly, this mechanism provides a rationale for the dominant-negative effect because a degree of inclusion of microexon below the minimum can lead to the cooperative and irreversible aggregation of both CPEB4 isoforms and, potentially, of other CPEB proteins found in the same condensates. This finding suggests that neurons, unlike non-neuronal cells, require a minimal inclusion of microexon 4 in CPEB4 to preserve the liquid character of phase-separated protein over extended periods.

Given that the nominal pK_a of the His side chain is close to physiological pH, small fluctuations in pH can alter the phase separation propensity of proteins enriched in this residue. This effect can be used to design pH-sensitive polymers for applications in bioengineering[37,38] and it explains how changes in pH alter the phase separation propensity of Fillagrin in skin barrier formation[39]. Similarly, stress conditions lead to pH fluctuations that trigger a phase transition in a yeast prion protein via protonation of Glu residues[40]. In nCPEB4, the changes in phase separation propensity and activity that occur upon depolarization are caused by local changes in intracellular pH. This mechanism contrasts with how CPEB4 activity is regulated during the cell cycle in non-neuronal cells, where condensate dissolution is caused by phosphorylations by cell-cycle-specific kinases such as Cdk1[11]. Interestingly, CPEB2 and 3 are also enriched in His residues (Supplementary Fig. 6a,b) and they contain the same neuron-specific microexon, thus suggesting that our conclusions regarding CPEB4 may apply to other members of this family of translational regulators[41,42].

Our work provides a striking example of how alternative splicing provides a means to tissue-specifically control the material properties of biomolecular condensates to ensure that key biological functions are properly regulated in different contexts[43]. In addition, it reveals how mis-splicing can alter the material properties of biomolecular condensates, thus leading to

disease, and also provides proof of concept of how this process can be prevented or even reversed. More specifically, our understanding of how a reduced inclusion of microexon 4 in CPEB4 generates a dominant-negative effect, as well as our observation that the normal function of the microexon can be restored in trans, opens up new therapeutic opportunities for ASD based on the regulation of the dynamics of biomolecular condensates by drug-like small molecules and peptides[44].

Acknowledgements

We thank Benedetta Bolognesi, Isabel Fariñas, Ernest Giralt, Jordina Guillén-Boixet, Denes Hnisz, Manuel Irimia, José J. Lucas and Guillermo Montoya for critically reading the manuscript and making useful suggestions. We acknowledge support from the Advanced Digital Microscopy Facility and Mass Spectrometry and Proteomics Facility at IRB Barcelona and the Spanish ICTS *Red de Laboratorios de RMN de biomoléculas* (R-LRB). We thank Antonio García de Herreros and Antonio Zorzano for kindly providing NmumG and C2C12 cell lines, respectively. We thank Isabel Latorre and Joan Miquel Valverde for help with protein expression and purification. C.G-C. acknowledges a graduate fellowship from MINECO (PRE2018-084684). A.Bar. acknowledges a graduate fellowship from MINECO (PRE2018-083790). K.L.-L. acknowledges support from the Novo Nordisk Foundation (NNF18OC0033950). This project has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101025063. R.M. was supported by grants from the Spanish Ministry of Science, Innovation and Universities (PID2020-119533GB-I00 and PDC2021-121716-I00), and the Spanish Association Against Cancer (AECC) (GCB15152955MÉND), the Worldwide Cancer Research Foundation (20-0284), World Cancer Research Fund International (2020_021), the BBVA Foundation (BBVABIOMED/18), La Caixa Foundation (LCF/PR/HR19/52160020), and the Fundacio Marato TV3 (201926-30-31). X.S. was supported by AGAUR (2017 SGR 324), MINECO (BIO2015-70092-R and PID2019-110198RB-I00), and the European Research Council (CONCERT, contract number 648201). IRB Barcelona is the recipient of a Severo Ochoa Award of Excellence from MINECO (Government of Spain).

Author contributions

Conceptualization: C.G-C., A.Bar., R.M., X.S.; Methodology: C.G-C., A.Bar., A.Bal., G.T., B.D-A., M.F-A., J.M., C.D.P.; Investigation and formal analysis: C.G-C., A.Bar., A.Bal., G.T., B.D-A., M.F-A., C.D.P., L.R-P.; Supervision: J.G., L.R-P., G.B., K.L-L., R.M., X.S.; Writing - Original draft: C.G-C., A.Bar., R.M., X.S.; Writing - Review and editing: all authors; Funding acquisition: G.B., K.L-L., R.M., X.S.

Declaration of interests

X.S. is a scientific founder and advisor of Nuage Therapeutics.

Main references

1. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
2. Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell* **76**, 329–345 (2019).
3. Gonatopoulos-Pournatzis, T. *et al.* Autism-Misregulated eIF4G Microexons Control Synaptic Translation and Higher Order Cognitive Functions. *Mol. Cell* **77**, 1176–1192.e16 (2020).
4. Gonatopoulos-Pournatzis, T. & Blencowe, B. J. Microexons: at the nexus of nervous system development, behaviour and autism spectrum disorder. *Curr. Opin. Genet. Dev.* **65**, 22–33 (2020).
5. Parras, A. *et al.* Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing. *Nature* **560**, 441–446 (2018).
6. Ollà, I. *et al.* Pathogenic mis-splicing of CPEB4 in schizophrenia. *bioRxiv* 2022.09.22.508890 (2022) doi:10.1101/2022.09.22.508890.
7. Duran-Arqué, B. *et al.* Comparative analyses of vertebrate CPEB proteins define two subfamilies with coordinated yet distinct functions in post-transcriptional gene regulation. *Genome Biol.* **23**, 192 (2022).
8. Ivshina, M., Lasko, P. & Richter, J. D. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annu. Rev. Cell Dev. Biol.* **30**, 393–415 (2014).
9. Weill, L., Belloc, E., Bava, F.-A. & Méndez, R. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.* **19**, 577–585 (2012).

10. Afroz, T. *et al.* A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev.* **28**, 1498–1514 (2014).
11. Guillén-Boixet, J., Buzon, V., Salvatella, X. & Méndez, R. CPEB4 is regulated during cell cycle by ERK2/Cdk1-mediated phosphorylation and its assembly into liquid-like droplets. *Elife* **5**, (2016).
12. Wang, X.-P. & Cooper, N. G. F. Comparative in silico analyses of cpeb1-4 with functional predictions. *Bioinform. Biol. Insights* **4**, 61–83 (2010).
13. Tsang, B., Pritišanac, I., Scherer, S. W., Moses, A. M. & Forman-Kay, J. D. Phase Separation as a Missing Mechanism for Interpretation of Disease Mutations. *Cell* **183**, 1742–1756 (2020).
14. Mathieu, C., Pappu, R. V. & Taylor, J. P. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* **370**, 56–60 (2020).
15. Shin, J., Salameh, J. S. & Richter, J. D. Impaired neurodevelopment by the low complexity domain of CPEB4 reveals a convergent pathway with neurodegeneration. *Sci. Rep.* **6**, 29395 (2016).
16. Marzinke, M. A. & Clagett-Dame, M. The all-trans retinoic acid (atRA)-regulated gene Calmin (Clmn) regulates cell cycle exit and neurite outgrowth in murine neuroblastoma (Neuro2a) cells. *Exp. Cell Res.* **318**, 85–93 (2012).
17. Raj, B. *et al.* A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* **56**, 90–103 (2014).
18. Martin, E. W. & Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **57**, 2478–2487 (2018).
19. Dreier, J. P. & Reiffurth, C. The stroke-migraine depolarization continuum. *Neuron* **86**, 902–922 (2015).
20. Chesler, M. Regulation and modulation of pH in the brain. *Physiol. Rev.* **83**, 1183–1221 (2003).
21. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
22. Basu, S. *et al.* Rational optimization of a transcription factor activation domain inhibitor. *bioRxiv* (2022) doi:10.1101/2022.08.18.504385.
23. Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**, (2018).
24. Tesei, G., Schulze, T. K., Crehuet, R. & Lindorff-Larsen, K. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
25. Tesei, G. & Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Res. Eur.* **2**, 94 (2023).

26. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622 (2004).
27. Bremer, A. *et al.* Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* (2021) doi:10.1038/s41557-021-00840-w.
28. Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).
29. Martin, E. W. *et al.* A multi-step nucleation process determines the kinetics of prion-like domain phase separation. *Nat. Commun.* **12**, 4513 (2021).
30. Kar, M. *et al.* Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202222119 (2022).
31. Christodoulou, J. *et al.* Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10949–10954 (2004).
32. Baldwin, A. J. *et al.* Cytochrome display on amyloid fibrils. *J. Am. Chem. Soc.* **128**, 2162–2163 (2006).
33. Klein-Seetharaman, J. *et al.* Long-range interactions within a nonnative protein. *Science* **295**, 1719–1722 (2002).
34. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
35. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
36. Peskett, T. R. *et al.* A liquid to solid phase transition underlying pathological huntingtin Exon1 aggregation. *Mol. Cell* **70**, 588–601.e6 (2018).
37. Mackay, J. A., Callahan, D. J., Fitzgerald, K. N. & Chilkoti, A. Quantitative model of the phase behavior of recombinant pH-responsive elastin-like polypeptides. *Biomacromolecules* **11**, 2873–2879 (2010).
38. Callahan, D. J. *et al.* Triple stimulus-responsive polypeptide nanoparticles that enhance intratumoral spatial distribution. *Nano Lett.* **12**, 2165–2170 (2012).
39. Quiroz, F. G. *et al.* Liquid-liquid phase separation drives skin barrier formation. *Science* **367**, (2020).
40. Franzmann, T. M. *et al.* Phase separation of a yeast prion protein promotes cellular fitness. *Science* **359**, (2018).
41. Picó, S. *et al.* CPEB alteration and aberrant transcriptome-polyadenylation lead to a treatable SLC19A3 deficiency in Huntington's disease. *Sci. Transl. Med.* **13**, eabe7104 (2021).
42. Parras, A. *et al.* Polyadenylation of mRNA as a novel regulatory mechanism of gene

expression in temporal lobe epilepsy. *Brain* **143**, 2139–2153 (2020).

43. Garcia-Cabau, C. & Salvatella, X. Regulation of biomolecular condensate dynamics by signaling. *Curr. Opin. Cell Biol.* **69**, 111–119 (2021).
44. Biesaga, M., Frigolé-Vivas, M. & Salvatella, X. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr. Opin. Chem. Biol.* **62**, 90–100 (2021).

Main figures and legends

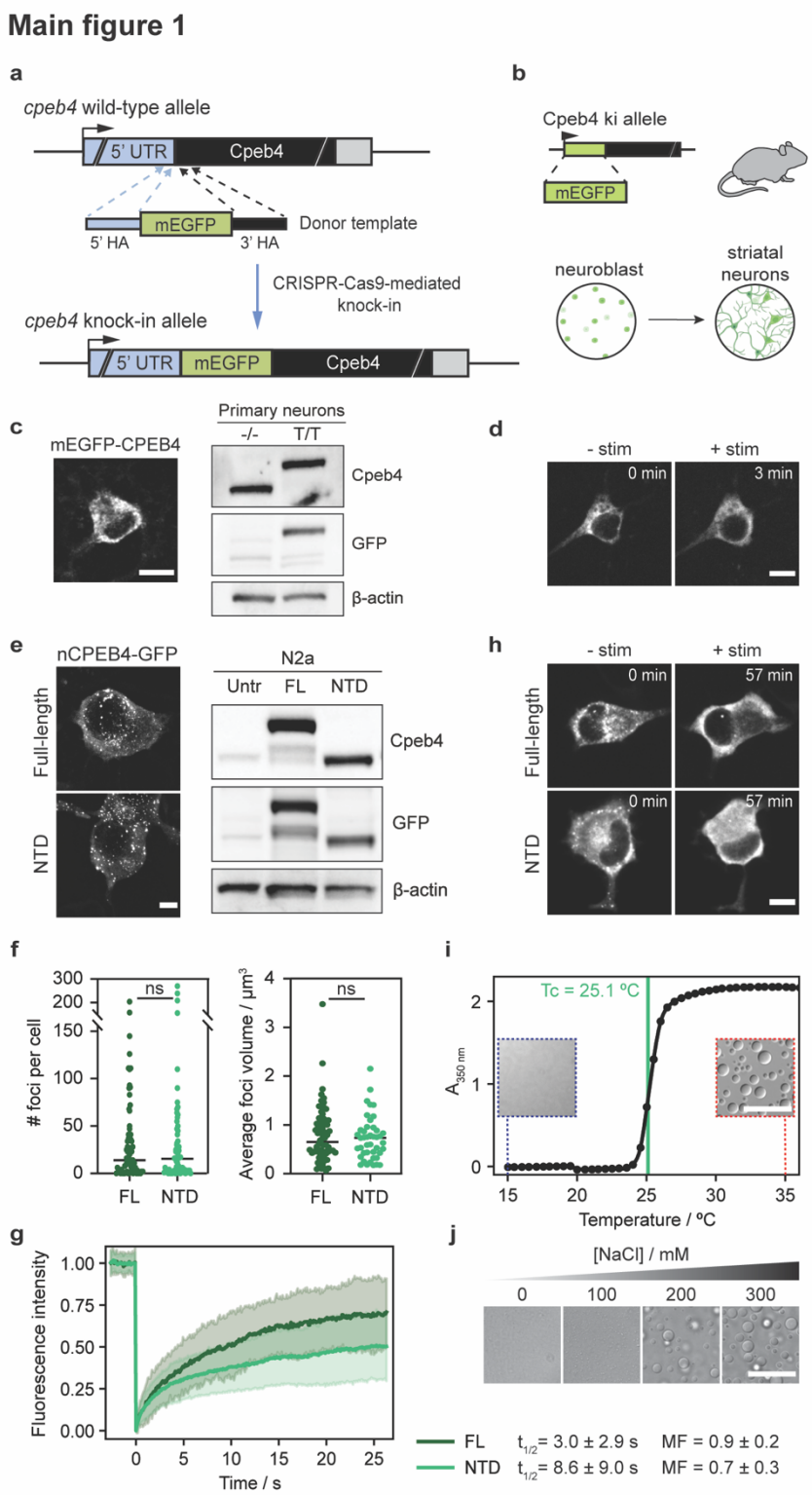


Figure 1 | nCPEB4 forms condensates that dissolve upon neuron depolarization. a) Schematic representation of CRISPR-Cas9-mediated mEGFP knock-in strategy in mice. **b)** Schematic representation of the generation of *ex vivo* differentiated striatal neurons from mEGFP-CPEB4 knock-in (ki) mice. Created with BioRender.com. **c)** Representative

fluorescence microscopy image of live *ex vivo* striatal neurons from mEGFP-CPEB4 mice. The scale bar represents 10 µm. Representative western blot of the expression of the endogenous GFP-tagged CPEB4 from mEGFP-CPEB4 mice (T/T) and untagged littermates (-/-). **d)** Representative fluorescence microscopy images from *ex vivo* live-cell time-lapse movies of striatal neurons from mEGFP-CPEB4 knock-in mice. Time series show pre (-stim, 0 min) and post (+stim, 3 min) stimulation foci disassembly. The scale bar represents 10 µm. **e)** Representative fluorescence microscopy images of differentiated Neuro2a (N2a) cells overexpressing either full-length (FL) nCPEB4-GFP or N-terminal domain (NTD). The scale bar represents 10 µm. Representative western blot of the expression of GFP-tagged CPEB4 FL and NTD. Untr: untransfected cells**. f)** Left panel: number (#) of cytoplasmic foci per cell of nCPEB4-GFP FL and NTD. Right panel: average foci volume (in μm^3) of nCPEB4-GFP FL and NTD (line represents the median, n = 3). **g)** Representative recovery curve from fluorescence recovery after photobleaching experiments of nCPEB4-GFP FL or NTD overexpressed in N2a cells. Half-life ($t_{1/2}$) and mobile fraction (MF) are calculated from three independent experiments. **h)** Representative fluorescence microscopy images from time-lapse movies of differentiated N2a cells overexpressing either nCPEB4-GFP FL or NTD. Time series show pre (-stim, 0 min) and post (+stim, 57 min) stimulation foci disassembly. The scale bar represents 10 µm. **i)** Apparent absorbance at 350 nm as a function of temperature of a solution of nCPEB4-NTD, where the cloud point (Tc) is represented by a green vertical line. Insets show DIC microscopy images of the same sample at the indicated temperatures. The scale bar represents 20 µm. **j)** DIC microscopy images of nCPEB4-NTD at increasing NaCl concentrations. The scale bar represents 20 µm.

Main figure 2

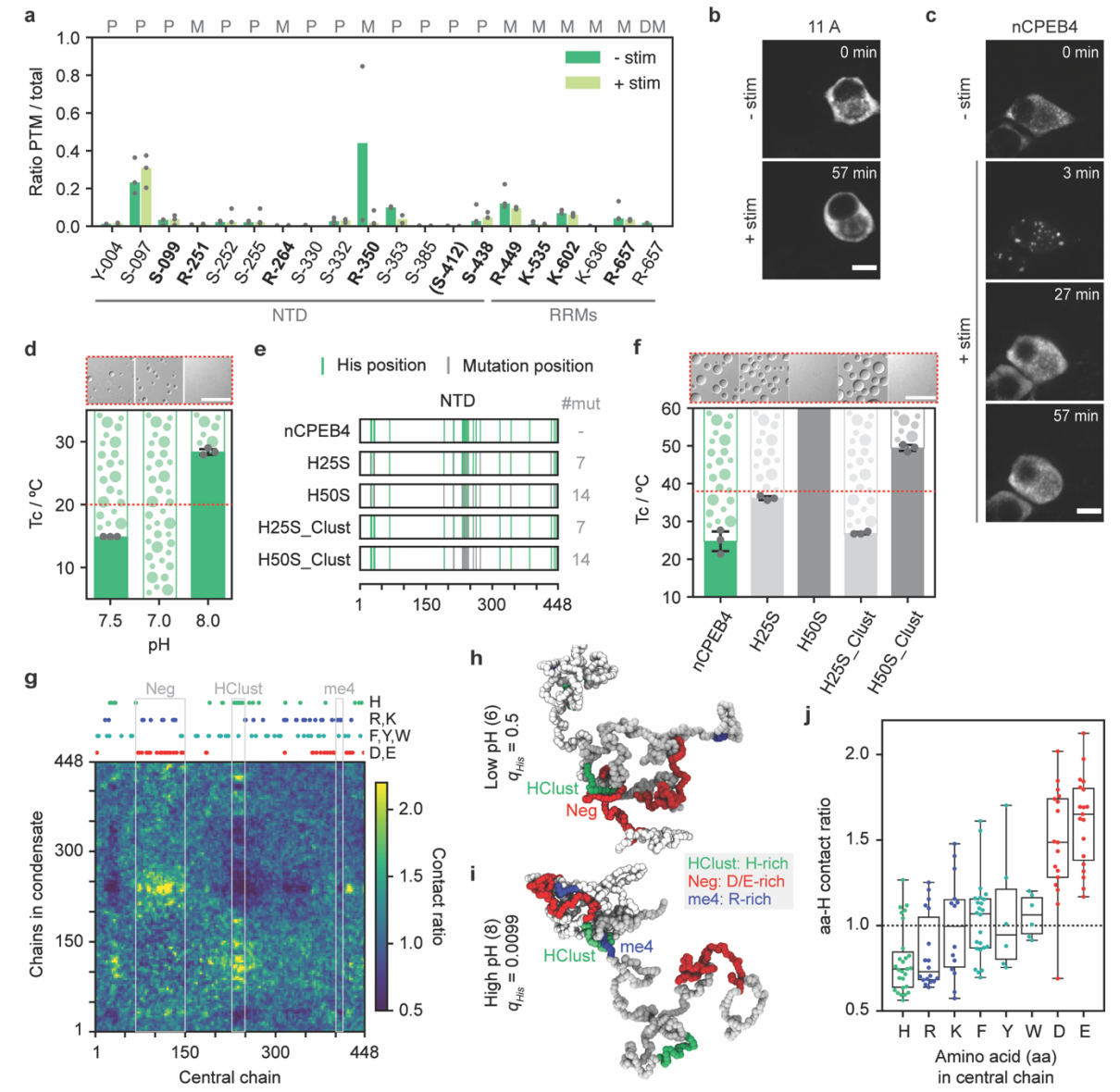


Figure 2 | The phase equilibrium of nCPEB4 is regulated by pH changes. a) nCPEB4 modified-to-total occurrence ratios determined by mass spectrometry for the indicated residue positions on basal (-stim) and stimulated (+stim) N2a cells overexpressing full-length nCPEB4-GFP. Ratios were calculated from 3 independent biological replicates. Bars represent the median. Previously undescribed post-translationally modified sites are highlighted in bold. P: phosphorylation, M: methylation, DM: dimethylation. **b)** Representative fluorescence microscopy images from live-cell time-lapse movies of N2a cells overexpressing the NTD of nCPEB4-GFP non-phosphorylatable mutant (11A). Time series showing pre (-stim, 0 min) and post (+stim, 57 min) stimulation foci disassembly. The scale bar represents 10 µm. **c)** Representative fluorescence microscopy images from live-cell time-lapse movies of N2a cells overexpressing full-length nCPEB4-GFP. Time series show pre (-stim, 0 min) and post (+stim,

3, 27, and 57 min) stimulation foci disassembly. The scale bar represents 10 µm. **d)** Cloud points (mean ± SD, n = 3) of 10 µM nCPEB4-NTD at pH 7.5, 7.0, and 8.0; and DIC microscopy images of the same samples at 20 ºC. The scale bar represents 20 µm. **e)** Comparison of the sequences of the His to Ser mutants with that of nCPEB4-NTD, where the positions of the mutations are indicated with gray vertical lines, the number of mutations (#mut) in each construct is indicated on the right, and the His residues remaining in the sequence are indicated with cyan vertical lines. **f)** Comparison of the cloud points (mean ± SD, n = 3) of nCPEB4-NTD with those of the His to Ser mutants, and DIC microscopy images of the same samples at 38 ºC. The scale bar represents 20 µm. **g)** Ratio of intermolecular contact maps at pH 6 and 8 calculated between a chain in the middle of the condensate and the surrounding chains. **h)** Simulation snapshot of two interacting nCPEB4-NTD chains in the center of a condensate at pH 6.0 showing contacts between HClust (green) and the Asp- and Glu-rich region at positions 72–147 (Neg, red). **i)** Simulation snapshot of two interacting nCPEB4-NTD chains in the center of a condensate at pH 8.0 showing contacts between HClust (green) and the Arg-rich microexon 4 (blue). **j)** Ratio of contacts at pH 6 and 8 between His, Arg, Lys, Phe, Tyr, Trp, Asp, and Glu residues on a chain in the middle of the condensate and the histidine residues on the surrounding chains.

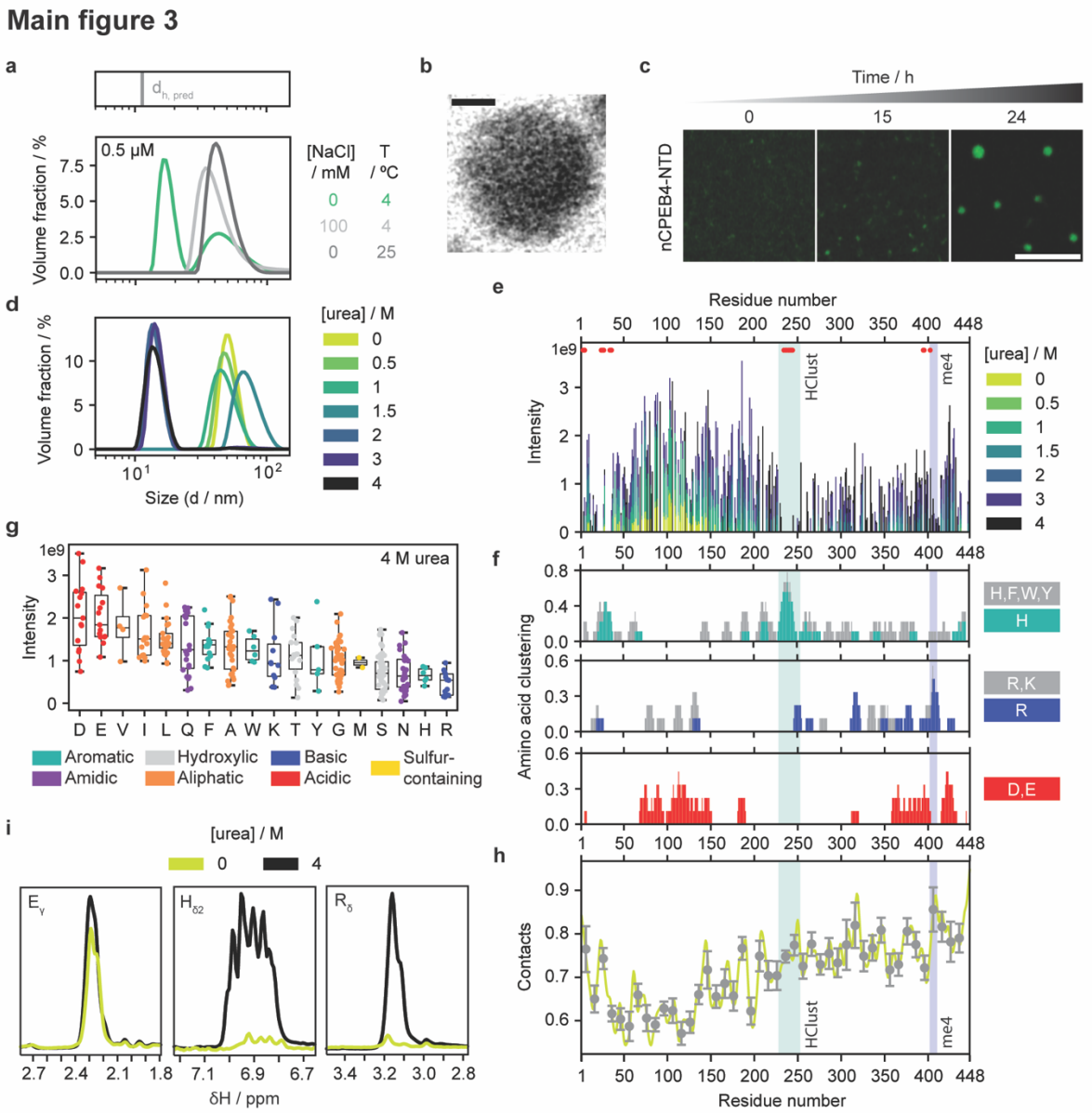


Figure 3 | nCPEB4-NTD forms multimers stabilized by interactions involving His and Arg residues. a) Top panel: predicted diameter of nCPEB4-NTD. Bottom panel: effect of protein concentration, temperature, and ionic strength on multimer formation monitored by DLS. **b)** Micrograph obtained via liquid-phase TEM of a multimeric particle in solution displaying spherical morphology. Dose 20 e-/Å^2. The scale bar represents 10 nm. **c)** Super-resolution microscopy images of the evolution of the multimers up to 24 h from sample preparation. The scale bar represents 5 μm. **d)** Urea titration, from 0 to 4 M, of the NMR samples in panel *e)* monitored by DLS. **e)** ^{1}H,^{15}N-CP-HISQC backbone amide peak intensities as a function of residue number at different urea concentrations. Shaded regions correspond to the His-rich cluster (HClust, 229–252, in cyan) and the microexon 4 (me4, 403–410, in blue). Not assigned residues are indicated as red dots. **f)** Amino acid clustering on the sequence using a nine amino acid window. Top panel: aromatic residues (gray) and His (cyan). Middle

panel: positively charged residues (gray) and Arg (blue). Bottom panel: negatively charged residues (red). **g)** Intensities of the peaks in the $^{1}H,^{15}N$-CP-HISQC spectrum at 4 M urea grouped by amino acid type ordered by the mean values. **h)** Total contacts formed between residues on a chain in the middle of a multimeric particle and all the residues in the surrounding chains. Contacts are obtained from molecular simulations averaging over multimers of 130–170 chains with an average radius of gyration of 17.5±0.9 nm. Circles and error bars are, respectively, averages and standard errors calculated over windows of ten residues. **i)** $^{1}H,^{13}C$-HSQC traces of representative amino acids (one letter code) side chain atoms (as Greek letters) in the absence (lime green) and presence (black) of 4 M urea.

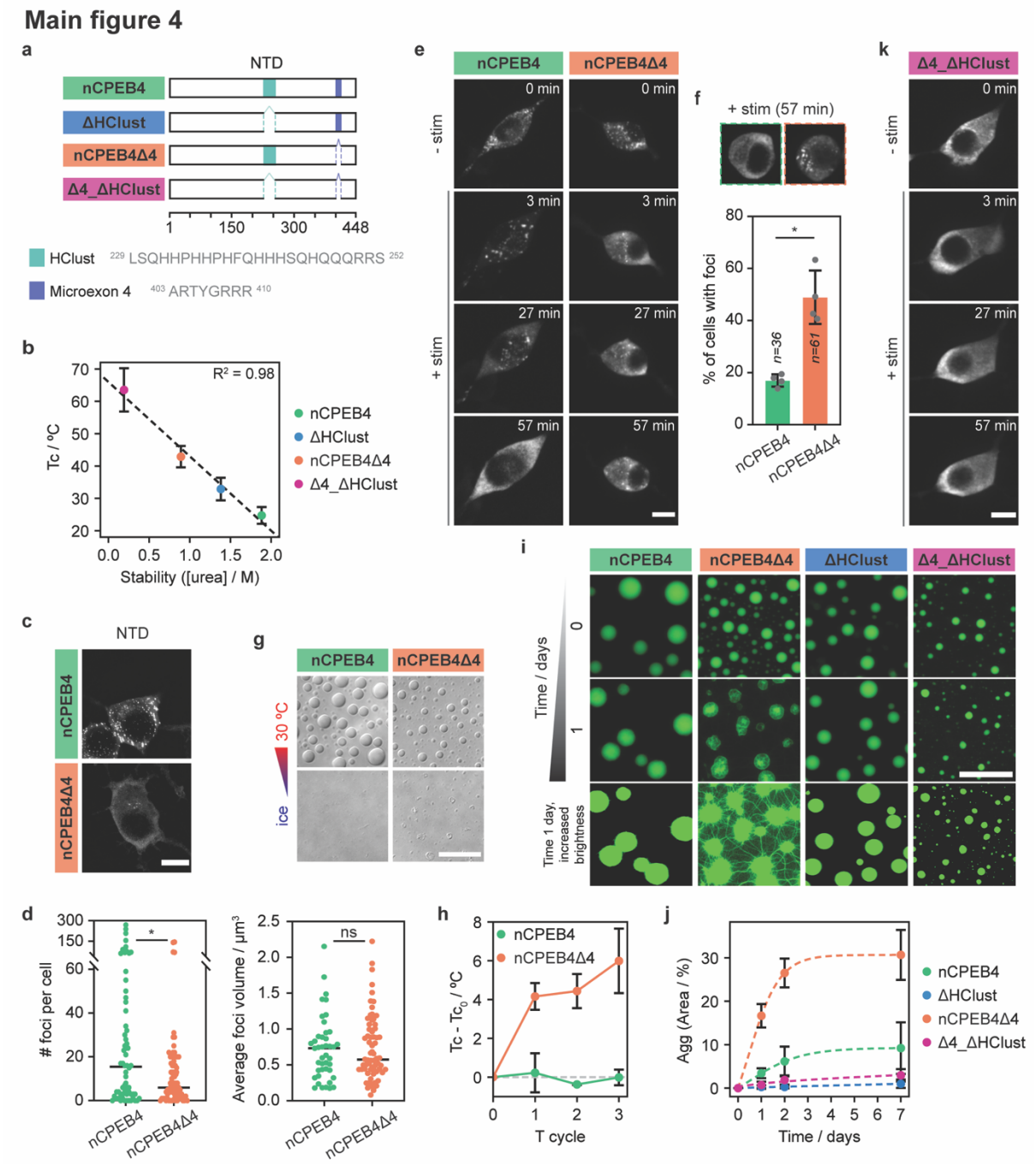


Figure 4 | Microexon 4 kinetically stabilizes the nCPEB4 condensates against aggregation. a) Scheme of nCPEB4-NTD isoforms and mutants. **b)** Correlation between the cloud points (mean ± SD, n = 3) and the stability of the 50 nm multimers (represented by the urea concentration needed for their dissolution, *Supplementary Fig. 4a*). The dotted line represents the fitting to a linear relationship. **c)** Representative fluorescence microscopy images of N2a cells overexpressing the GFP-tagged NTD of either nCPEB4 (with microexon 4) or nCPEB4Δ4 (the isoform lacking microexon 4). The scale bar represents 10 µm. **d)** Left panel: number (#) of cytoplasmic foci per cell of GFP-tagged nCPEB4-NTD and nCPEB4Δ4-NTD. Right panel: average foci volume (in μm^3) of GFP-tagged nCPEB4-NTD and

nCPEB4Δ4-NTD (line represents the median, n = 3). **e)** Representative fluorescence microscopy images from live-cell time-lapse movies of N2a cells overexpressing full-length nCPEB4-GFP and nCPEB4Δ4-GFP. Time series show pre (-stim, 0 min) and post (+stim, 3, 27, and 57 min) stimulation foci disassembly. The scale bar represents 10 µm. **f)** Top panel: representative fluorescence microscopy images of nCPEB4-GFP (left) and nCPEB4Δ4-GFP (right) at the end time point of the time-lapse in panel *e)*. Bottom panel: percentage of cells displaying cytoplasmic foci of nCPEB4-GFP and nCPEB4Δ4-GFP after depolarization (t = 57 min). Dots represent independent measurements analyzed by blind analysis from a pool of 6 experiments (mean ± SD). **g)** DIC microscopy images of nCPEB4-NTD and nCPEB4Δ4-NTD with 150 mM NaCl at 30 ºC and after decreasing the temperature with ice. The scale bar represents 20 µm. **h)** Temperature reversibility experiment by apparent absorbance measurement (mean ± SD, n = 3). Difference of cloud point (Tc) at each temperature cycle compared to the Tc value at 0 cycles ($Tc - Tc_0$). The gray dotted line indicates an ideally fully reversible process. **i)** Time evolution of nCPEB4-NTD, nCPEB4Δ4-NTD, ΔHClust, and Δ4_ΔHClust followed by fluorescence microscopy. The scale bar represents 20 µm. **j)** Quantification of the fluorescence stemming from aggregated material (Agg) in the images in panel *i)* and up to one week after sample preparation (mean ± SD, n = 9). **k)** Representative fluorescence microscopy images from live-cell time-lapse movies of N2a cells overexpressing GFP-tagged Δ4_ΔHClust mutant. Time series show pre (-stim, 0 min) and post (+stim, 3, 27, and 57 min) stimulation foci disassembly. The scale bar represents 10 µm.

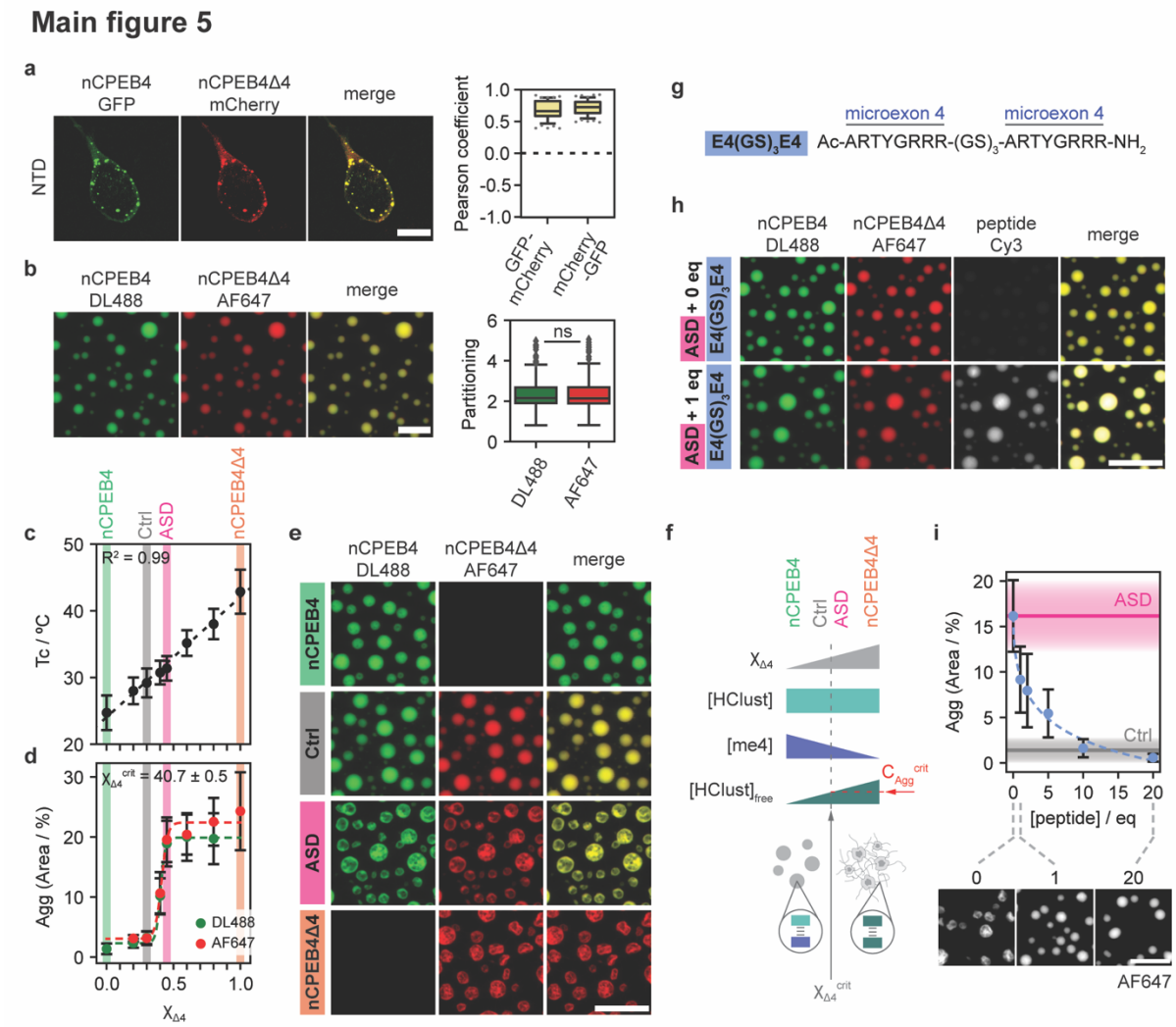


Figure 5 | A minimal concentration of microexon 4 in the nCPEB4 is sufficient to prevent aggregation. a) Left panel: representative fluorescence co-localization images of nCPEB4-GFP with nCPEB4Δ4-mCherry in N2a cells. The scale bar represents 10 µm. Right panel: magnitude of GFP-mCherry co-localization shown as Pearson coefficient determined with JaCoP (n = 3). **b)** Left panel: fluorescence microscopy images of a solution containing both nCPEB4-NTD and nCPEB4Δ4-NTD isoforms in a Δ4 mole fraction (χΔ4) of 0.5. The scale bar represents 20 µm. Right panel: partitioning of each isoform in the droplets (n = 378). **c)** Cloud point (Tc) of samples at different χΔ4 ranging from 0 to 1 (mean ± SD, n = 3). **d)** Quantification of the fluorescence stemming from aggregated material (Agg) of samples at different χΔ4 24 h after sample preparation (mean ± SD, n = 9). **e)** Representative images of panel *d)* for nCPEB4 (χΔ4 = 0), Ctrl (χΔ4 = 0.30), ASD (χΔ4 = 0.45), and nCPEB4Δ4 (χΔ4 = 1) samples. The scale bar represents 20 µm. **f)** Schematic representation of the effect of increasing χΔ4 (in gray) on the aggregation propensity of nCPEB4-NTD. Free HClust (in green) represents the amount of HClust not interacting with microexon 4 (me4, in blue). C_{Agg}^{crit} represents the critical concentration of free HClust necessary for aggregation, which defines

the boundary between the liquid and aggregated regimes ($\chi\Delta4^{crit}$). **g)** E4(GS)$_3$E4 peptide sequence. **h)** Fluorescence microscopy images of an ASD sample ($\chi\Delta4$ = 0.45) in the absence and presence of 1 molar equivalent of E4(GS)$_3$E4 peptide. The scale bar represents 20 µm. **i)** Top panel: quantification of the fluorescence stemming from aggregated material (Agg) of an ASD sample ($\chi\Delta4$ = 0.45) at increasing molar equivalents of E4(GS)$_3$E4 peptide 24 h after sample preparation (mean ± SD, n = 9). Bottom panel: representative images at 0, 1, and 20 molar equivalents of peptide. The scale bar represents 20 µm.

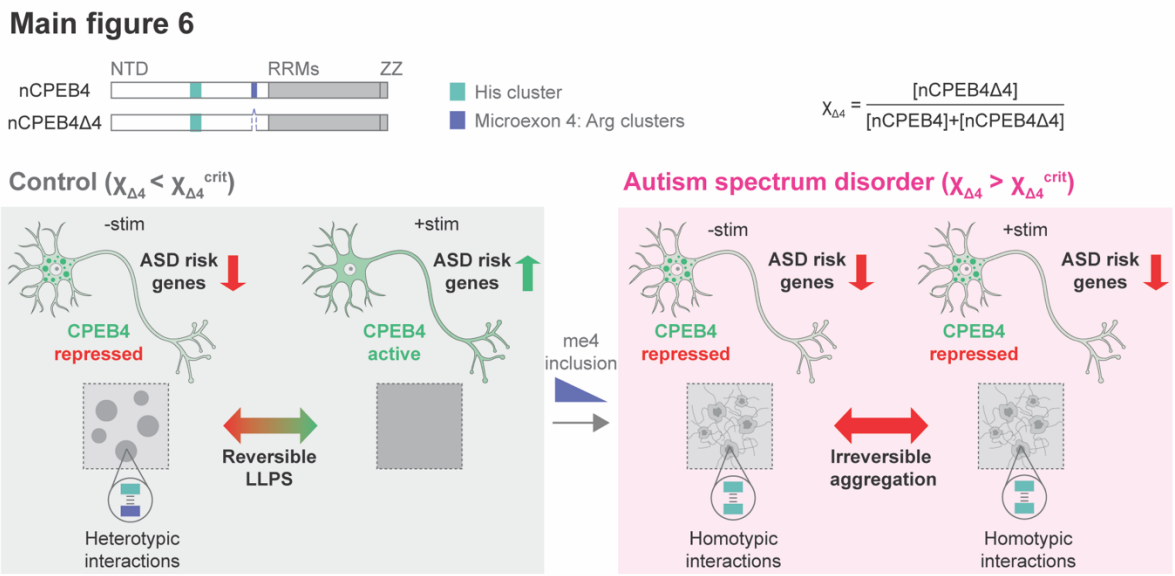


Figure 6 | Schematic illustration of how a low inclusion of microexon 4 (me4) in nCPEB4 can lead to the onset of autism spectrum disorder (ASD).

Methods

Generation of mEGFP-CPEB4 mice

A synthetic sequence consisting of mEGFP-linker sequence flanked by short regions of 5' homologous (238bp) and 3' homologous (99bp) DNA was obtained (Twist bioscience). PCR was carried out on the synthetic sequence using the following primers: Fw ssDNA-mGFP-CPEB4 (phosphorylated) *tacttcaagcaaacatatttgagatacagggga*; Rv ssDNA-mGFP-CPEB4 (thiol–protected) *GGTGATGGTGTGGAGGCTGC*. Single-stranded DNA (ssDNA) was generated from double-stranded DNA by lambda exonuclease digestion of the phosphorylated strand, followed by gel purification and column extraction. Animals were generated by electroporation of isolated mouse zygotes with ssDNA combined with Cas9 protein and guide/tracr RNA ribonuclear protein (RNP) complexes (guide; C45gRNA *ATCCTAAAAATAATAAATGG*). The correct integration of the knock-in cassette was confirmed by PCR and sequencing of the region. The resulting positive mice were crossed with C57BL6/J mice to confirm germline transmission. The offspring were maintained in a C57BL/6J background, and routine genotyping was performed by PCR using the following genotyping primers: *5' ACGTAGGGTGATAAGCTGTGAT 3' (Fw)* and *5' AGGGTCTTGTTGTTCTTGCTGT 3' (Rv)*. Mice were maintained in a specific pathogen-free (SPF) facility with a 12-h light–dark cycle at 21ºC ±1 and given ad *libitum* access to standard diet and water. Animal handling and all experimental protocols were approved by the Animal Ethics Committee at the Barcelona Science Park (PCB) and by the Government of Catalonia.

mEGFP-CPEB4 mice striatal neuron extraction and culture

mEGFP-CPEB4 mice over 6 weeks of age were crossed in timed matings. Females were weighed weekly to monitor gestation progression. Females with an increment over 3 grams up to 18 days after a positive plug were euthanized and embryos were collected at E18.5 on cold buffer containing 1x HBSS, 10 mM glucose, and 10 mM HEPES. A tail sample was also collected for embryo genotyping. Brains were dissected in the aforementioned buffer on an ice-cold plate and the striatum was extracted and chopped. Samples were centrifuged and digested with a previously heated solution containing 1x HBSS, 10 mM glucose, 10 mM HEPES, 12 U/mL papain (Worthington LS003180), and 5 mM L-cysteine for 15 min at 37ºC. Samples were then disaggregated in a buffer containing 1x DMEM F-12, 2 mM glutamine, 1 mM sodium pyruvate, 20 mM glucose, and 10% inactivated horse serum. Cells were seeded at a confluence of 25,000 cells per well in µ-Slide 8-well ibiTreat imaging plates (Ibidi, 80826) previously coated with poly-D-lysine. Cells were then incubated at 37ºC for 1 h. After this time, medium was exchanged by previously tempered media containing 1x Neurobasal (Gibco,

21103049), 1x B27 with vitamin A (Gibco, 17504044), 2 mM glutamine, and 0.5% penicillin/streptomycin (P/S). Medium was refreshed every 2/3 days. Neurons were considered differentiated after 7 days of culture. After genotyping, homozygous mEGFP-CPEB4 mice and wild-type littermates were selected for imaging. When specified, cell depolarization was induced by the addition of 50 mM potassium chloride (KCl) with 1:3 media dilution.

mEGFP-CPEB4 distribution in neurons

Primary striatal neurons from mEGFP-CPEB4 mice were imaged at 7 days of differentiation using a LIPSI Spinning Disk microscope (Nikon). Image acquisition was performed using a fully incubated, high-content, high-speed screening LIPSI platform (Nikon) equipped with an Eclipse Ti2 inverted microscope and a Yokogawa W1 confocal spinning disk unit. The spinning disk unit with Apo LWD 40x water lens of 1.15 NA and a 488 nm (20%) laser was used for acquisition on a Prime BSI Photometrics sCMOS camera. NIS Elements AR 5.30.05 software was used for acquisition, whereas Fiji-/ImageJ software was used to adjust images for visualization.

Plasmids for expression in N2a cells

Human nCPEB4 (UniProt ID: Q17RY0-1) full-length open reading frame (ORF) was cloned in pBSK vector. Microexon 4 sequence (nucleotides 1258-1281) was deleted by PCR on the pBSK-nCPEB4 plasmid using Gibson Assembly® Master Mix (New England Biolabs, E2611S), following the manufacturer's instructions. Mutagenesis of nCPEB4 phosphorylation sites was performed using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies, 210513), with oligonucleotides purchased from Sigma-Aldrich, following the manufacturer's instructions. ΔHClust mutants were generated by PCR mutagenesis on the pBSK-nCPEB4 plasmid, with oligonucleotides purchased from Sigma-Aldrich. For cell transfection, nCPEB4 ORF, full-length, NTD and mutants, were cloned in pPEU4 and pPEU5 vectors, which contain a C-terminal eGFP or mCherry tag, respectively, by In-Fusion (BD Clontech) cloning reaction[45]. For BioID, xCPEB4 or BirA ORF were cloned in pBSK vector. Microexon 4 was added to xCPEB4 sequence by PCR on the pBSK-xCPEB4 plasmid, using Gibson Assembly® Master Mix (New England Biolabs, E2611S) and following the manufacturer's instructions. A Myc-tag and BirA ORF were added at the N-terminus of pBSK-xCPEB4 plasmid. For competition experiments, xCPEB1 RRMZZ and xCPEB4 RRMs domains were cloned in pBSK and an HA-tag was added at the N-terminus. pBSK-Emi2 3'UTR was obtained from Belloc and Méndez, 2008[46].

N2a cell culture, differentiation, and DNA transient transfection

Neuro2a (N2a) cells were grown in Dulbecco's modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (P/S), and 2 mM L-glutamine for maintenance. For fixed-cell imaging, cells were seeded on 6-well plates with 12 mm-diameter poly-lysine-coated glass coverslips (Marienfeld Superior). For live-cell-imaging, cells were seeded on µ-Slide 8-well ibiTreat plates. For differentiation, medium was exchanged by DMEM with 0.5% FBS, 1% P/S, 2 mM L-glutamine, and 1 µM retinoic acid (RA) and cells were grown for 48 h. They were then transfected at 60% confluence with 1.25 µg of DNA using Lipofectamine LTX and Plus Reagent (Thermo Fisher, 15338100), following the manufacturer's protocol. When specified, N2a depolarization was induced as described for striatal neurons, specified in the "*mEGFP-CPEB4 mice striatal neuron extraction and culture*" section.

nCPEB4-GFP distribution in N2a cells

Twenty-four hours post-transfection, N2a cells were fixed with 4% paraformaldehyde (Aname, 15710) in phosphate-buffered saline (PBS) for 10 min at room temperature (RT). They were then washed with PBS and incubated with 0.5 µg/µl DAPI (Sigma) for 15 min. Coverslips were rinsed with PBS and mounted on a glass slide with Prolong Gold Antifade Mountant (P36934, Invitrogen). Image acquisition was performed with a Leica SP5 confocal microscope (Leica Microsystems), and Z-series stacks were acquired at 1024 x 1024 pixels using the 63 x/1.4NA oil immersion objective with a zoom factor of 2. Argon 488 nm (20%) and Diode 405 nm (10%) lasers were used. Hybrid detectors for GFP (500-550 nm with 33% gain) and DAPI (415-480 nm, 33% gain) were used for acquisition. LAS AF Leica software was used to acquire 10-20 Z-stack slices per cell with a z-step size of 0.5 µm. Fiji-/ImageJ software was used to perform the image analysis. A tailor-made macro using BioVoxxel Toolbox and 3D object counter plugins was used to accurately segment and obtain the number and volume of foci per cell. The statistical analysis of foci features was carried out using the non-parametric Mann-Whitney test. Differences were considered significant when the p-value was lower than 0.05 (*), 0.01 (**), 0.001 (***), or 0.0001 (****).

Live cell imaging of GFP-tagged CPEB4 variants in N2a cells

Live imaging of overexpressed nCPEB4-GFP variants in N2a cells was performed 20 h post-transfection, whereas primary striatal neurons from mEGFP-CPEB4 mice were imaged at 7 days of differentiation. For both types of cells, image acquisition was performed using a Spinning Disk microscope (Andor Revolution xD, Andor). A total of 24 images were taken per experiment (4 before the addition of the stimulus and 20 after), with 13 z-stacks at 512 x 512

pixels of format resolution. Images were acquired with a step size of 0.5 µm. For acquisition, typical frame rate was adjusted to 5 images per second at 50 ms integration time of the EMCCD camera (Andor). The Argon 488 nm laser (20%) was used for acquisition with a 1.4 NA/60x oil immersion objective. Fiji-/ImageJ software was used to obtain a z projection of the z-stacks and subsequent concatenation of images. The obtained time-lapses were subsequently used for manual quantification of nCPEB4 dissolution events. Cells were manually classified in two categories depending on the existence of cytoplasmic foci at t = 57 min: cells with remaining foci or cells without. For nCPEB4 and nCPEB4Δ4 comparison, the percentage of cells with remaining cytoplasmic foci after the depolarizing stimuli (t = 57 min) was calculated from a pool of 6 experiments. When specified, blind analysis and classification was performed independently by a group of 4 different people. For HClust-mutant experiments, the percentage of cells with remaining cytoplasmic foci after the depolarizing stimuli (t = 57 min) was calculated per each experiment and was normalized to nCPEB4. The statistical analysis of the percentage of cells with foci was carried out using the non-parametric Mann-Whitney test. Differences were considered significant when the p-value was lower than 0.05 (*), 0.01 (**), 0.001 (***), or 0.0001 (****).

Fluorescence recovery after photobleaching (FRAP) in N2a cells

A Spinning Disk microscope from Andor, equipped with a FRAPPA module was used for FRAP experiments. 350 images were taken by experiment (50 images before the bleaching and 300 after) at 512 x 512 pixels. The typical frame rate was set to the fastest (88 ms) with an exposure time of 50 ms on an EMCCD camera. An AOTF 488 nm laser (20%) was used for acquisition, while 50% laser intensity was set for bleaching in 2 repeats with a dwell time of 40 ms. The Fiji-/ImageJ software was used for FRAP analysis. Three regions of interest (ROIs) were defined per video: background, cell, and bleaching area. The mean fluorescence intensity was obtained for the three ROIs for all 350 frames and the output was exported in tabular format. Outputs were then entered on the easyFRAP website[47]. Full-scale normalization was selected, “initial values to discard” was set to 20, and the curves obtained were fitted to a single exponential model. Fluorescence recovery curves, mobile fraction, and half time of recovery were obtained for each experimental condition.

Mapping of nCPEB4 post-translational modified sites by mass spectrometry

Overexpressed nCPEB4-GFP and nCPEB4Δ4-GFP were immunoprecipitated from basal (-stim) and stimulated (+stim) N2a differentiated cells. Cells were lysed in ice-cold RIPA buffer containing 50 mM Tris HCl pH 8, 1% Nonidet P-40 (NP40), 0.1% SDS, 1 mM EDTA, 150 mM NaCl, 1 mM $MgCl_2$, 1x EDTA-free complete protease inhibitor cocktail (Roche, 5056489001) and phosphatase inhibitor cocktails (Sigma, P5726, and P0044). Cells were subsequently

sonicated for 5 minutes at low intensity with a Standard Bioruptor Diagenode. Following centrifugation (4ºC for 10 min at max speed), supernatants were collected, precleared and immunoprecipitated overnight at 4 ºC with 50 µL of GFP-conjugated Dynabeads Protein A (Invitrogen). Beads had previously been conjugated with 5 µL anti-GFP antibody (Invitrogen, A6455) for 2 hours at RT. After immunoprecipitation, beads were washed with cold RIPA buffer and eluted with Laemmli sample buffer. Eppendorf LoBind microcentrifuge tubes (Eppendorf, 30108116) were used for the whole protocol. The immunoprecipitated elutions were run on precast 4-20% gradient gels (Midi Criterion™ TGX™, Biorad) and stained with Coomassie blue for 1 h at RT. Bands at the expected nCPEB4-GFP molecular weight were cut, washed, with 50 mM NH_4HCO_3 and acetonitrile, reduced with 10 mM DTT and alkylated with 50 mM IAA. Samples were digested with trypsin and digestion was stopped by the addition of 5% formic acid. Following evaporation, samples were reconstituted 15 µL of 1% formic acid and 3% acetonitrile. Mass spectrometry analysis of nCPEB4 PTM sites was performed as previously described[7] with some modifications. Briefly, samples were loaded in a µ-precolumn at a flow rate of of 250 nL/min using Dionex Ultimate 3000. Peptides were separated using a NanoEase MZ HSS T3 analytical column with a 60-min run and eluted with linear gradient from 3 to 35% B in 60 min (A = 0.1% formic acid in H_2O and B = 0.1% formic acid in acetonitrile). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific). Spray voltage in the NanoMate source was set to 1.7 kV. The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. Survey MS scans were acquired in the orbitrap with the resolution (defined at 200 m/z) set to 120,000. The top speed (most intense) ions per scan were fragmented in the HCD cell and detected in the orbitrap.

For peptide identification, searches were performed using MaxQuant v1.6.17.0 software and run against target and decoy database to determine the false discovery rate (FDR). Database included proteins of interest sequences (nCPEB4-GFP, nCPEB4Δ4-GFP) and contaminants. Search parameters included trypsin enzyme specificity, allowing for two missed cleavage sites, oxidation in methionine, phosphorylation in serine/threonine/tyrosine, methylation and demethylation in lysine/arginine and acetylation in protein N-terminus as dynamic modifications, and carbamidomethyl in cysteine as a static modification. Peptides with a q-value lower than 0.1 and FDR < 1% were considered as positive identifications with a high confidence level. MS spectra were searched against contaminants (released in 2017) and user proteins using Andromeda and MaxQuant v1.6.17.0 software. To accept a site as modified, post-translational modification (PTM) localization probability was set above 75%. For the differential expression analysis, a T-test on PTM site intensities from MaxQuant was applied for each site within nCPEB4 variants. For data visualization, two parameters were

used for each PTM site, namely the sum of intensities of modified peptides that contain the specific PTM-site (*Int_mod*), and the PTM-to-base ratio, the latter calculated as: *Int_mod / Int_unmod*, where *Int_unmod* is the sum of intensities of unmodified peptides that contain the site. For data visualization, the PTM-to-total ratio was calculated for each site as follows:

PTM-to-total = Int_mod / (Int_mod + (Int_mod / PTM-to-base))

RNA extraction and real-time quantitative RT-PCR

For N2a RNA extraction, cells were scraped in an ice-cold plate, collected, and centrifuged at 500 g for 5 min at 4ºC. For mouse tissue RNA extraction, organs were ground with a liquid nitrogen-cooled mortar to obtain tissue powder. Total RNA was extracted from both cells and tissue powder by TRIsure reagent (Bioline, Ecogen, BIO-38033) following the manufacturer's protocol and using phenol-chloroform. RNA concentration was determined using Nanodrop Spectrophotometer (Nanodrop Technologies). 1 µg of total RNA was reverse transcribed using RevertAid Reverse Transcriptase (Themo Fisher, EP0442), following the manufacturer's recommendations and using oligodT and random hexamers as primers. Quantitative real-time PCR (qPCR) was performed in triplicates in a QuantStudio 6flex (Thermo Fisher) using PowerUp SYBR green master mix (Thermo Fisher, A25778). All quantifications of mRNA levels were first normalized to an endogenous housekeeping control (Tbp), and then mRNA relative quantities to a reference sample (Brain, N2a undifferentiated) were calculated by the $2^{-\Delta\Delta Ct}$ method. The primers used for qPCR are the following: 5' TGATTCCATTAAAGGTCGTCTAAACT 3' (Fw) and 5' GAAACAATGAAGACTGACCTCTCCTT 3' (Rv) for Mm Cpeb4 isoform containing exons 3 and 4; 5' TGATTCCATTAAAGCAAGGACTTATG 3' (Fw) and 5' GCTGTGATCTCATCTTCATCAATATC 3' (Rv) for Mm Cpeb4 isoform lacking exon 3; 5' TGATTCCATTAAAGGTCGTCTAAACT 3' (Fw) and 5' GGAAACAATGAAGACTGACCATTAAT 3' (Rv) for Mm Cpeb4 isoform lacking exon 4; 5' ATTCCATTAAAGGTCAGTCTTCATTG 3' (Fw) and 5' GCTGTGATCTCATCTTCATCAATATC 3' (Rv) for Mm Cpeb4 isoform lacking exons 3 and 4; 5' GGAAAGGGACCTTCAAAGCAGT 3' (Fw) and 5' CTCTGTCCTTGGCATCGGCT 3' (Rv) for Mm Srrm4; 5' ACTTCTATGCAGGCACGGTG 3' (Fw) and 5' AGCCAGGCATTGCAGAAGTAT 3' (Rv) for Mm Rbfox1. Mm JunB and cFos primers were obtained from Gonatopoulos-Pournatzis et al., 2020[3]. The statistical analysis of Srrm4 levels was carried out using one-way ANOVA. The statistical analysis of early depolarization markers was carried out using two-way ANOVA. Differences were considered significant when the p-value was lower than 0.05 (*), 0.01 (**), 0.001 (***), or 0.0001 (****).

Real-time semi-quantitative RT-PCR

For CPEB4 splicing isoforms amplification, specific primers were used in CPEB4 exon 2 (Fw primer) and exon 5 (Rv primer), as previously described[5]. PCR products conforming to the four isoforms of CPEB4 were resolved on a 2% agarose/GelRed™ gel run at 130V for 2 h.

Western blot

Cells were lysed in ice-cold buffer containing 1% NP40, 150 mM NaCl, 50 mM Tris HCl (pH 7.5), 2 mM EDTA, 2 mM EGTA, 20 mM sodium fluoride, 2 mM PMSF, 2 mM sodium orthovanadate, 1 mM DTT and 1x EDTA-free complete protease inhibitor cocktail. Lysed samples were then sonicated at medium intensity for 5 min with Standard Bioruptor Diagenode, and total protein content was quantified using DC Protein Assay (Biorad, 5000113). 15 to 30 µg of total protein lysate were resolved on SDS-PAGE gel and transferred to a nitrocellulose membrane (Cytiva, 10600001). After 1 h of blocking at RT in 5% non-fat milk, membranes were incubated overnight at 4ºC with primary antibodies and subsequently with secondary antibodies for 2 h at RT. Specific proteins were labeled using the following primary antibodies: CPEB4 (homemade, mouse monoclonal, clone 149C), GFP (Invitrogen, rabbit polyclonal, A6455), myc (Abcam, goat polyclonal, ab9132), and -actin (Abcam, mouse monoclonal, ab20272); or streptavidin (Thermofisher, S911). Membranes were then incubated for 3 min with Amersham ECL TM Western Blotting detection reagents (Sigma, GERPN2106) or for 5 min with Clarity Western ECL Substrate (Biorad, 1705061).

nCPEB4 to nCPEB4Δ4 co-localization experiments in N2a cells

mCherry red signal was acquired with a DPSS 561 excitation laser (9%) and a HyD2 detector set to 570-650 nm with a gain of 33%. DAPI and GFP signals were acquired using the settings specified in the "*nCPEB4-GFP distribution in N2a cells*" section. For measuring the extent of co-localization between the two channels, the ImageJ JaCoP plugin was used to obtain Pearson's correlation coefficient and Mander's overlap coefficient per cell.

Immunohistochemistry

Mouse embryos at embryonic day 13.5 were fixed in 10% neutral-buffered formalin solution and embedded in paraffin. Rabbit polyclonal primary antibody anti-CPEB4 (Abcam, ab83009) was used at 1:1000 dilution. Embryo sections were counterstained with hematoxylin.

Xenopus laevis oocyte preparation

Stage VI oocytes were obtained from full-grown *Xenopus laevis* females as previously described[48]. Briefly, ovaries were treated with collagenase (2 mg/mL; StemCell Technologies) and incubated in Modified Bath Saline (MBS) 1x media with 0.7 mM $CaCl_2$. Animal handling and all experimental protocols were approved by the Animal Ethics Committee at the Barcelona Science Park (PCB) and by the Government of Catalonia.

Xenopus laevis BioID

BioID was performed as previously described[7]. Briefly, 150 stage VI *X. laevis* oocytes were microinjected with 50.6 nL of 50 ng/µL *in vitro* transcribed and polyadenylated RNAs corresponding to Myc-BirA-xCPEB4 variants. Oocytes were then incubated in 20 µM Biotin (Merck) at 18ºC for 40 h. Oocytes were lysed in cold lysis buffer and centrifuged twice at 16,000 *g* at 4 °C for 15 min. Cold BioID lysis buffer was added to cleared extract, and the resulting mixture was subjected to clearing with PD MiniTrap G-25 columns (GE Healthcare). 1.6% Triton-X100 and 0.04% SDS were added, and extracts were incubated with MyOne Dynabeads Streptavidin C1 (Invitrogen). The beads were then washed with a subsequent sequence of wash buffers. The beads were resuspended in 3 M urea, 50 mM NH_4HCO_3 pH 8.0, 5 mM DTT for 1 h at RT with orbital shaking and subsequently incubated in 10 mM iodoacetamide for 30 min at RT, and then DTT was added. Proteins were on-bead digested with trypsin (Promega) at 37 °C for 16 h with orbital shaking. Digestion was stopped by the addition of 1% formic acid. Mass spectrometry analysis of biotinylated proteins in xCPEB4 and xCPEB4+Ex4 was carried out at the Mass Spectrometry (MS) Facility at IRB Barcelona as previously described[7]. Briefly, samples were analyzed using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific). The MS/MS spectra obtained were searched against the UniProt (Xenopodinae, release 2017_02) and contaminants databases, and proteins of interest sequences using Proteome Discoverer v2.1.0.81.

Competition experiments

Competition experiments were performed as previously described[10] using 23 nL of 500 ng/µL in vitro transcribed and polyadenylated RNAs encoding for HA-tagged xCPEB1 and xCPEB4 RRMs and variants. Not injected was considered as 0% competition whereas HA-xCPEB1 RRM was considered as 100% competition.

Plasmids for protein expression in E. coli

An insert codifying for the nCPEB4-NTD protein sequence (UniProt ID: Q17RY0-2, residues 1–448) was ordered in GenScript subcloned in a pET-30a(+) vector. The His_6- and

S- tags from the plasmid were removed by PCR using the NEB Q5 Site-directed mutagenesis kit, with oligonucleotides purchased from Sigma-Aldrich. The His to Ser mutants were ordered in GenScript subcloned in a pET-30a(+) vector in the NdeI and XhoI restriction enzymes positions. The nCPEB4Δ4, ΔHClust, and Δ4_ΔHClust mutants were generated by PCR mutagenesis on the nCPEB4-NTD plasmid using the NEB Q5 Site-directed mutagenesis kit, with oligonucleotides purchased from Sigma-Aldrich.

Protein expression and purification for in vitro experiments

E. coli B834 cells were transformed with the pET-30a(+) plasmids. For non-isotopically labeled protein, the cells were grown in LB medium at 37 ºC until $OD_{600\ nm}$ = 0.6, and then the cultures were induced with 1 mM IPTG for 3 h at 37 ºC. For single (^{15}N) or double (^{15}N,^{13}C) isotopically labeled protein, the cells were grown in LB medium until $OD_{600\ nm}$ = 0.6 and then transferred into M9 medium[49] (3 L LB for 1 L M9) containing $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-glucose, respectively, and then the cultures were induced with 1 mM IPTG overnight at 37 ºC. The cultures were then centrifuged for 30 min at 4,000 rpm, and the cells were resuspended with the *lysis* buffer (50 mM Tris-HCl, 1 mM DTT, 100 mM NaCl, 0.05% TritonX-100, at pH 8.0, and supplemented with 500 µL of PIC and 500 µL of 100 mM PMSF).

The cells were lysed by sonication and centrifuged for 30 min at 20,000 rpm. The pellet was washed first with the *wash-1* buffer (20 mM Tris-HCl, 1 mM DTT, 1 M NaCl, 0.05% TritonX-100, at pH 8.0, and supplemented with 500 µL of PIC, 500 µL of 100 mM PMSF, and 50 µL of 5 mg/mL DNAse) and then with the *wash-2* buffer (20 mM Tris-HCl, 1 mM DTT, 0.1 M L-Arginine, at pH 8.0). The pellet was resuspended with the *nickel-A* buffer (25 mM Tris-HCl, 1 mM DTT, 50 mM NaCl, 8 M urea, 20 mM imidazole, at pH 8.0) and centrifuged for 30 min at 20,000 rpm. The supernatant was injected at RT into a nickel affinity column and eluted with a gradient from 0 to 100% of the *nickel-B* buffer (25 mM Tris-HCl, 1 mM DTT, 50 mM NaCl, 8 M urea, 500 mM imidazole, at pH 8.0). The fractions with protein were pooled, and 1 mM EDTA was added. The sample was injected into a size exclusion Superdex 200 16/600 (GE Healthcare) column, running at 4 ºC in the *size exclusion* buffer (25 mM Tris-HCl, 1 mM DTT, 50 mM NaCl, 2 M urea, at pH 8.0). The fractions with protein were pooled and concentrated to *ca* 150 µM. The sample was dialyzed against the *final* buffer (20 mM sodium phosphate, 1 mM TCEP, 0.05% NaN_3, at pH 8.0), fast frozen in liquid nitrogen, and stored at -80 ºC.

Peptide for in vitro experiments

The E4$(GS)_3$E4 synthetic peptide with amidated or Cy3 modified C-terminus and acetylated N-terminus was obtained as lyophilized powder with >95% purity from GenScript. The

peptides were dissolved in 6 M guanidine thiocyanate and incubated with agitation overnight at 25 ºC. The samples were then centrifuged at 15,000 rpm for 10 min. The supernatant was extensively dialyzed against the *final* buffer (20 mM sodium phosphate, 1 mM TCEP, 0.05% NaN_3, at pH 8.0)[50]. The peptide samples were then manipulated in the same way as the protein samples, as detailed below.

Sample preparation for in vitro experiments

All samples were prepared on ice as follows. First, a buffer stock solution consisting of 20 mM sodium phosphate buffer with 1 mM TCEP and 0.05% NaN_3 was pH adjusted to 8.0 (unless otherwise indicated) and filtered using 0.22 µm sterile filters (*Buffer Stock*). A 1 M NaCl solution in the same buffer was also pH adjusted to 8.0 (unless otherwise indicated) and filtered (*Salt Stock*). The protein samples were then thawed from -80 ºC on ice, pH adjusted to 8.0 (unless otherwise indicated), and centrifuged for 5 min at 15,000 rpm at 4 ºC. The supernatant (*Protein Stock*) was transferred to a new Eppendorf tube, and the protein concentration was determined by measuring its absorbance at 280 nm. The samples were prepared by mixing the right amounts of *Buffer Stock*, *Protein Stock*, and *Salt Stock*, as well as other indicated additives in the experiments, to reach the desired final protein and NaCl concentrations. All samples contained 30 µM protein and 100 mM NaCl concentrations unless otherwise indicated.

Apparent absorbance measurement as a function of temperature

The absorbance of the samples was measured at 350 nm ($A_{350\ nm}$) using 1 cm pathlength cuvettes and a Cary100 ultraviolet–visible spectrophotometer equipped with a multicell thermoelectric temperature controller. The temperature was increased progressively at a ramp rate of 1 ºC·min^{-1}. The cloud temperatures (Tc) were determined as the maximum of the first-order derivatives of the curves, and the absorbance increase (ΔA) represents the difference between the maximum and the minimum absorbance values of the samples during the temperature ramp.

For the experiment to quantify the reversibility of phase separation, a 20 µM protein with 100 mM NaCl sample was prepared on ice. It was then split into four Eppendorf tubes and a temperature ramp was carried out with the first one after centrifugation for 2 min at 5 ºC and 15,000 rpm. Once the Tc and ΔA for phase separation had been determined, the other three samples were heated 10 ºC above the Tc for 2.5 min and then cooled to 10 ºC below the Tc for 5 more min. This procedure was repeated 1, 2, or 3 times for each sample. Next, the samples were centrifuged for 2 min at 5 ºC and 15,000 rpm, and a temperature ramp was carried out to determine their respective Tc and ΔA (Supplementary Fig. 4f).

Microscopy in vitro

For microscopy imaging, 1.5 µL of sample was deposited in a sealed chamber comprising a slide and a coverslip sandwiching double-sided tape (3M 300 LSE high-temperature double-sided tape of 0.17 mm thickness). The coverslips used had been previously coated with PEG-silane following a published protocol[51]. The imaging was always performed on the surface of the coverslip, where the droplets had sedimented.

The DIC microscopy images were taken using an automated inverted Olympus IX81 microscope with a 60x/1.42 oil Plan APo N or a 60x/1.20 water UPlan SAPo objective using the Xcellence rt 1.2 software.

For fluorescence microscopy experiments, purified nCPEB4-NTD was labeled with DyLight 488 dye (DL488, Thermo Fisher Scientific) or Alexa Fluor 647 dye (AF647, Thermo Fisher Scientific). The labeling, as well as the calculation of the labeling percentage and the determination of the protein concentration, was performed following the provider's instructions. The final samples contained 0.5 µM of labeled protein and/or peptide out of the total indicated concentrations.

Fluorescence recovery after photobleaching (FRAP) experiments of a 20 µM protein sample with 150 mM NaCl were recorded at 37 ºC by using a Zeiss LSM780 confocal microscope system with a Plan ApoChromat 63x/1.4 oil objective. Droplets of similar size were selected and the bleached region was 30% of their diameter. The intensity values were monitored for different regions of interest (ROI): ROI1 (bleached area), ROI2 (entire droplet), and ROI3 (background signal). The data were fitted using EasyFrap software[47] to extract the kinetic parameters of the experiment (recovery half-time and mobile fraction).

Super-resolution microscopy images of the multimers (10 µM nCPEB4-NTD with 100 mM NaCl) and their time evolution were taken at 25 ºC in a Zeiss Elyra PS1 LSM 880 confocal microscope with the Fast Airyscan mode using an alpha Plan ApoChromat 100x/1.46 oil objective. The pixel size was kept constant at 40 nm.

Fluorescence microscopy images of the droplets and aggregates were taken at 37 ºC in a Zeiss Elyra PS1 LSM 880 confocal microscope with Airyscan detector using a Plan ApoChromat 63x/1.4 oil objective. Samples in Fig. 4i,j, Supplementary Fig. 4i, Fig. 5b,d,e, and Supplementary Fig. 5a contain 200 mM NaCl. The quantification of the aggregation process was done by image analysis using FIJI (ImageJ). The regions with fluorescence signal not stemming from the droplets were selected. The percentage of the area of the field of view occupied by this selection corresponds to the aggregation (Agg) value of the sample. The

partitioning of the proteins or peptides in the droplets was calculated by image analysis using FIJI (ImageJ). The partitioning for each droplet was calculated by dividing the mean intensity of the droplet by the mean intensity of a ring of 1 µm thickness around the droplet. The statistical analysis of the partitioning values was carried out with a t-test. The differences were considered significant when the p-value was lower than 0.05 (*), 0.01 (**), 0.001 (***), or 0.0001 (****).

Saturation concentration measurements

Saturation concentration measurements of nCPEB4-NTD and the His to Ser mutants were carried out incubating the samples at 40 ºC for 5 min, followed by centrifugation at 5,000 rpm for 1.5 min at 40 ºC. The concentration of protein in the supernatant (c_{sat}) was determined by absorbance measurement at 280 nm.

Nuclear magnetic resonance spectroscopy

The samples were prepared as indicated in the "*Sample preparation for in vitro experiments*" section, using isotopically labeled protein (^{15}N- or ^{15}N,^{13}C-labeled). The prepared final samples were again pH adjusted to the desired value immediately before measurement. All the measurements were acquired at 5 ºC using 3 mm NMR tubes with a sample volume of 200 µL.

All NMR experiments, except the diffusion measurements, were carried out on a Bruker Avance NEO 800 MHz spectrometer equipped with a TCI cryoprobe. All NMR samples contained 100 µM protein concentration (unless otherwise indicated) in a 20 mM sodium phosphate buffer with 1 mM TCEP, 0.05% NaN_3, 7% D_2O, and 2.5 µM DSS for referencing, at pH 8.0 (unless otherwise indicated). Samples with denaturant agent contained the indicated concentrations of d_4-urea.

A ^{15}N,^{13}C-labeled nCPEB4-NTD sample at 280 µM with 4 M urea at pH 7.0 was used for backbone resonance assignment. A series of non-linear sampled 3D triple resonance experiments were recorded, including the BEST-TROSY version[52] of $^{1}H_N$-detected HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, and (H)N(CA)NH. Also, additional ^{1}Hα-detected HA(CA)CON and (HCA)CON(CA)H experiments[53] were measured. Backbone resonance assignments were performed using ccpnmr[54]. The NMR assignment of nCPEB4-NTD will be available on BMRB (ID 51875).

pH titrations from 7.0 to 8.0 were carried out to transfer NH assignments to the final experimental conditions. Standard 2D ^{1}H,^{15}N-HSQC, or BEST-TROSY experiments were measured at 7.0 ≤ pH ≤ 7.25. 2D ^{1}H,^{15}N-CP-HISQC[55] experiments were used at pH ≥ 7.5 to

reduce the effects of chemical exchange with water. For the urea titrations from 0 to 4 M at pH 8.0, $^{1}H,^{15}N$-CP-HISQC experiments were measured. In the $^{1}H,^{15}N$-CP-HISQC for the detection of Arg side chain resonances the ^{15}N carrier was placed at 85 ppm and ^{13}C pulses for decoupling were centered at the chemical shift of $^{13}C_{\delta}$ (42 ppm) and $^{13}C_{\zeta}$ (158 ppm) Arg side chains.

Standard 2D $^{1}H,^{13}C$-HSQC experiments of 500 µM $^{15}N,^{13}C$-labeled nCPEB4-NTD were measured in the absence and presence of 4 M urea to monitor specific amino acid side chains easily identifiable by their typical ^{1}H and ^{13}C random coil chemical shifts.

Non-uniform sampled experiments were processed using qMDD[56]. 2D ^{15}N-correlations ($^{1}H,^{15}N$-HSQC, $^{1}H,^{15}N$-CP-HISQC, BEST-TROSY) and 2D $^{1}H,^{13}C$-HSQC were processed with NMRPipe[57] and Topspin 4.0.8 (Bruker), respectively.

^{15}N-edited X-STE diffusion experiments[58] of 100 µM ^{15}N-labeled nCPEB4-NTD in the absence and presence of 2 M urea were performed on a Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe. An encoding/decoding gradient length of 4.8 ms and a diffusion delay of 320 ms were used. The hydrodynamic diameter of nCPEB4-NTD was estimated using dioxane as a reference molecule. Diffusion measurements under identical experimental conditions were carried out for dioxane, using in this case the PG-SLED sequence[59]. A gradient time (δ) of 1.6 ms and a diffusion time (Δ) of 70 ms were used. Diffusion coefficients were obtained by fitting the data to a mono-exponential equation using the MestreNova software.

Dynamic light scattering

DLS measurements were taken with a Zetasizer Nano-S instrument (Malvern) equipped with a He-Ne of 633 nm wavelength laser. Immediately before the measurement, the prepared samples were centrifuged for 5 min at 15,000 rpm at 4 ºC and only the supernatant was subjected to measurement. Three measurements were performed for each sample, each of the measurements consisting of 10 runs of 10 s each. The experiments were carried out at 5 ºC, unless otherwise indicated. The deconvoluted hydrodynamic diameters arise from the mean of the peak of the intensity deconvolution of the data. The multimers time evolution experiment was taken on a 10 µM nCPEB4-NTD sample with 0 mM NaCl.

Size exclusion chromatography coupled to multi-angle light scattering

SEC-MALS experiments were performed by loading a 160 µM nCPEB4-NTD with 0 mM NaCl sample into a Superose 6 increase 10/300 GL column (GE Healthcare) mounted on a Shimadzu Prominence Modular HPLC with an SPD-20 UV detector (Shimadzu) coupled to a

Dawn Heleos-II multi-angle light scattering detector (18 angles, 658 nm laser beam) with an Optilab T-rEX refractometer (Wyatt Technology). The SEC-UV/MALS/RI system was equilibrated at 25 ºC with a 20 mM sodium phosphate buffer with 1 mM TCEP and 0.05% NaN_3 at pH 8.0. Data acquisition and processing were performed using the Astra 6.1 software (Wyatt Technology).

Liquid-phase transmission electron microscopy

TEM experiments were performed on a 50 µM nCPEB4Δ4-NTD sample with 0 mM NaCl. The sample was first imaged in solid-state TEM using the same set-up as described below for liquid-phase TEM (LPEM). Pre-screening samples in solid-state TEM before the LPEM imaging procedure is routinely done to pre-screen the structures that will be imaged in liquid later. The sample with no stain was deposited onto 400 mesn Cu grids, which were plasma discharged for 45 seconds to render them hydrophilic and allow optimal sample wettability.

LPEM imaging was performed using a JEOL JEM-2200FS TEM microscope. The system was equipped with a field emission gun operating at 200 kV and an in-column Omega filter. The images were acquired with a direct detection device, the in-situ K2-IS camera (Gatan). The *Ocean* liquid holder, from DENSsolutions (Netherlands), was used to image the specimens' structure and dynamics. The liquid samples were encased into two silicon nitride (Si_XN_y) chips. These chips had a 50 nm-thick Si_XN_y electron-transparent window of dimensions 10 µm x 200 µm. One of these chips had a 200 nm spacer that acts as a pillar and defines the liquid cell thickness, i.e Z height, and hence the liquid thickness in the experiments. The chips were cleaned in HPLC-graded acetone followed by isopropanol for 5 min each to remove their protective layer made of poly(methyl methacrylate) (PMMA). Afterwards, the chips were plasma-cleaned for 13 min to increase their hydrophilicity. 1.5 µL of non-stained sample was deposited on the previously prepared 200 nm-spacer chip. The drop casted sample was enclosed by the spacer-less chip, thus sealing the liquid chamber. 300 µL of the sample solution was flushed in the liquid holder at 20 µL/min with a syringe pump to ensure the liquid cell inlet and outlet pipes were filled out with the solution. We waited five minutes after collecting the sample solution from the outlet tube to minimize the convection effects from the flowing process. The liquid holder was introduced in the microscope where imaging was performed in TEM mode and static conditions, i.e. not in flow. In order to limit the electron beam dose on the specimen, images were collected at the minimum spot size (#5) with a small condenser lens aperture (CLA #3). For our investigations, dose fractionation videos were recorded in counted mode at 20 frames per second with the K2. Every image was recorded in the format of 4-byte grayscale and required the full size of the detector.

The images and videos recorded were corrupted by noise, which significantly reduces the quality of the images, complicating any further analysis. Therefore, the Noise2Void (N2V) machine learning-based approach has been adopted to overcome this problem[60]. Unlike the conventional machine learning-based approach, the N2V reduces the noise of an image without any need for the corresponding noiseless image. This requirement makes the N2V perfect to process LPEM images, as noiseless images in liquid TEM are impossible to record. However, the N2V requires an extensive dataset (also known as a training set) in order to fulfill its task, a common requirement to machine learning-based approaches. Conventionally, the training set has to contain thousands of hundreds of images recorded with the same imaging settings to produce high-quality estimation of the noise distribution. Unfortunately, in our case, only single videos (image sequence) recorded at different imaging conditions were available. Therefore, the training set was created by sampling the image sequence every two frames in order not to bias the training process of the N2V. The remaining half of the image sequence was processed and used to derive the presented results.

To train the N2V, 3050 frames were selected, extracting 128 different non-overlapping squared patches (i.e. portions of the pixels of the image) of 64x64 pixels from each of them. Moreover, the N2V was iterated for 100 epochs, a trade-off value between performances and processing time. The training was performed by extracting 128 random patches (64x64 pixels) from each training image. These values produced the best results in a short processing time.

Molecular simulations

Molecular dynamics simulations were performed using the single-bead-per-residue model CALVADOS 2[24,25] implemented in openMM v7.5[61]. All simulations were conducted in the *NVT* ensemble at 20 ºC using a Langevin integrator with a time step of 10 fs and friction coefficient of 0.01 ps^{-1}. Configurations were saved every 5 ns. In our simulations, pH is modelled through its effect on the charge on the histidine residues, which was estimated to be 0.0099 at pH 8 and 0.5 at pH 6 using the Henderson–Hasselbalch equation and assuming a pKa value of 6[62].

Direct-coexistence simulations were performed with 100 chains in a cuboidal box of side lengths $[L_x,L_y,L_z]$ = [25, 25, 300] nm for at least 20 µs. The systems readily formed a protein-rich slab spanning the periodic boundaries along the x- and y-coordinates. The initial 1 µs of the simulation trajectories were discarded and time-averaged concentration profiles along the z-axis were calculated after centering the condensates in the box as previously described[24]. To model protein multimers, 400 chains were initially placed at random positions in a cubic box of side length 188 nm and simulations were run in two replicas for at least 16 µs. The formation and dissolution of protein multimers was monitored via a cluster analysis implemented in OVITO[63]. Proteins were clustered based on the distances between their

centers of mass, using a cutoff of 1.5 times the average radius of gyration of the protein. Contact maps were calculated between a chain in the middle of the condensate or multimer and the surrounding chains. Contacts were quantified via the switching function $c(r) = 0.5 - 0.5 \times tanh[(r - \sigma)/r_0]$, where *r* is the intermolecular distance between two residues, σ is the arithmetic average of the diameters of the two residues, and $r_0 = 0.3$ nm.

To match the conditions of the experiments with which the simulation results are compared, direct-coexistence (Fig. 2 and Supplementary Fig. 2) and multimer (Fig. 3) simulations were performed at ionic strengths of 150 mM and 60 mM, respectively.

Data representation

All box-plot elements in the figures are defined as follows: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

Data availability

The NMR assignment will be available on BMRB (ID 51875). The simulation data and code generated in this study are available via https://github.com/KULL-Centre/_2023_Garcia-Cabau_CPEB4 and https://doi.org/10.5281/zenodo.7727848. The rest of the data generated in this study will be included in this article (and its supplementary information files).

Code availability

The simulations were performed with CALVADOS (v2; available at https://github.com/KULL-Centre/CALVADOS) using openMM (v7.5, https://openmm.org/) an open source toolkit for molecular simulations. The code for repeating the simulations analyses is available at https://github.com/KULL-Centre/_2023_Garcia-Cabau_CPEB4 and at https://doi.org/10.5281/zenodo.7727848.

Methods references

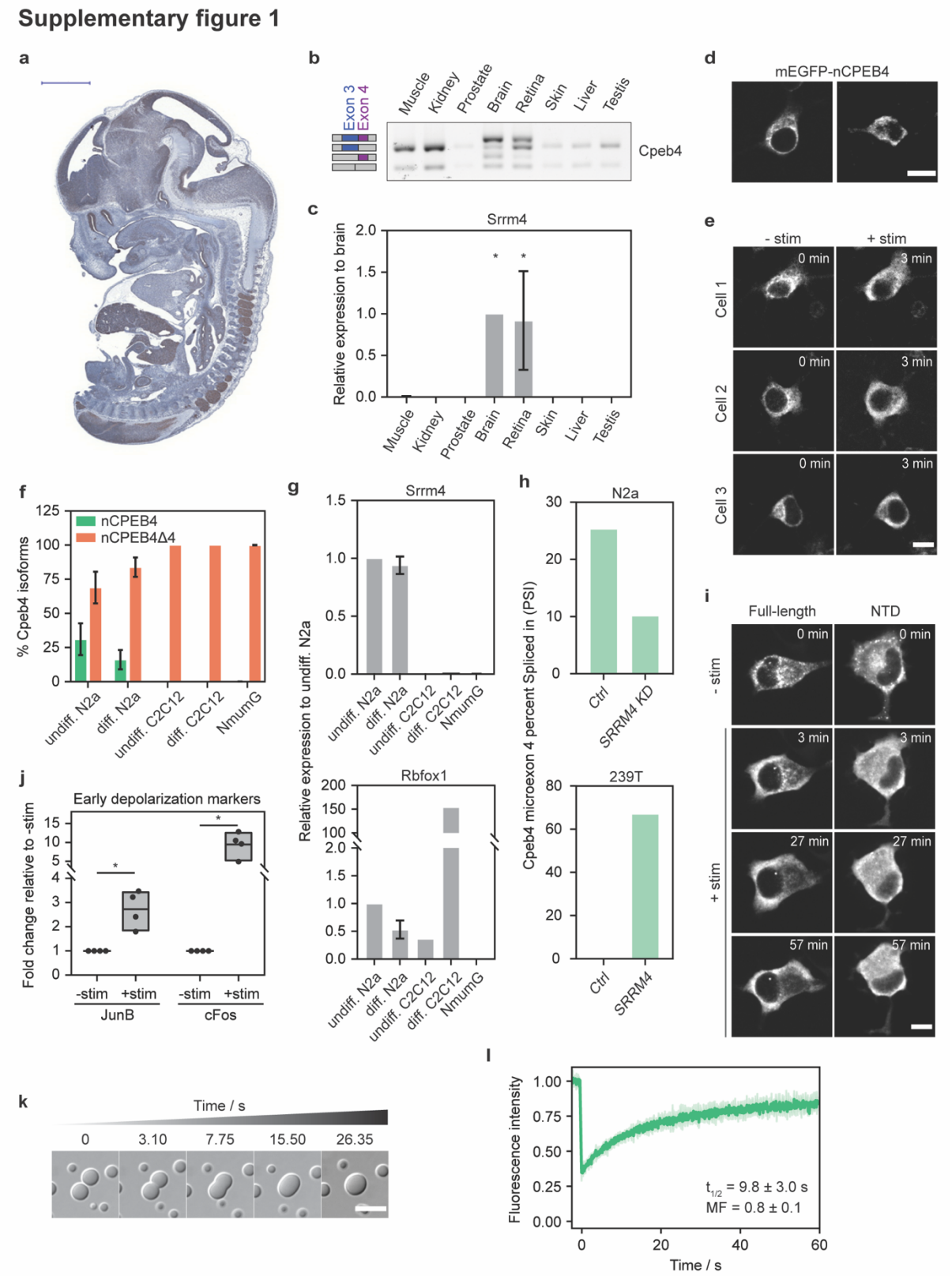
45. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
46. Belloc, E. & Méndez, R. A deadenylation negative feedback mechanism governs meiotic metaphase arrest. *Nature* **452**, 1017–1021 (2008).

47. Rapsomaniki, M. A. *et al.* easyFRAP: an interactive, easy-to-use tool for qualitative and quantitative analysis of FRAP data. *Bioinformatics* **28**, 1800–1801 (2012).
48. de Moor, C. H. & Richter, J. D. Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. *EMBO J.* **18**, 2294–2303 (1999).
49. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).
50. Escobedo, A., Chiesa, G. & Salvatella, X. Recombinant Production of Monomeric Isotope-Enriched Aggregation-Prone Peptides: Polyglutamine Tracts and Beyond. in *Intrinsically Disordered Proteins: Methods and Protocols* (eds. Kragelund, B. B. & Skriver, K.) 211–231 (Springer US, 2020).
51. Alberti, S. *et al.* A User's Guide for Phase Separation Assays with Purified Proteins. *Journal of Molecular Biology* vol. 430 4806–4820 Preprint at https://doi.org/10.1016/j.jmb.2018.06.038 (2018).
52. Solyom, Z. *et al.* BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR* **55**, 311–321 (2013).
53. Mäntylahti, S., Hellman, M. & Permi, P. Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. *J. Biomol. NMR* **49**, 99–109 (2011).
54. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
55. Yuwen, T. & Skrynnikov, N. R. CP-HISQC: a better version of HSQC experiment for intrinsically disordered proteins under physiological conditions. *J. Biomol. NMR* **58**, 175–192 (2014).
56. Mayzel, M., Rosenlöw, J., Isaksson, L. & Orekhov, V. Y. Time-resolved multidimensional NMR with non-uniform sampling. *J. Biomol. NMR* **58**, 129–139 (2014).
57. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
58. Ferrage, F., Zoonens, M., Warschawski, D. E., Popot, J.-L. & Bodenhausen, G. Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J. Am. Chem. Soc.* **125**, 2541–2545 (2003).
59. Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431 (1999).
60. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - learning denoising from single noisy images. *arXiv [cs.CV]* (2018).
61. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
62. Nagai, H., Kuwabara, K. & Carta, G. Temperature Dependence of the Dissociation

Constants of Several Amino Acids. *J. Chem. Eng. Data* **53**, 619–627 (2008).

63. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **18**, 015012 (2009).
64. Sickmeier, M. *et al.* DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **35**, D786–93 (2007).
65. Vacic, V., Uversky, V. N., Dunker, A. K. & Lonardi, S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**, 211 (2007).

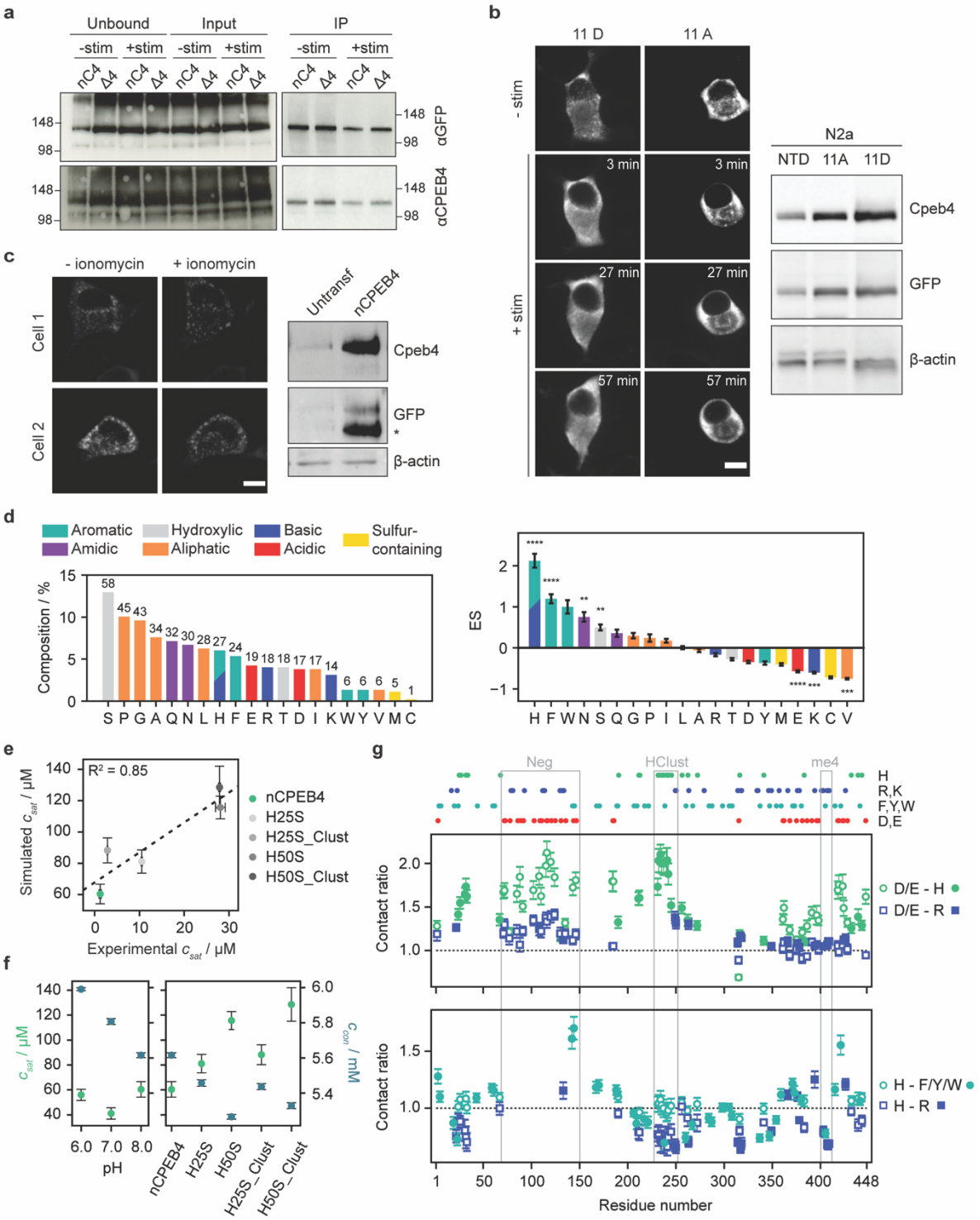
Supplementary figures and legends



Supplementary Figure 1 | a) Representative CPEB4 immunohistochemistry in a mouse embryo (E13.5). The scale bar represents 1000 µm. **b)** Representative RT-PCR assay

monitoring Cpeb4 microexon 4 splicing in different mouse organs. Amplified Cpeb4 splice variants are represented next to the shown gel (alternative exon 3 and microexon 4 are highlighted in blue and purple, respectively). **c)** qRT-PCR assays monitoring Srrm4 relative transcript levels in different mouse organs (mean ± SD, n ≥ 2). **d)** Fluorescence microscopy images of live *ex vivo* striatal neurons from mEGFP-CPEB4 mice. The scale bar represents 10 µm. **e)** Fluorescence microscopy images from *ex vivo* live-cell time-lapse movies of striatal neurons from mEGFP-CPEB4 knock-in mice. Time series show pre (-stim, 0 min) and post (+stim, 3 min) stimulation foci disassembly. The scale bar represents 10 µm. **f)** qRT-PCR assays monitoring the percentage of Cpeb4 isoforms harboring exon 4 (nCPEB4) or not (nCPEB4Δ4) in different cell lines (mean ± SD, n ≥ 1). Undiff.: undifferentiated, diff.: differentiated. **g)** Top panel: qRT-PCR assays monitoring Srrm4 transcript levels in different cell lines relative to undifferentiated (undiff.) N2a. Bottom panel: qRT-PCR assays monitoring Rbfox1 transcript levels in different cell lines relative to undiff. N2a (mean ± SD, n ≥ 1). **h)** Top panel: percentage spliced in (PSI) of Cpeb4 exon4 in N2a in either Control (Ctrl) or SRRM4 knock-down (SRRM4 KD) cells. Bottom panel: PSI of Cpeb4 exon 4 in 239T cells in either Control (Ctrl), or SRRM4 over-expressed (SRRM4) cells. Data from B. Blencowe[17]. **i)** Fluorescence microscopy images from full time-lapse movies of differentiated N2a cells overexpressing either full-length (FL) nCPEB4-GFP or its N-terminal domain (NTD) in *Fig. 1h*. Time series show pre (-stim, 0 min) and post (+stim, 3, 27, and 57 min) stimulation foci disassembly. The scale bar represents 10 µm. **j)** Fold change increase in the transcript levels of early depolarization markers (JunB, cFos) in stimulated cells (+stim) relative to unstimulated (-stim). Dots represent independent experiments (n = 4). **k)** DIC microscopy images showing a fusion event of nCPEB4-NTD droplets *in vitro*. The scale bar represents 10 µm. **l)** Fluorescence recovery after photobleaching experiment of nCPEB4-NTD *in vitro* (mean ± SD, n = 3).

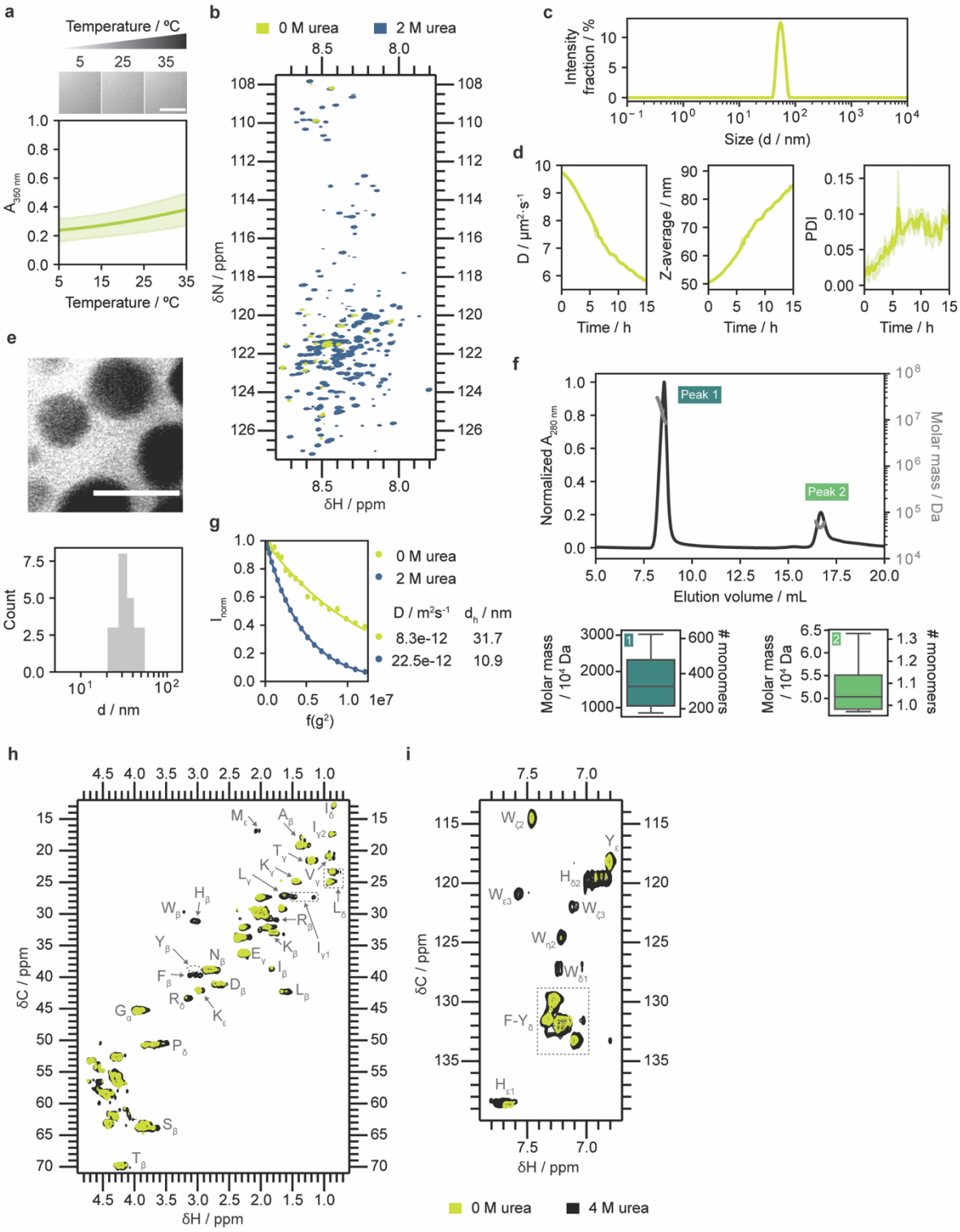
Supplementary figure 2



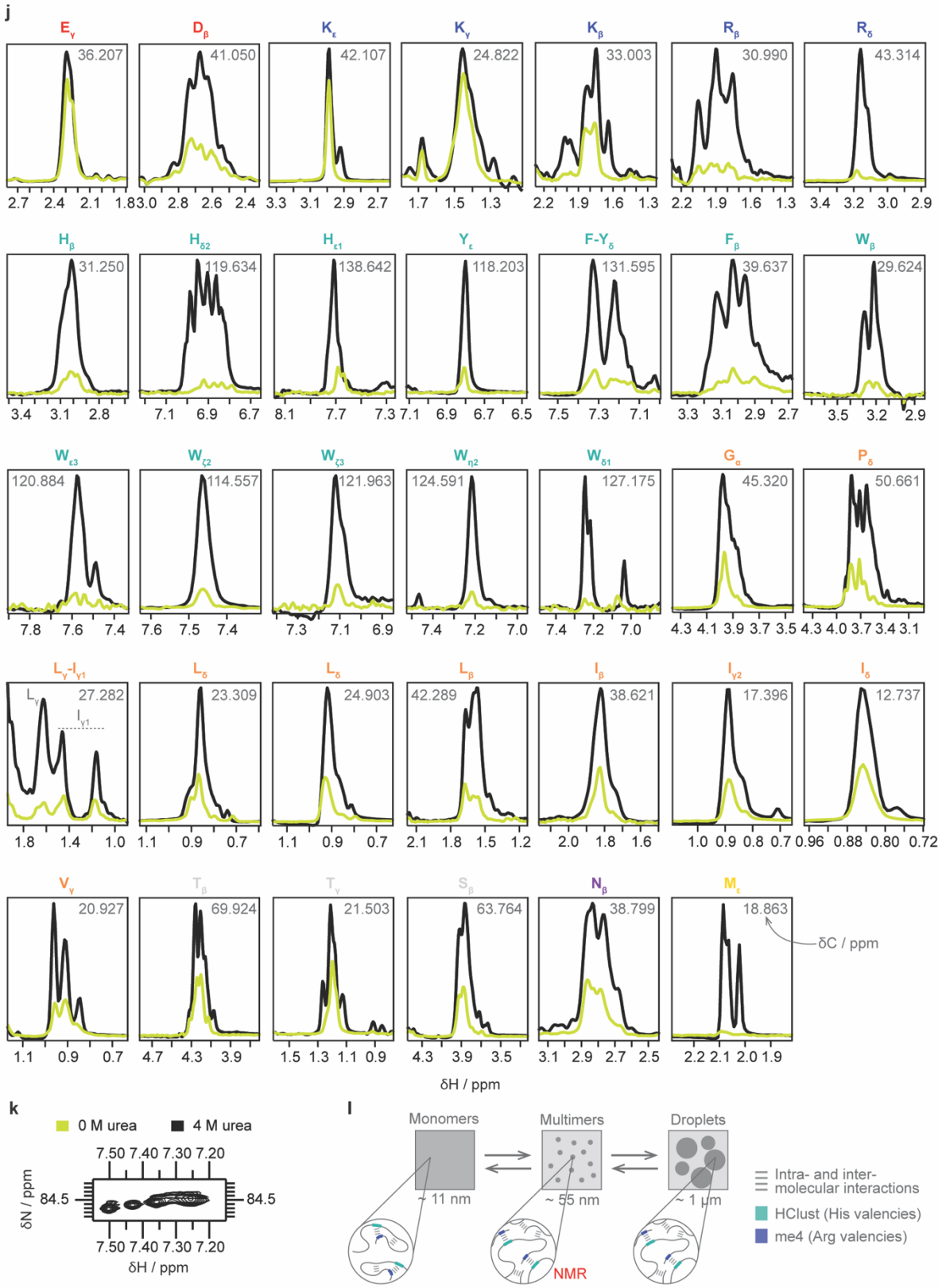
Supplementary Figure 2 | a) Western blot of nCPEB4 immunoprecipitation for mass spectrometry post-translational modification analyses. **b)** Left panel: Representative fluorescence microscopy images from time-lapse movies of N2a cells overexpressing the NTD of nCPEB4-GFP phospho-mimetic (11D) and non-phosphorylatable (11A) mutants in *Fig. 2b*. Time series show pre (-stim, 0 min) and post (+stim, 3, 27, and 57 min) stimulation foci

disassembly. The scale bar represents 10 µm. Right panel: Representative western blot of the expression of NTD of nCPEB4-GFP, 11D and 11A mutants. **c)** Left panel: Representative fluorescence microscopy images from time-lapse movies of N2a cells overexpressing nCPEB4-GFP. Time series show pre (-ionomycin) and post (+ionomycin) 1 µM ionomycin addition. The scale bar represents 10 µm. Right panel: Representative western blot of the expression of nCPEB4-GFP. Untransf: untransfected cells. **d)** Left panel: amino acid composition of nCPEB4-NTD. Right panel: enrichment score (ES) of each amino acid type in nCPEB4-NTD compared to the DisProt3.4 database[64,65]. **e)** Comparison between saturation concentrations from experiments (at 40 ºC) and molecular simulations (at 20 ºC) of nCPEB4-NTD and the His to Ser variants at pH 8. Experimental saturation concentrations are represented as mean ± SD of three measurements of the same sample. **f)** Saturation concentrations (green, left axis) and dense phase concentrations (blue, right axis) from molecular simulations of nCPEB4-NTD at pH 6, 7, and 8 (left panel) and of the His to Ser variants at pH 8 (right panel). **g)** Top panel: position of His, Arg, Lys, Phe, Tyr, Trp, Asp, and Glu residues in the nCPEB4-NTD sequence. Middle panel: ratio of contacts at pH 6 and 8 between negatively charged residues (open symbols) and histidine (green circles) or arginine residues (blue triangles). Bottom panel: ratio of contacts at pH 6 and 8 between histidine residues (open symbols) and aromatic (cyan circles) or arginine residues (blue squares). Both open and closed symbols show contacts between residues on a chain in the middle of the condensate and the surrounding chains.

Supplementary figure 3 (I)



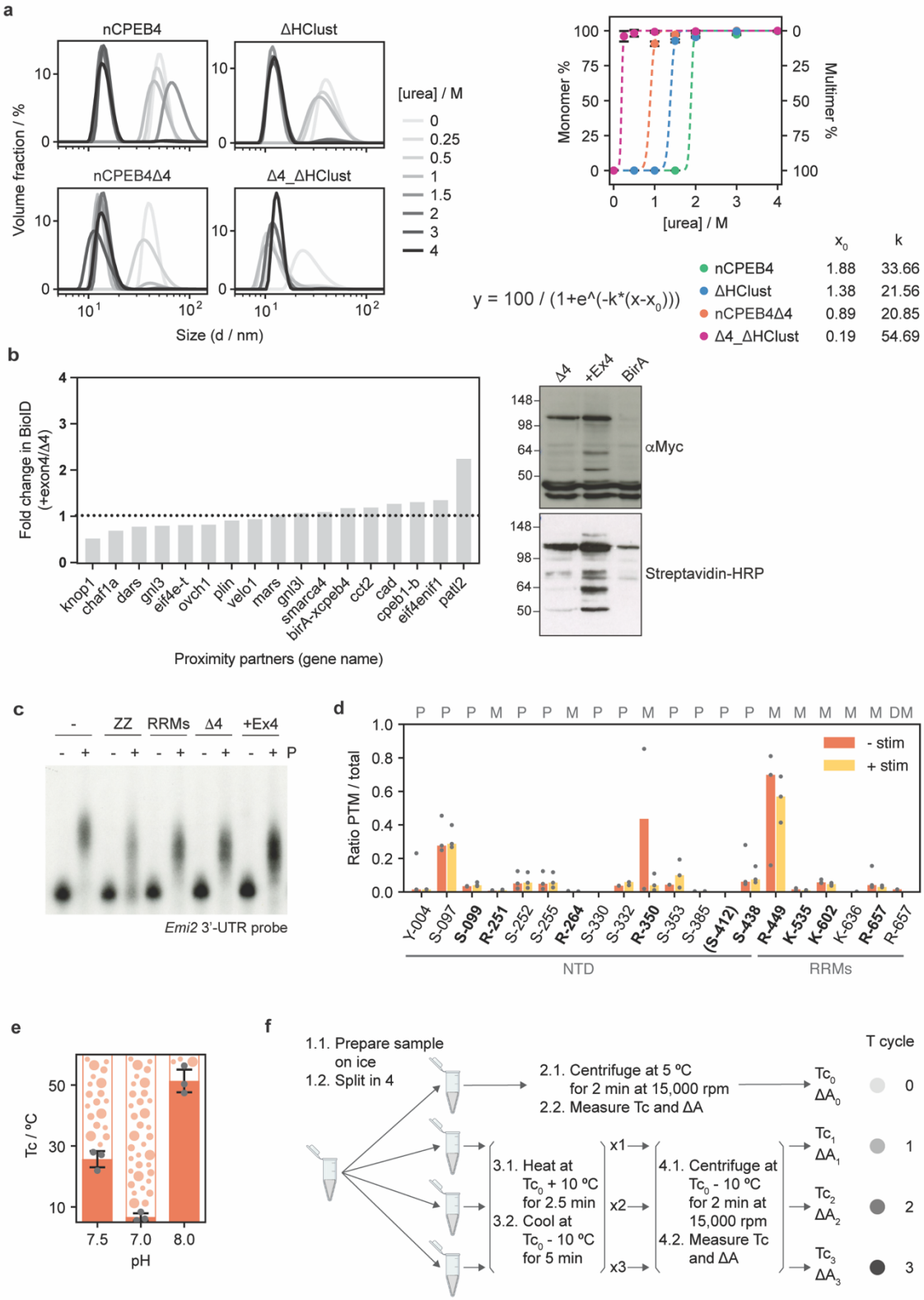
Supplementary figure 3 (II)

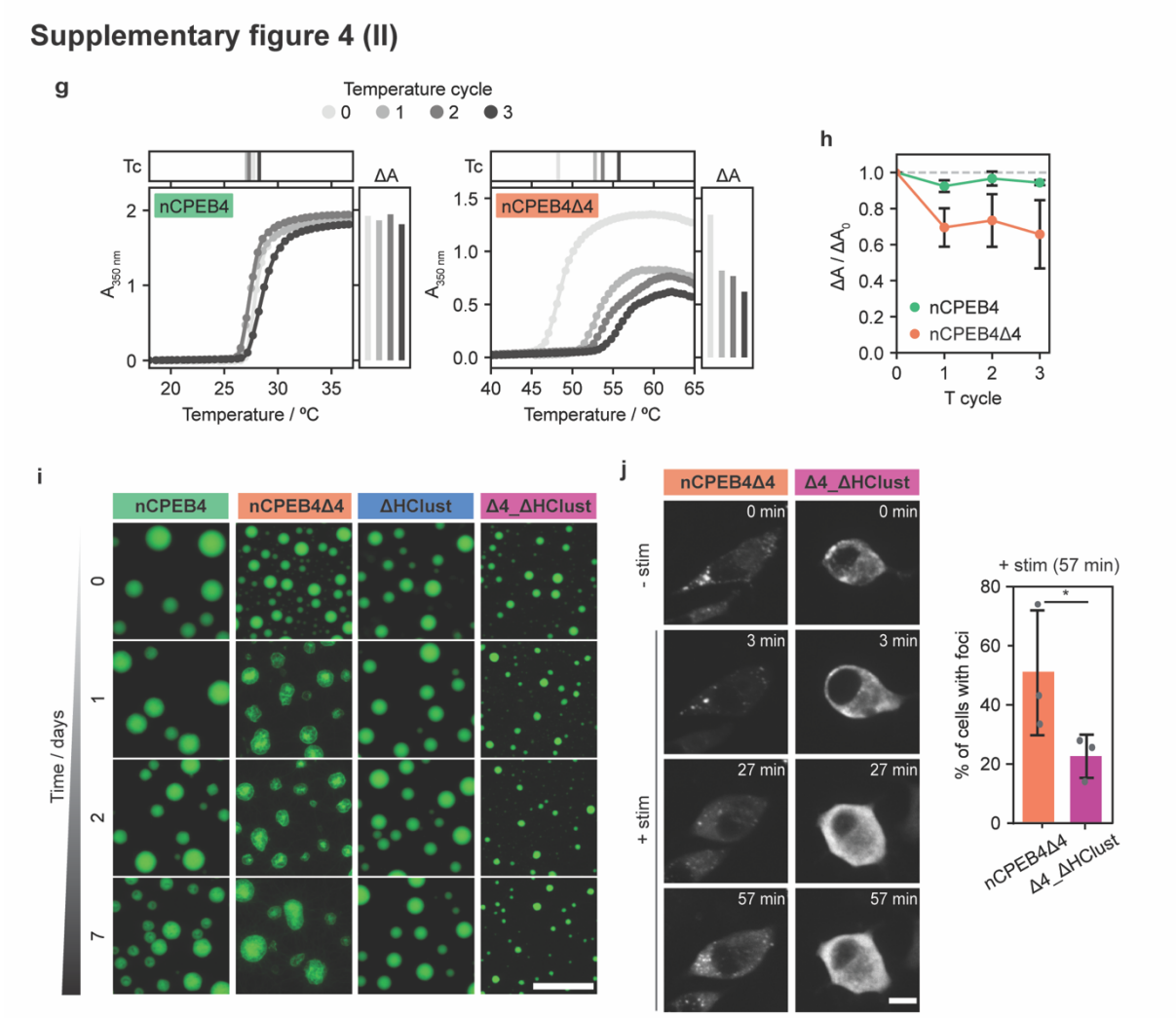


Supplementary Figure 3 | a) Apparent absorbance measurement (mean ± SD, n = 3) of the NMR sample in panel *b)* without urea at increasing temperature from 5 to 60 ºC and

corresponding DIC microscopy images at 5, 25 and 35 ºC. The scale bar represents 20 µm. **b)** $^{1}H,^{15}N$-CP-HISQC spectrum of nCPEB4-NTD in the absence (lime green) and presence (black) of 4 M urea. **c)** DLS intensity fraction of the sample in panel *a)* at 5 ºC. **d)** Time evolution of the multimers followed by DLS over 15 h at 25 ºC. Left panel: diffusion coefficient (D). Center panel: Z-average. Right panel: polydispersity index (PDI). Data represents mean ± SD of three measurements of the same sample. **e)** Micrograph obtained via liquid-phase TEM showing a population of multimers, with the quantification of their sizes. Dose 20 e-/Å^2. The scale bar represents 50 nm. **f)** Top panel: chromatogram profile of the normalized absorbance at 280 nm and the molar mass integration of the SEC-MALS experiment. Bottom panel: boxplots for each peak of the chromatogram corresponding to the molar mass distribution and the deconvoluted number of monomers, taking into account that one nCPEB4-NTD molecule weighs 48444.87 Da. **g)** Calculation of diffusion coefficients (D) by fitting the gradient strength-dependent decay in intensity in the absence (lime green) and presence (dark blue) of 2 M urea. The hydrodynamic diameters (d_h) are calculated using the values of dioxane as reference. **h)** Superposition of the aliphatic region of the $^{1}H,^{13}C$-HSQC spectra in the absence (lime green) and presence (black) of 4 M urea. The specific residue types and ^{1}H position within the side chain are indicated with arrows. **i)** Superposition of the aromatic region of the $^{1}H,^{13}C$-HSQC spectra in the absence (lime green) and presence (black) of 4 M urea. The specific residue types and ^{1}H position within the side chain are indicated with arrows. **j)** $^{1}H,^{13}C$-HSQC traces of identified amino acids (one letter code) side chain atoms (as Greek letters) in the absence (lime green) and presence (black) of 4 M urea. The inset value indicates the ^{13}C chemical shift of the trace (δC / ppm). **k)** Superposition of the $N_\varepsilon/H_\varepsilon$ Arg resonances region of the $^{1}H,^{15}N$-CP-HISQC spectra in the absence (lime green) and presence (black) of 4 M urea. **l)** Schematic representation of the monomer-multimer-droplet equilibrium of nCPEB4-NTD in solution, indicating the interactions stabilizing each state.

Supplementary figure 4 (I)

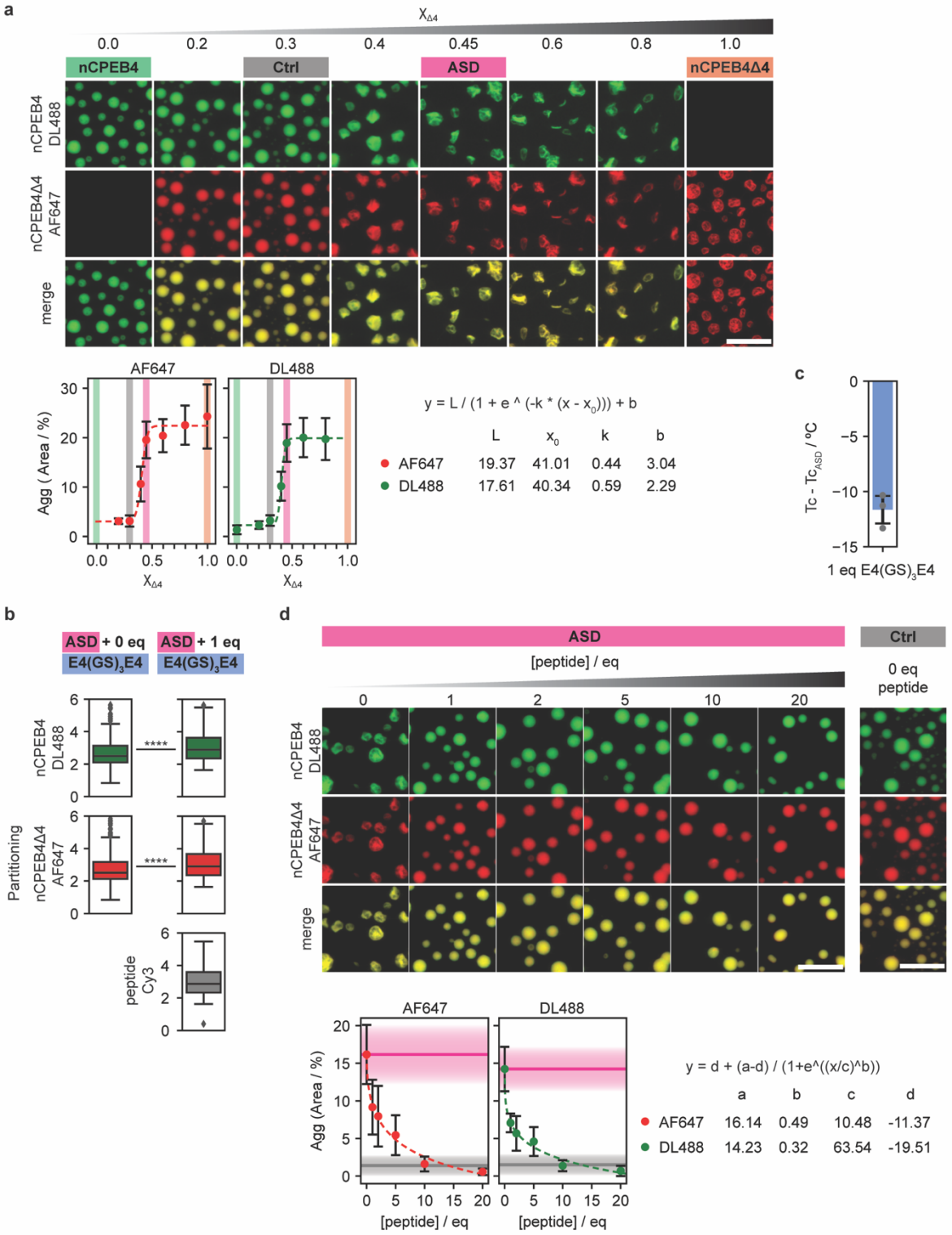




Supplementary Figure 4 | a) Left panel: urea titration, from 0 to 4 M, followed by DLS of multimeric samples (100 μM) of nCPEB4-NTD, nCPEB4Δ4-NTD, ΔHClust, and Δ4_ΔHClust. Right panel: relative volume fractions of the monomer and multimer peaks versus urea concentration (mean ± SD of three measurements of the same sample), and fitting parameters to a sigmoidal curve. **b)** xCPEB4 and xCPEB4+Ex4 complex(es) composition. Left panel: xCPEB4 and xCPEB4+Ex4 proximity partners determined by BioID in *Xenopus laevis* Profase I oocytes (n = 1). Fold increase in association of proximity partners with xCPEB4+Ex4 relative to the xCPEB4 variant is shown as peptide-spectrum match (PSM) ratios. Identified hits shown include proteins enriched in xCPEB4-BirA and xCPEB4+Ex4-BirA, relative to BirA alone. The hits include the N-terminal and C-terminal BirA fusions. Right panel: western blot of *X. laevis* oocytes injected with BirA alone or myc-tagged xCPEB4-BirA and xCPEB4+Ex4-BirA used in BioID assay shown in the left panel. Top panel: reblot against myc tag. Bottom panel: biotinylated proteins as shown with streptavidin-HRP. The indicated molecular weights are expressed in kDa. **c)** *In vivo* competition assay. Polyadenylation of *Emi2* 3′-UTR radioactive probe in the absence (–) or presence (+) of progesterone (P) in *X. laevis* oocytes injected with

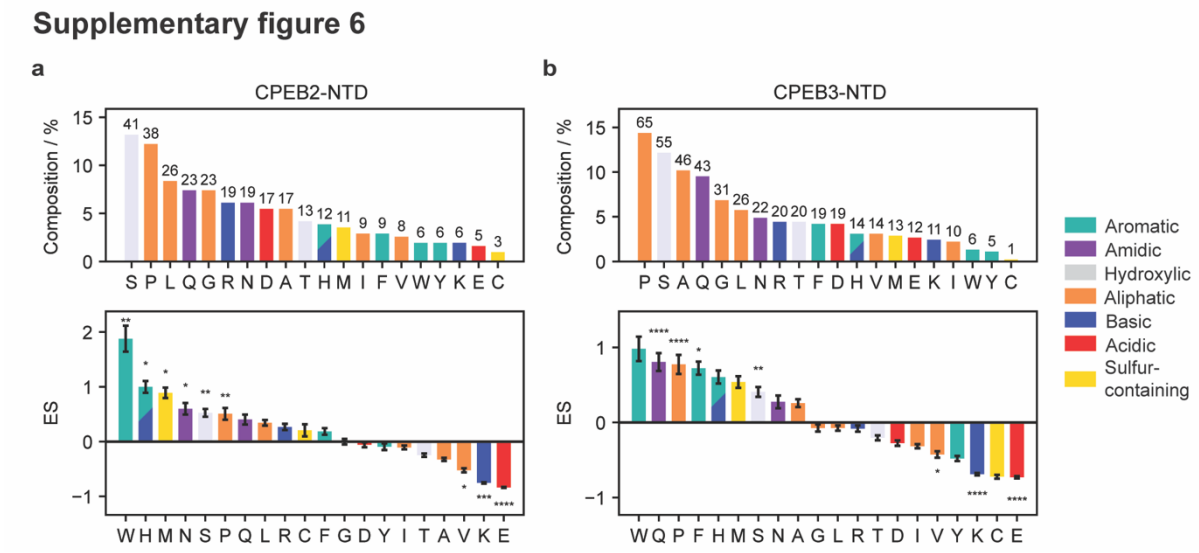
CPEB1 RNA-binding domain (ZZ), CPEB4 RNA-binding domain (RRMs), xCPEB4 (Δ4), xCPEB4+Ex4 (+Ex4), or not injected (–). **d)** nCPEB4Δ4 modified-to-total occurrence ratios determined by mass spectrometry for the indicated residue positions on basal (-stim) and stimulated (+stim) N2a cells overexpressing full-length nCPEB4Δ4-GFP. Ratios were calculated from 3 independent biological replicates. Bars represent the median. Previously undescribed post-translationally modified sites are highlighted in bold. P: phosphorylation, M: methylation, DM: dimethylation. **e)** Cloud points (mean ± SD, n = 3) of 10 μM nCPEB4Δ4-NTD at pH 7.5, 7.0 and 8.0. **f)** Schematic representation of the *in vitro* temperature reversibility experiment. Created with BioRender.com. **g)** Example of one replicate of the temperature reversibility experiment. Apparent absorbance measurement for nCPEB4-NTD (left) and nCPEB4Δ4-NTD (right) over three temperature cycles. **h)** Temperature reversibility experiment by apparent absorbance measurement (mean ± SD, n = 3). Absorbance increase (ΔA) ratio at every temperature cycle relative to the ΔA value at 0 cycles ($\Delta A / \Delta A_0$). The gray dotted line indicates an ideally fully reversible process. **i)** Time evolution up to one week from sample preparation of nCPEB4-NTD, nCPEB4Δ4-NTD, ΔHClust, and Δ4_ΔHClust followed by fluorescence microscopy. The scale bar represents 20 μm. **j)** Left panel: representative fluorescence microscopy images from time-lapse movies of N2a cells overexpressing GFP-tagged nCPEB4Δ4 and Δ4_ΔHClust mutants. Time series show pre (-stim) and post (+stim, 3, 27, and 57 min) stimulation foci disassembly. The scale bar represents 10 μm. Right panel: percentage of cells displaying cytoplasmic foci of GFP-tagged nCPEB4Δ4 and Δ4_ΔHClust mutants after depolarization (t = 57 min). Dots represent independent experiments (mean ± SD, n = 3).

Supplementary figure 5



Supplementary Figure 5 | a) Fluorescence microscopy images of samples at increasing Δ4 mole fractions (0 ≤ χΔ4 ≤ 1) 24 h after sample preparation and quantification of the fluorescence stemming from aggregated material (mean ± SD, n = 9). The scale bar represents 20 µm. **b)** Partitioning of each nCPEB4-NTD isoform and the peptide in the droplets

formed by an ASD sample ($\chi\Delta4$ = 0.45) in the absence (n = 438) and presence (n = 783) of 1 molar equivalent of E4(GS)$_3$E4 peptide, related to *Fig. 5h*. **c)** Cloud point difference (mean ± SD, n = 3) of an ASD sample with 1 molar equivalent of peptide compared to the sample without peptide (Tc - Tc_{ASD}). **d)** Fluorescence microscopy images of an ASD sample ($\chi\Delta4$ = 0.45) with increasing molar equivalents of peptide 24 h after sample preparation, and quantification of the fluorescence stemming from aggregated material (mean ± SD, n = 9). The scale bar represents 20 µm.



Supplementary Figure 6 | a) Amino acid composition (top panel) and enrichment score (ES) of each amino acid type in CPEB2-NTD (UniProt ID: Q7Z5Q1-3, residues 1–311) compared to the DisProt3.4 database (bottom panel)[64,65]. **b)** Amino acid composition (top panel) and enrichment score (ES) of each amino acid type in CPEB3-NTD (UniProt ID: Q7TN99-1, residues 1–452) compared to the DisProt3.4 database (bottom panel)[64,65].