Preferred Conformations in the Intrinsically Disordered Region of Human CPEB3 Explain its Role in Memory Consolidation

Daniel Ramírez de Mingo[1], David Pantoja-Uceda[2], Rubén Hervás[3], Mariano Carrión Vázquez[1*] and Douglas V. Laurents[2,*]

Running Title: NMR Reveals Nascent hCPEB3 Structures

[1]Instituto Cajal, IC-CSIC, Avda. Doctor Arce 37, E-28002 Madrid, Spain.
[2]Instituto de Química-Física Rocasolano, IQFR-CSIC, Serrano 119, E-28006 Madrid, Spain.
[3]Stowers Institute for Medical Research, Kansas City, MO, 64110, USA.

[†]To whom correspondence should be addressed:
MCV (mcarrion@cajal.csic.es), DVL (dlaurents@iqfr.csic.es)

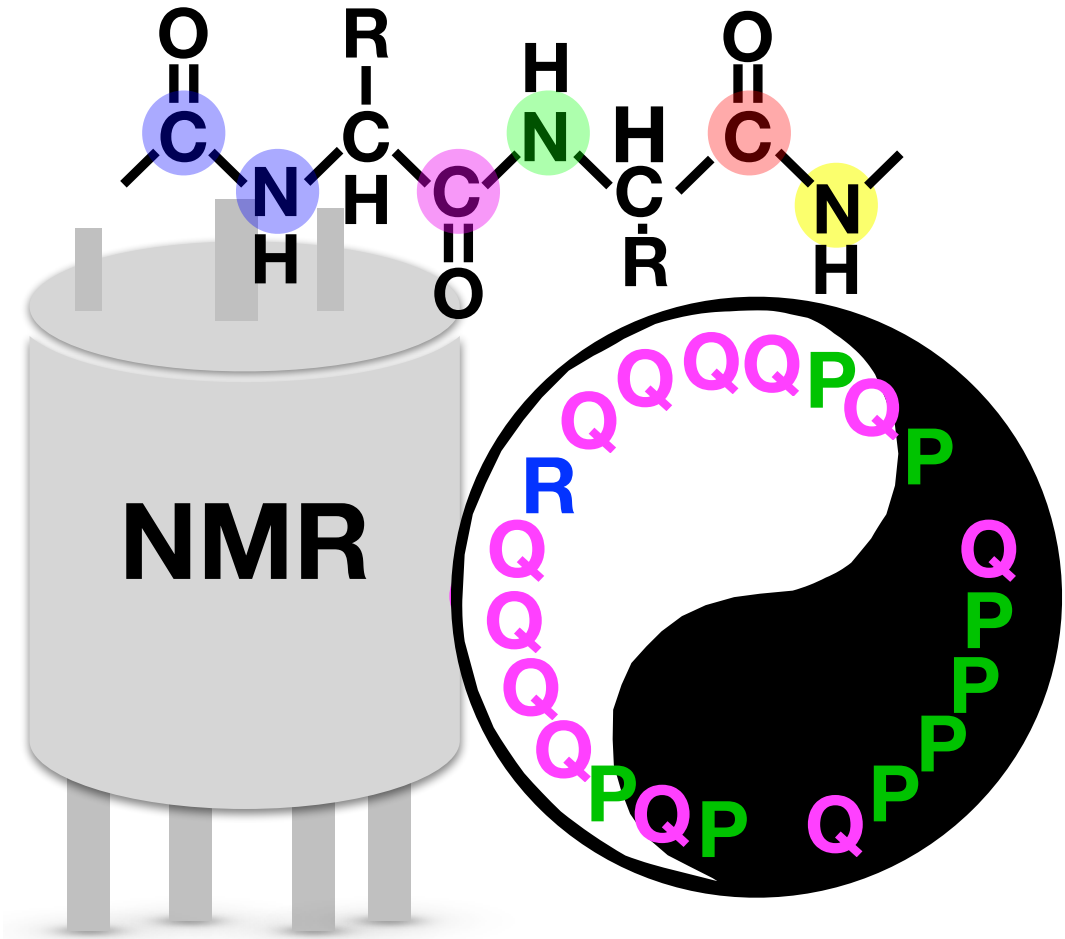
Highlights:
· Human CPEB3's aggregation-prone disordered region is studied by NMR.
· In the monomeric state, the polyQ stretch is disordered and flexible.
· 5 α-helices, 2 PPII helices and a rigid nonpolar stretch are identified.
· Association of the first 4 α-helices could promote functional amyloid formation.
· The PPII helices could negatively regulate this process.
· The last α-helix forms the NES and putatively regulates nucleocytoplasmic transport via STAT5B.

Key Words: nuclear magnetic resonance, memory consolidation, functional amyloids, intrinsically disordered proteins, amyloidogenic proteins

Short Summary

Amyloids are implicated in neurodegenerative diseases, but they are also essential to several physiological processes, including memory consolidation, neuronal-specific isoforms of the Cytoplasmic Polyadenylation Element Binding (CPEB) protein family. Here, we characterize the atomic level conformation and ps-ns dynamics of the 426-residue intrinsically disordered region (IDR) of human CPEB3 (hCPEB3), which has been associated with episodic memory in humans, by NMR spectroscopy. The first 29 residues: M_1QDDLLMDKSKTQPQPQQQQRQQQQQPQP$_{29}$, adopt a helical+disordered motif, which is key for *in vitro* amyloid formation. Residues 86-93: P_{83}QQPPPP$_{93}$, and 166-175: P_{166}PPPAPAPQP$_{175}$ form polyproline II (PPII) helices. We advance that the presence of amyloid breaker residues, such as proline, is a key difference between functional versus pathological amyloids. While the $(VG)_5$ repeat motif is completely disordered, residues 200-250 adopt three partially populated α-helices. Residues 345 – 355, which comprise the nuclear localization signal, form a modestly populated α-helix and border a phosphoTyr which may mediate STAT5B binding. These findings allow us to advance a model for hCPEB3 structural transitions in memory consolidation in humans.

Table of Contents Graphic:



Abstract

Amyloids play key pathological roles in a score of neurodegenerative diseases, but they are also essential to several physiological processes. Perhaps the most fascinating functional amyloid is formed by the Cytoplasmic Polyadenylation Element Binding protein (CPEB) protein family as its structural conversion is essential for memory consolidation in different animals. *In vitro*, CPEB amyloid formation is driven by the intrinsically disordered region (IDR) present in neuronal-specific isoforms. However, the underlying conformational transitions from the monomeric state to the amyloid state remain poorly understood. Here, we characterize the residue level conformational preferences and ps-ns dynamics for human CPEB3 (hCPEB3) by high field, heteronuclear NMR spectroscopy. At 426 residues, this is the second longest IDR characterized to date in such detail. We find that the residues 1–29: M_1QDDLLMDKSKTQPQPQQQQRQQQQQPQP$_{29}$, adopt an α-helical+disordered Q-rich motif. Similar helix + Q/N-rich motifs are observed in CPEB homologs as well as other RNA-binding proteins like TDP-43 that form pathological amyloids. Residues 86-93: P_{83}QQPPPP$_{93}$, and residues 166-175: P_{166}PPPAPAPQP$_{175}$ form polyproline-II (PPII) helices, and we propose NMR chemical shift-based criteria for identifying this conformation. We advance that the presence of helix and amyloid breaker residues, such as proline, in hCPEB3 and its absence in TDP-43 may be a key difference between the functional and pathological amyloids in higher eukaryotes. While the $(VG)_5$ repeat motif (residues 272-282) appears to be completely disordered, residues S221-A235 form a highly populated, rather rigid α-helix and two nearby segments adopt partially populated α-helices. Residues 345–355 also form a partially populated α-helix. These residues comprise the nuclear export signal and border a putative phosphoTyr site which may mediate STAT5B binding. Thus, nascent hCPEB3 protein is not fully disordered like a blank sheet of paper; instead, it contains creases which guide the engraving of our memories. Based on these findings and previous results, a working model for hCPEB3 structural transitions in human memory consolidation is advanced.

Introduction

The molecular basis of long-term memory, which endures decades despite being built by ephemeral biomolecules, has long fascinated biochemists (Crick, 1984). Seminal findings by Si, Lindquist and Kandel showed that long-term changes in synaptic efficacy require a self-perpetuating aggregative, and possibly amyloid, state in the *Aplysia* Cytoplasmic Polyadenylation Element Binding protein (CPEB) (Si *et al.,* 2003). This change in CPEB's conformation leads to permanent alterations at the synapse constituting a physical basis of memory storage. *Aplysia* CPEB contains an IDR that is very rich in glutamine residues. This Q-rich region, essential for acquiring the self-perpetuating state necessary to stabilize memory, losses α-helix and gains β-sheet structure during amyloid formation *in vitro* (Raveendrea *et al.,* 2013). However, polyQ expansions in Huntingtin protein (Jayaraman *et al.,* 2012) and the Androgen Receptor (Escobedo *et al.,* 2019) are known for their ability to form toxic aggregates in Huntington's disease and in spinobulbar muscular atrophy, respectively. Interestingly enough, Huntingtin protein has been proposed to play a role in memory consolidation (Choi *et al.,* 2014).

More recently, the *Drosophila* homolog of CPEB, called Orb2A, has been shown to behave in a similar fashion despite its lower glutamine residue content and more highly regulated amyloid formation (Majumdar *et al.,* 2012; Khan *et al.,* 2015). Indeed, inhibition of Orb2 amyloid formation specifically impairs memory consolidation, but not short-term memory in *Drosophila* (Majumdar *et al.,* 2012; Hervás *et al.,* 2016). Due to numerous His residues in the Orb2 amyloid core, pH regulates this structure's stability as shown in its recently elucidated structure (Hervás *et al.,* 2020). In mammals, the N-terminal region of neuronal-specific isoform of CPEB3 is considered to be crucial for amyloid formation and memory consolidation (Fioriti *et al.,* 2015; Stephan *et al.,* 2015). The regulation of functional amyloid formation in mammalian CPEB3 appears to be even more sophisticated due to multiple variants, post-translational modifications (Drisaldi *et al.,* 2015), and feedback loops to maintain hCPEB3 expression levels (Fiumara *et al.,* 2015). Compared to the *Aplysia* and *Drosophila* homologs, hCPEB3's overall content of glutamine residues is lower. Its IDR is very long (426 residues) and contains diverse segments which are enriched for certain residues such as Gln, Ser, Ala, Pro, Gly+Val and hydrophobic residues (**Table 1**).

Beyond the fundamental biological interest in comprehending the molecular basis of long-term human memory, this understanding could also guide the development of treatments for Post-Traumatic Stress Disorder and other diseases associated with the storage of "pathological memories" (Hyman, 2005). Furthermore, like its homologs in *Aplysia* and *Drosophila*, hCPEB3 resembles the abundant and diverse superfamily of RNA-binding proteins that contain RNA-Recognition Motifs (RRM) and/or Zinc Fingers (ZnF) domains as well as intrinsically disordered prion-like regions, such as Fused in Sarcoma (FUS) or Transactive Response DNA Binding Protein of 43 kDa (TDP-43). FUS and TDP-43 are essential proteins, but their anomalous aggregation has been implicated in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)

(Mackenzie *et al.,* 2010). In addition, TDP-43 has been linked to the recently described Limbic-Predominant Age-Related TDP-43 Encephalopathy (LATE) (Nelson *et al.,* 2019). Thus, the comparison of the physiologically relevant amyloid formation by hCPEB3 versus the aberrant amyloid formation by TDP-43 may reveal why the latter can become pathological.

Mammalian CPEB3 travels to several distinct neuronal regions to carry out multiple functions (**Figure 1A**). Following its synthesis, CPEB3 is SUMOlyated; which has been reported to block CPEB3 aggregation (Drisaldi *et al.,* 2015). Upon neuronal stimulation, CPEB3, which is mostly cytoplasmic, travels to the nucleus. This process is mediated by the karyopherin IPO5 through interactions the nuclear localization signal (NLS) in the first RRM of hCPEB3 (Chao *et al.,* 2012). Inside the nucleus, CPEB3 interacts with STAT5B, which normally activates the transcription of genes such as *EGFR* (Terlau & Seifert 1989), that trigger signaling cascades thought to promote memory consolidation (Peng *et al.,* 2010). CPEB3-STAT5B binding downregulates STAT5B-dependent transcription. This union is driven by interactions between the IDR of hCPEB3 and residues 639–700 of STAT5B, but the details of this interaction have not yet been addressed. By contrast, the 3D solution structure of the first CPEB3 RRM domain has been elucidated and revealed an additional β-hairpin (W471-G485) proposed to play a key role in RNA recognition (Tsuda *et al.,* 2014). This domain, as well as the second RRM domain and ZnF motif, were reported to bind specifically to the 3'UTR of the GluR2 mRNA (Huang *et al.,* 2006). Together, CPEB3 and its target mRNA eventually exit the nucleus. CPEB3 and its associated mRNA can join distinct biomolecular condensates, such as the stress granule during cellular stress, neuronal granules, which provide physiological transport to dendritic spines or to dendritic P-body-like granules (Ford *et al.,* 2019) where hCPEB3 stores and downregulates GluR2 mRNA translation (Huang *et al.,* 2006). Upon neuronal stimulation, CPEB3 converts into an active state, through the IDR, that promotes target mRNA translation. This process is associated with a decreased SUMOylation of CPEB3 and an increased CPEB3 aggregation in hippocampal neurons (Drisaldi *et al.,* 2015). This leads to structural modifications, including a more robust actin network, which fortify the spine and permanently enhance neurotransmission at this particular synapse (Stephan *et al.,* 2015). The hypothesis that CPEB functional amyloid formation is key for memory persistence (Si *et al.,* 2003) is supported by numerous *in vitro* experiments as well as studies in *Aplysia*, *Drosophila* and mammalian cells which have been reviewed (Si & Kandel, 2016). The causal role of CPEB3 in memory is corroborated by observations that persons carrying a rare *CPEB3* allele, which leads to a decreased production of hCPEB3 protein, have episodic memory impairments (Vogler *et al.,* 2009).

The 426 residue long IDR of hCPEB3 contains an amyloid-forming prion like domain (PLD) spanning residues 1-253, while the remaining residues, 254-426, promote liquid-liquid phase separation (Ramírez de Mingo *et al.,* 2020). The complete IDR (residues 1-426) is followed by two folded RRM which bind RNA and finally a ZZ-type ZnF domain. The complete domain structure as well as segments which are found here to be structurally and functionally important are shown in **Figure 1B**. Recent sequence and deletion mutational analyses of the IDR have begun to identify subregions key for

aggregation, such as the first 30 residues (Stephan *et al.,* 2015). Programs to predict secondary structure tendencies give different outputs and to date, no high-resolution experimental data on the partial structures or motions of the hCPEB3 IDR have been reported. Here, motivated by the key roles of the IDR in CPEB3 functions in transcription via association with STAT5B, as well as CPEB aggregation-dependent synaptic protein synthesis, we characterize the atomic level conformation and dynamics of the complete IDR of hCPEB3 by NMR spectroscopy and propose a mechanism for CPEB3 aggregation.

Materials & Methods

1. Materials: $^{15}NH_4Cl$ and ^{13}C-glucose were purchased Tracertec (Madrid, Spain), D_2O, was a product of Euroisotop, deuterated acetic acid was from Sigma/Aldrich and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as the internal chemical shift reference, was from Stolher Isotopes Chemical Company.

A twelve residue peptide, called hCPEBpep1, whose sequence corresponds to the protein's first 12 residues (M_1QDDLLMDKSKT_{12}) and a twenty residue peptide, called hCPEBpep2 whose sequence corresponds to residues P_{91}PPQEPAAPGASLSPSFGST_{110} in hCPEB3 were purchased from Genscript. hCPEBpep2's sequence overlaps with the C-terminal of Segment 1 and the N-terminus of Segment 3. The peptides were over 95% pure, as assessed by HPLC, and their identities were confirmed by mass spectrometry and NMR spectroscopy.

2. Sample Production: Coding mRNAs of hCPEB3 vary in length due to an embedded human delta virus-like ribozyme which slowly splices out introns, leading to the generation of multiple isoforms (Salehi-Ashtiani *et al.,* 2006). It is also noteworthy that Orb2A's pre-mRNA contains an intron with multiple stop codons which is only spliced out when certain "memorable" stimuli are experienced (Gill *et al.,* 2017). Here, the hCPEB3 isoform 2, Uniprot Q8NE35-2 / Genebank CAI14105.1, is studied.

Plasmid construction, protein expression and purification. The hCPEB3 IDR, which corresponds to the first 426 residues of the protein whose sequence is shown in **Table 1**, was expressed at eight highly overlapping one-hundred residue segments. To control for end effects as well as to check the reproducibility of the results, each segment overlapped by 50 residues with the preceding and successive segment.

Each segment was cloned into the pET-28a(+) by PCR using the full length human CPEB3-2 in pLL3.7 plasmid as the template kindly provided by Dr. Yi-Shuian Huang (Huang *et al.,* 2014). The DNA amplified fragments were digested with XhoI and NheI. Expression of the resulting clones led to fusion proteins containing a His_6 tag and a TEV NIα protease cleavage site. Thus, each segment studied had the sequence: MGSSHHHHHHSSGLVPRGSHMASENLYFQ, at its N-terminus.

All overlapping segments were expressed in the *E. coli* BL21 Star (DE3) strain using the T7 expression system (Novagene). $^{13}C/^{15}N$ isotopic labelling of each segment

was done by using a previously published protocol (Marley *et al.,* 2001). Briefly, cells were grown in 1 L of LB at 37 ºC by shaking at 280 rpm upon reaching optical cell densities at 595 nm (OD_{595}) ~ 0.6-0.7. Cells were pelleted by a 30 min centrifugation at 5000 × g and washed using a M9 salt solution (15.0 g/L KH_2PO_4, 34.0 g/L Na_2HPO_4 and 2.5 g/L NaCl for 1 L of 5 × M9 salts) excluding nitrogen and carbon sources. Cell pellets were resuspended in 250 mL of isotopically labelled minimal media M9 salt solution supplemented with ^{13}C D-glucose 4.0 g/L and $^{15}NH_4Cl$ 1.0 g/L (Cambridge Isotope Laboratories, Inc.), then incubated to allow the recovery of growth and clearance of unlabeled metabolites. Protein expression was induced after 1 h by addition of IPTG to a concentration of 1 mM. After a 4.5 h incubation period, the cells were harvested.

Cell pellets were lysed with buffer with the following composition: 50 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 50 mM imidazole, 6 M guanindium chloride (GdmCl), pH 7.4 and then sonicated. Each recombinant segment was purified by Ni^{2+}-affinity chromatography using HisTrap HP purification columns with a FPLC system (ÄKTA Purifier, GE Healthcare) with elution buffer consisting in 50 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, 6 M GdmCl, pH 7.4.

If necessary, the pure segments were then incubated with TEV NIα protease O/N at 4ºC (Nallamsetty *et al.,* 2004). The cleaved protein was then subjected to dialysis and recovered. Protein samples were stored at -80ºC until use. Eventually, they were desalted to 1 mM DAc, pH 4.0 by gel filtration chromatography using a PD-10 column (GE Healthcare), and concentrated to a final protein concentration of 1.0 – 1.5 mM in 250 µL using a Vivaspin microfiltration device and placed in a 5 mm Shigemi reduced volume NMR tube for measurements.

Non-labeled full length hCPEB3-IDR expression and purification was carried out essentially as described in the companion paper (Ramírez de Mingo *et al.* 2020). Briefly, cells were grown in 1 L of LB medium at 37 ºC until reaching an OD_{600} = 0.6 – 0.7 and protein expression was induced for 4 hours by adding IPTG at 1 mM final concentration. Cells were harvested and following sonication, then lysed with buffer 50 mM Na_2HPO_4, 500 mM NaCl, 50 mM imidazole, 6 M GdmCl pH 7.4. After centrifugation at 18000 rpm for 45 min, supernatants were purified with Ni^{++} affinity chromatography and elution was performed in buffer 50 mM $NaPO_4$, 500 mM NaCl, 500 mM imidazole, 3 M GdmCl, pH 7.4. Purified CPEB3-IDD was diluted to PBS, 1 M GdmCl pH 7.4, dialysed against PBS pH 7.4 at 4ºC and finally concentrated by using Amicon Ultra-15 centrifugal filters 10 kDa cut-off.

3. Sequence alignment: The conservation of vertebrate CPEB3 protein sequences was assessed using the programs T-coffee (Di Tiomasso *et al.,* 2011) and Clustal Omega (Sievers *et al.,* 2011) using the default settings. The sequences chosen as representative are human isoform 1 NP_001171608.1, human isoform 2 CAI14105.1, mouse NP_001277755.1, chicken XP_105144323.1, turtle (*Chrysemys picta belli*) XP_005301348.1, frog NP.001015925.1 (*Xenopus tropicalis*) and fish (*Danio rerio*) XP_009305819.1.

4. NMR Spectroscopy: Instrumentation: All spectra for the hCPEB3 segments were recorded on a Bruker 800 MHz (^{1}H) Avance spectrometer fitted with a triple resonance TCI cryoprobe and Z-gradients. The ^{1}H chemical shift was referenced to 50 μM DSS measured in the same buffer and at the same temperatures as those used for the CPEB3 segments. Since DSS can sometimes bind to instinsically disordered proteins (Diez-García *et al.,* 2012) the DSS signal was recorded in an independent reference tube containing the same buffer at the same temperature. The ^{13}C and ^{15}N chemical shift references values were calculated by multiplying by their respective gyromagnetic ratios with ^{1}H; that is Ξ ^{13}C / ^{1}H = 0.251449530 and Ξ ^{15}N / ^{1}H = 0.101329118 (Markley *et al.,* 1998). NMR spectra were recorded and transformed using TOPSPIN (versions 2.1) (Bruker Biospin).

Nonconventional Assignment Strategy: To speed and improve the assignment of the backbone, we used a "proton-less" NMR approach for segments 1, 4, 5, 6 and 8 based on 2D CON spectra in which successive ^{15}N-^{13}CO nuclei correlations are obtained in two 3D spectra called hacacoNcaNCO and hacaCOncaNCO (Pantoja-Uceda & Santoro, 2014). For segments 7 and 8, which tend to form condensates (Ramírez de Mingo *et al.,* 2020) and seem to be more rigid, this strategy afforded less intense spectra and in particular about 35 CON crosspeaks were missing. Therefore, an additional strategy based on ^{13}CO connectivities from 3D HNCO and HNcaCO spectra, as well as ^{1}HN and ^{15}N connectivities of consecutive residues from 3D HncocaHN and hNcocaNH spectra (Sun *et al.,* 2005; Pantoja-Uceda & Santoro, 2009) was utilized to check and complete the backbone assignments. The latter strategy was also employed for segment 3, which was less soluble. For all segments, further corroboration was obtained by conventional 2D ^{1}H-^{15}N HSQC and 3D HNCO spectra as well as 3D CCCON to confirm the residue identity and obtain the chemical shift values of ^{13}C nuclei of the side chain.

Of the eight segments, only segment 2 failed to yield a soluble sample. Whereas the sequence assignments are complete thanks to the analysis of segments 1 and 3, to test for possible end effects, a 20 residue peptide, hCPEBpep2, corresponding to residues 91-110 of the hCPEB3 sequence was assigned and characterized structurally by 2D ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOSCY, ^{1}H-^{1}H NOESY and 2D ^{1}H-^{13}C HSQC NMR spectra at 5.0 ºC on a Bruker 600 MHz spectrometer fitted with a cryoprobe and Z-gradients. A 2D ^{1}H-^{15}N HSQC was also recorded on the 800 MHz Bruker spectrometer. Both the ^{1}H-^{15}N HSQC and the ^{1}H-^{13}C HSQC spectra were recorded at the natural abundance of ^{15}N and ^{13}C. The program NMRFAM-Sparky (Lee *et al.,* 2015) was used to facilitate manual spectral assignment. The NMR spectral parameters are summarized in **Sup. Table 1**.

Theoretical chemical shift values for statistical coil ensembles were calculated using the parameters tabulated by Kjaergaard and Poulsen (2011) and Kjaergaard *et al.,* (2011), as implemented on the https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/ server at the Bax laboratory. These values were used to calculate conformational chemical shifts. When appropriate, representative preferred conformers were calculated using the program CYANA 3.98 (Günter & Buchner, 2015).

Coupling Constants: For segment 5, as an additional, independent test, a 3D HNHA spectrum were recorded and the ratio of the $^{1}H\alpha$-^{1}HN crosspeak to the ^{1}HN-^{1}HN diagonal peak intensities in the 3D HNHA spectrum was utilized to calculate the $^{3}J_{HNCH\alpha}$ coupling constants following the procedure of Vuister and Bax (1993). $^{3}J_{HNCH\alpha}$ coupling constants were also measured for hCPEBpep2 using the 2D ^{1}H-^{1}H COSY spectrum. Utilizing the Karplus equation, (Karplus, 1963), these $^{3}J_{HNCH\alpha}$ coupling constants can be related at the backbone ϕ angle, which is different for α-helical, statistical coil and β-strands.

Relaxation: To assess the dynamics on the ps–ns time scales, the heteronuclear $^{15}N\{^{1}H\}$ NOE (hNOE) of backbone amide groups was registered as the ratio of spectra recorded with and without saturation in an interleaved mode. Long recycling delays of 13 s were used. Two sets of experiments, one at 25 ºC and one at 5 ºC were recorded at 800 MHz. Uncertainties in peak integrals were determined from the standard deviation of intensities from spectral regions devoid of signal which contain only noise.

In addition, $R_1\rho$ relaxation rates, which are sensitive to the presence of preferred, rigid conformers on slower µs-ms timescales were measured by recording two sets of ten ^{1}H-^{15}N correlation spectra with relaxation delays at 8, 300, 36, 76, 900, 100, 500, 156, 200 and 700 ms. One set of experiments was recorded at 25 ºC and the second was recorded at 5 ºC. The relaxation rates were calculated by least-squares fitting of an exponential decay function to the data using NMRPipe (Delagio *et al.,* 1995). As an additional check, the data were also analyzed independently by using the program DynamicsCenter 2.5.2 (Bruker Biospin).

Results

hCPEB3's IDR is chiefly disordered. As a first step to experimentally characterize hCPEB3's IDR, we probed the complete 426 residue IDR of hCPEB3 by biophysical techniques and homonuclear NMR. Its fluorescence emission spectra, recorded at temperatures ranging from 2 to 70 ºC, show emission maximum > 350 nm. This is consistent with its six Trp residues being solvent exposed and not buried in the hydrophobic core of a folded domain (**Sup. Fig. 1A**) (Alston *et al.,* 2004). The far UV CD spectrum of the hCPEB3 IDR also shows the hallmarks of a disordered protein, namely a minimum near 200 nm (Denning *et al.,* 2003). No spectral features indicative of α-helix and β-sheet; namely, minima at 208, 218 or 222 nm and no maximum at 195 nm, are evident (**Sup. Fig. 1B**). The 1D ^{1}H and 2D ^{1}H-^{1}H NOESY spectra show ^{1}H signals clustered into a narrow bands near the values observed for unstructured short peptides (**Sup Fig. 1C**) (Bundi & Wüthrich, 1979; López-Alonso *et al.,* 2010). The sequence alignment of several representative vertebrate CPEB3 proteins using the T-Coffee program is shown in **Sup. Fig. 2**. Very similar results were obtained from the Omega Clustal program (not shown). Whereas most IDPs show poor levels of sequence conservation, some stretches rich in hydrophobic residues, such residues M1-T12, W111-F139 and Y341-I357, are highly conserved. By contrast, glutamine rich, alanine rich and

some proline rich segments are present only in mammals. Taking all these data together, the presence of large, stably folded domains in the IDR can be ruled out, but short segments with partly populated secondary structures could still be present.

Atomic level characterization reveals partially structured elements in hCPEB's N-terminal "disordered" region. To discover and characterize possible segments with partial secondary structure, we applied multidimensional heteronuclear NMR. As the full length IDR is too long to characterize by this methodology, we have followed the "divide and conquer" approach developed by Zweckstetter *et al.* to characterize tau, a similarly sized IDP implicated in Alzheimer's disease and other tauopathies (Mukrasch *et al.,* 2009). As described in the **Material & Methods** section, and shown in **Fig. 1**, eight overlapping segments of 100 residues were characterized.

Using our powerful ^{13}CO, ^{15}N, ^{1}HN based assignment strategy, over 99% of the main chain ^{13}CO, ^{13}Cα, ^{15}N, ^{1}HN and the ^{13}Cβ resonances were assigned for residues 1 – 450 of hCPEB3. The assigned 2D ^{1}H-^{15}N HSQC and 2D ^{13}CO^{15}N spectra of segment 4 are shown in **Sup. Fig. 3** and **Sup. Fig. 4**, respectively. The chemical shifts of the complete IDR of hCPEB3 are reported in the **BMRB (**entry number **50256)**. Multiple attempts to express and purify hCPEB3 segment 2, which spans residues 51-150, by recombinant methods were unsuccessful. Nevertheless, all the residues within segment 2 are present and have been characterized structurally in the context of segments 1 and 3. To test if there might be some structure in the neighborhood of residues 90-110 which corresponds to the middle of segment 2 and the C- and N-termini of segments 1 and 3, respectively, we studied the conformation of a twenty residue peptide corresponding to residues 91 - 110 of hCPEB3 by NMR spectroscopy. No significant trends towards structure formation were detected (**Sup. Fig. 5**).

The first residues of hCPEB3 adopt a partly populated α-helix which precedes the Q-rich stretch. The conformational chemical shifts point to the formation of partly populated α-helix in the first ten residues of the protein (**Figure 2A,B**). Increased conformational chemical shifts are observed at 5 ºC, reflecting a higher amount of helical structure upon cooling. Relaxation measurements detect that these residues are the most rigid part of segment 1 (**Figure 2 C,D**).

The helix detected for these first residues extends N-terminally into the His/Tev tag. Although the presence of this tag does not appear to increase the population of helical structures in the other segments studied, to rule out a possible structure-promoting effect on segment 1, we tried to remove it by proteolytic cleavage with the TEV protease. Multiple attempts failed, which suggests that the helix spanning the last residues of His/Tev tag and the first residues of the hCPEB3 IDR is present and impedes the proteolytic cleavage. Therefore, we characterized a dodecamer peptide whose sequence corresponds to the first twelve residues: M_1QDDLLMDKSKT_{12}, of the hCPEB3 IDR. The observation of a series of weak $^{1}HN_i$ - $^{1}HN_{i+1}$ nuclear Overhauser enhancement (NOE) crosspeaks reveals that this peptide has a slight tendency to form α-helix in aqueous buffer (**Sup. Fig. 6A**). Fluorinated alcohols like trifluoroethanol (TFE) and

hexafluoroisopropanol (HFIP) are known to increase the population of helical conformations in peptides which have an α-helix forming tendency, but not in peptides which prefer to adopt β-strands or random coil (Muñoz *et al.,* 1995). In the presence of 20% HFIP, the population of helix in this peptide increases strongly, based on the observation of stronger and more numerous NOE crosspeaks as well as ^{1}Hα and ^{13}Cα conformational chemical shifts (**Sup. Fig. 6B**). These findings confirm that the first 12 residues of hCPEB3 do tend to adopt an α-helix.

Interestingly enough, the polyQ segment, Q_{16}QQQRQQQQ_{24}, does not form an α-helix or a β-strand and appears to be thoroughly disordered and flexible (**Figure 2A,B**). This polyQ segment is preceded and followed by Pro-Gln-Pro residue triplets ($P_{13}QP_{15}$ and $P_{25}QP_{27}$). Considering the inhibitory effect of proline residues previously observed for polyQ amyloid formation in Huntingtin by Wetzel and co-workers (Bhattacharyya *et al.,* 2006), it is likely that these PQP mini-motifs check amyloidogenesis by the polyQ segment. The first 100 residues also contain a predicted SUMOylation site (Ramírez de Mingo *et al.,* 2020) at Lys 47 and ends with a proline-rich segment P_{86}PQQPPPPQEPAAPG_{100}, which is associated with solubility. Whereas recently reported NMR criteria (Treviño *et al.,* 2018), allow us to rule out that this stretch folds into a stable polyproline II (PPII) helical bundle, residues 86-93 adopt an isolated, partly populated PPII helix (**Figure 2E**). In fact, the consecutive proline residues show a distinct pattern of conformational chemical shifts; namely +0.6 ppm, -1.0 and -0.3 for ^{13}Cα, ^{13}Cβ and ^{13}CO, respectively (**Sup. Fig. 7**). Not observed in isolated proline residues, we advance that they are hallmarks of a PPII helical conformation.

Residues 101 - 200 of hCPEB3 contains a rigid nonpolar segment and a PPII helix. Regarding residues 101 – 200, no strong trends to adopt α-helical or β-structures are detected. Nevertheless, the stretch composed of residues, W_{111}STGTTNAVEDSFFQGITPVNGTMLFQNF_{139} which contain numerous aliphatic and aromatic residues, shows relatively high rigidity, both on fast ns/ps as well as slower μs/ms timescales (**Sup. Fig. 8)**. This finding is interesting considering that this segment also appears to be essential for hCPEB3 amyloid formation *in vitro* (Ramírez de Mingo *et al.,* 2020). In addition, the stretch of residues 161-190: Q_{161}HHQQPPPPA_{170}PAPQPAQPAQ_{180}PPQAQPPQQR_{190}, has a very high Q/P content and Pro and Gln are the residues with the highest intrinsic tendencies to adopt PPII helices (Kelly *et al.,* 2001). The consecutive proline residues, P_{166}PPPAPAPQP_{175}, also display the characteristic PPII pattern of conformational chemical shifts (**Sup. Fig. 7**) seen for residues 86-93. We hypothesize that these segments may bind profilin, which has specific domains to interact with proline-rich polypeptides (Metzler *et al.,* 1994) and to mediate interactions with the actin, whose levels rise at the synapse following CPEB3 aggregation (Stephan *et al.,* 2015).

Residues 201-300 contain three α-helical segments and a disordered $(VG)_5$ segment. Upon observing significant chemical shift deviations with respect to values predicted for a statistical coil, we utilized the programs TALOS+ and CYANA to calculate a family

for conformers for residues $\underline{A_{202}QRSAAAY_{210}}GHQPIMTSKP_{220}$-$\underline{S_{221}SSSAVAAAA_{230}AAAAA}SS\underline{ASS_{240}SWNTHQ}SVHAA_{250}$ (**Figure 3A**). The three segments of underlined residues have been detected to adopt partially populated α-helices. Based on the magnitude of the conformational chemical shifts, the helical populations are different; being about 30% for the A_{202}-Y_{210} α-helix, 80% for the S_{222}-A_{234} α-helix and 20% for the A_{238}-Q_{246} α-helix at 5 ºC; these populations decrease but are still significant at 25 ºC (**Figure 3B**). The presence of the first and second helices are confirmed by ^{1}HN-^{1}Hα coupling constants (**Sup. Fig. 9**). The helices are not especially rigid on fast ps-ns time scales (**Figure 3C**) or the slower μs-ms time domain (**Figure 3D**) at 25ºC, but do show a heightened stiffness at 5 ºC (**Sup. Fig. 8**). Helical wheel projections (**Sup. Fig. 10**) suggest that different stabilizing interactions contribute to these helices. Gly 211 and His 212 are positioned to stabilize the A_{202}-Y_{210} α-helix by a C-capping motif (Aurora & Rosa, 1998). Whereas Ala has a very high intrinsic helix forming propensity, the propensity of Ser is low (Chakrabartty *et al.,* 1994). In this segment, however, the Ser residues are positioned at the N-terminus of the α-helices, where adding negative charge via phosphorylation would increase the helical population, considering the well-known stabilizing effects of charge / macrodipole interactions and N-capping H-bonds (Chakrabartty & Baldwin 1995). The last α-helix is less populated, but its stability would be increased if W242 were to engage in long-range interactions.

One of the most striking features in hCPEB3's sequence is a short dipeptide protein motif $(Val\text{-}Gly)_5$ spanning residues 271 – 281, which is reminiscent of longer $(Ala\text{-}Gly)_N$ and $(Pro\text{-}Gly)_N$ and $(Arg\text{-}Gly)_N$ dipeptide repeat proteins encoded by mutant C9orf72 which have been implicated in ALS (Morón-Oset *et al.,* 2019); (Mizielinska *et al.,* 2014). Our *in silico* analysis identified this segment as having a high potential to form amyloid (Ramírez de Mingo *et al*., 2020). In the context of hCPEB3, however, this segment is among the most disordered and flexible of all the zones of the protein. We speculate that these exposed and flexible Val residues may be poised to interact with a still unidentified factor through the formation of hydrophobic interactions.

Just beyond the $(VG)_5$ segment, there is a stretch of 15 residues: S_{284}PLNPISPLKKPFSS$_{298}$, whose NMR parameters indicate disorder and flexibility (**Sup. Tables 1, 2**). Nevertheless, this stretch contains four Ser residues reported to phosphorylated by protein kinase A (PKA) or calcium/calmodulin-dependent protein kinase II (Kaczmarczyk *et al.,* 2016) (**Table 1**), and therefore might be important for the transition between short- and long-term memory. Residues P_{303}-PKFP**R**AAP$_{311}$ are proline rich. Predictions suggest that Arg 308 can be methylated (**Table 1**). This modification has been reported to fortify cation–π interactions, reduce interactions with RNA and destabilize condensates in other proteins (Chong *et al.,* 2018).

The residues forming the NES show a marked tendency to adopt α-helical structures. Significant conformational chemical shifts were also observed for residues L349-L353 which form the NES (**Figure 1**, **Figure 4, below),** indicating the presence of helical structure. Using the ^{13}CO, ^{15}N, ^{1}HN, ^{13}Cα and ^{13}Cβ chemical shift data as input, a family of conformers was calculated using the programs TALOS+ and CYANA for

residues P333-P363. This 31-residue segment is rich in aromatic (five) and aliphatic (six) residues, which is unusual for a disordered polypeptide. The resulting structures reveal that residues L346-L349 adopt one turn of α-helix and residues S352-M356 form a short α-helix (**Figure 4A**). The family of calculated conformations shows a high structural similarity in the helical regions (residues L346 – M356) with an RMSD of 0.9 Å for the backbone and 1.6 Å for all heavy atoms. It is notable that this conformer positions five nonpolar residues: L346, L349, L353, M354 and I357 on the same face of the α-helices. Y341, the putative phosphorylation site, is in an extended portion of the backbone and would be accessible for this PTM. Whereas the conformational ensemble will contain many other structures, based on conformational chemical shifts as illustrated by the $\Delta\delta^{13}C\alpha$ & $\Delta\delta^{13}CO$ values shown in **Fig. 4B** the α-helical population is about one third. The presence of rigid conformers is corroborated by relatively high $\{^{1}H\}$-^{15}N NOE ratios (**Fig 4C**) and elevated transverse relaxation ratios (**Fig. 4D**).

Beyond the NES α-helix, no segments with preferred secondary structure are detected. The last residues of the segment 8 construct: S_{426}-RKVFVGGLPPDIDEDEITASFRRF$_{450}$, belong to the RRM1 domain. According to the 3D structure (Tsuda *et al.,* 2014) residues K428-G432 adopt a β-strand and residues E440-R449 form an α-helix in the context of the complete RRM1 domain. Here these segments appear to be largely disordered. After a proline-rich zone ending around residue 380, the next fifty residues have a higher content of nonpolar residues and tend to be more rigid (**Sup. Fig 8**). Residues 400 – 412: SHGDQALSSGLSS contain five Ser residues reported to be phosphorylated (Kaczmarczyk *et al.,* 2016, **Table 1**). Like the four Ser residues of the 284-291 stretch mentioned above, this could be an element involved in the structural basis of memory consolidation.

Discussion

The biophysical analysis of the full length IDR of hCPEB3 shows that it lacks stable secondary structure. Compared to well-ordered protein domains, disordered regions tend to change their amino acid sequence much more rapidly over the course of evolution due to a lack of structural constraints for folding (Brown *et al.,* 2002). In hCPEB3, the N-terminal α-helix and the 350's (NES) α-helix as well as its accompanying putative phosphoTyr site, are well conserved throughout vertebrate CPEB3's, which strongly suggests that they are important for biological function (**Sup. Fig. 2**). Additional segments, like the stretch of hydrophobic residues key for amyloid formation (Ramírez del Mingo *et al.,* 2020) and a cluster of three Trp residues (W242, W252 and W259) are also strictly conserved from mammals to fish. In contrast, the Gln-rich segment at the N-terminus, the Pro-rich 'breaker' regions and the Ala-rich helices are well conserved in mammals but not across all vertebrates. In some lower vertebrates, there is an alternative Q-rich region positioned after the 100's hydrophobic segment. The relatively rapid evolution of these elements could be related to the development of the mammalian brain as compared to those of lower vertebrates. By contrast, the ability to move the polyQ

segment or substitute it for hydrophobic amyloidogenic segment and still retain function highlights the cassette or modular nature of PLDs which was previously established for the *Drosophila* homolog, Orb2A (Hervás *et al.,* 2016).

Over the past decade, the PLD of several proteins involved in transcription regulation and/or RNA metabolism have been characterized. Some of these, including Huntingtin (Jayaraman *et al.,* 2012), the Androgen Receptor (Eftekharzaedeh *et al.,* 2016), Ataxin-7 (Hong *et al.,* 2019) and TDP-43 (Lim *et al.,* 2016) (**Table 2**) have a hydrophobic, helix-forming segment which precedes the Q- or Q/N-rich amyloid-forming segment. These hydrophobic helices have been proposed to promote intermolecular associations, either through the formation of coiled-coils (Pelassa *et al.,* 2014) or by forming biomolecular condensates via liquid-liquid phase separation (Conicella *et al.,* 2016). In the case of Huntingtin, the polyQ segment is followed by a stretch of polyproline, which will be locked in a PPII helical conformation. This polyP stretch has been shown to strongly slow Huntingtin amyloid formation *in vitro* (Bhattacharyya *et al.,* 2006).

PPII helices may regulate amyloid formation and mediate interactions with actin. It is important to point out that whereas hCPEB3 and its homologs Orb2 and *Ap*CPEB have helix destabilizing residues between the N-terminal helix and the amyloidogenic Q-rich segment, such as proline in hCPEB3 and Orb2A or serine and valine in the case of *Ap*CPEB, no such helix busters are present in the analogous segments of Huntingtin, the Androgen Receptor or TDP-43 (**Table 2**). The absence of these breakers allows the polyQ segment to interact with the helix through the formation of sidechain to main chain hydrogen bonds in the case of the Androgen Receptor (Escobedo *et al.,* 2019) and to initiate as well as to form part of the amyloid structure as proposed recently for TDP-43 by Eisenberg and collaborators (Cao *et al.,* 2019). Based on these observations, we propose that the presence of helix breakers between the helix and the polyQ(/N) segments in functional amyloids and their absence in pathological amyloids is a fundamental difference that could be a molecular basis of their radically different toxicities.

PPII helices, well known in collagen and certain glycine-rich proteins (Pentelute *et al.,* 2008), also play key roles in mediating protein-protein interactions in biomolecular condensates (Guo *et al.,* 2019). Whereas conformation chemical shifts have proven to be extremely useful tools to identify α-helices and β-strands in folded and disordered proteins, conformational chemical shifts for PPII helices were unknown. The first 450 residues of hCPEB3 include no less than 66 proline residues (15% of the total). On the basis of the thorough set of chemical shifts obtained here, including Pro ^{15}N assignments which are rare in the literature, we propose a pattern of conformational chemical shifts that define PPII helices (**Table 3**). Combined with the recently published standards for glycine-rich PPII helical bundles (Treviño *et al.,* 2018), these values should aid the detection of PPII helices in biomolecular condensates and the elucidation of their roles in physiology and pathology.

α-Helices are stabilized by distinct interactions in pathological and functional amyloids. The disposition of the side chains in the N-terminal α-helices of TDP-43 and CPEB homologs is shown in **Sup. Fig. 10**. TDP-43's α-helix is almost six turns long. The first three are stabilized by numerous Ala and Met residues, which have a high intrinsic helix forming propensity. Most hydrophobic residues are clustered on the nonpolar face of the helix. In contrast, Trp334 is exposed on the polar face. This Trp is known to mediate intermolecular contacts which drive liquid/liquid phase separation (Li *et al.,* 2018). The α-helix of *Ap*CPEB is composed of four turns; the first two have a well defined hydrophobic face while the third and fourth are more polar. Leu and Asn are known to contribute stability and specificity to interhelical contacts in coiled-coils (Oakley and Kim, 1998). Orb2's putative α-helix is shorter with only three turns and may be less stable based on its content of four Asn and one Val residues, which have poor intrinsic helix propensities. It is amphiphilic with well defined polar and nonpolar faces. Phe 5 lies at the center of the hydrophobic face. This residue has been shown to be essential for proper amyloid formation and memory consolidation (Majumdar *et al.,* 2012). Interestingly enough, some residues following the poly-Q segment of Orb2; namely P_{54}-NLSALH$_{60}$, might adopt an α-helix stabilized by favorable *i, i+3* hydrophobic interactions (L58-L61), N-capping (P56, N57) and charge (H62) / helix macrodipole interactions (Chakrabartty & Baldwin, 1995).

The first 30 residues of CPEB3 are very well conserved in mammals and rather well conserved between mammals and reptiles, which speaks to their functional relevance (**Sup. Fig. 2**). The N-terminal helix of hCPEB3 is composed of residues whose intrinsic helix propensities are significantly lower than those of TDP-43. This suggests that long-range interactions might be necessary for it to become fully populated. Most of the hydrophobic residues of hCPEB3 are clustered on one side of the helix, to form a nonpolar face. The exception is Met 7, which is on the hydrophilic face. Such exceptionally exposed hydrophobic residues, like Trp 334 in TDP-43 (Cao *et al.,* 2019), and perhaps Trp 242 in the hCPEB3 A_{238}-Q_{246} α-helix, often mediate long-range interactions.

STAT5B binding may occlude hCPEB's NES leading to nuclear retention. The calculated conformers for the NES showed the preferred formation of α-helices which orient five hydrophobic residues on the same side (**Fig. 4**); this is a common feature on NES (Xu *et al.,* 2012). CPEB3 constantly shuttles in and out of the nucleus and its subcellular localization has important consequences as it regulates gene transcription via association with STAT5B in the nuclei of neurons as well as mRNA translation in dendritic spines. Previous studies reported that neural stimulation leads to hCPEB3, which is normally predominately cytoplasmic, to be concentrated in the nucleus. An earlier study (Peng *et al.,* 2010) revealed that the region spanning residues 639 – 700 of STAT5B binds the N-terminal IDR of hCPEB3. To learn about the nature of this interaction, we examined the recently elucidated structure of STAT5B (PDB: 6MBW), (de Araujo ED *et al.,* 2017). Remarkably, this structure reveals that the key hCPEB3-interacting residues in STAT5B form an SH2 domain. Based on this, we advance that hCPEB3 may contain a phospho-Tyr residue that is recognized and bound tightly by

STAT5B. Considering the results of previously reported very thorough motif analyses (Liu *et al.,* 2010 & Corwin *et al.,* 2017), the most promising site for phospho-Tyr modification in hCPEB3 is Tyr341, as this residue is bordered by SRPY$_{341}$DTF which closely matches a consensus substrate "[E/V/D]-[G]-[I/P/L]-Y-E-X-[F/T/V]" of Scr, which is one of the most relevant tyrosine kinases. Based on this proposed PTM site, and our finding that the preferred conformation of the nearby NES (L349-L353) is an α-helix (*vide supra*), we can advance a working hypothesis to account for how STAT5B binding retains hCPEB3 in the nucleus; namely binding of phospho-Tyr341 by the SH2 domain of STAT5B would occlude the neighboring NES of hCPEB3. This would explain why hCPEB3 becomes concentrated in the nucleus under conditions of neuronal stimulation (Chao *et al.,* 2012). Recent studies have shown that certain mutations in the nuclear localization and export signals or defects in nucleocytoplasmic transport of TDP-43, another aggregation-prone, RNA-binding protein, lead to an aberrant aggregation (Gasset-Rosa *et al.,* 2019). Based on this, we propose that mutations in the NES of hCPEB3 (residues L346–M356) would likely disrupt transport and lead to memory and learning defects.

Towards an atomic-level understanding of the role of hCPEB3 in memory consolidation. By combining the results obtained here with observations that residues 1-250 are necessary and sufficient for amyloid formation *in vitro* (Ramírez de Mingo *et al.,* 2020), we advance a working hypothesis for hCPEB3 conformational changes during memory consolidation (**Figure 5**). It is supported by analogous results on *Aplysia* CPEB (Raveendrea *et al.,* 2013) and polyalanine expansions (Polling *et al.,* 2015), which led to the proposal of similar mechanisms. Initially, the first 250 residues of CPEB3, which are known to be necessary and sufficient for aggregation and amyloid formation (Fioriti *et al.,* 2015; Ramírez de Mingo *et al.,* 2020), are mostly disordered except for the relatively stable α-helices formed by residues 222-234 and modestly populated α-helices at the N-terminus and spanning residues 202-212 and 237-245 (**Fig. 5A**). The Q_4RQ_4 motif and hydrophobic segment **F$_{123}$FQGIT**-PVNGT-**MLFQNF$_{139}$** are initially disordered and premature amyloidogenesis is discouraged by SUMOlyation (Drisaldi *et al.,* 2015) and proline breaker motifs.

Following deSUMOlyation and other putative PTMs such as phosphorylation, association among the α-helices would become possible, and might be strengthened by hydrophobic interactions between Met 7 and Trp 242 or cation-π interactions between Lys 11 and Trp 242 (**Fig. 5B**). The increased production of hCPEB3 upon neuronal stimulation (Fioriti *et al.,* 2015) as well as the proximity of the alanine rich α-helix (residues 222-234) may promote α-helix formation by the QQQQRQQQQ motif. These could then associate to form a coiled-coil, as has been demonstrated in model polypeptides (Fiumara *et al.,* 2010; Pelassa *et al.,* 2014).

These events would reinforce intermolecular contacts within the dendritic P-body like granule (Ford *et al.,* 2019) leading to gelification and, eventually, amyloid formation (**Fig. 5C**). The final amyloid structure could be comprised of the Q_4RQ_4 motif and the hydrophobic segment. In addition, the polyA segment of helices 202-212 and 222-234

may also transform into amyloid (**Fig. 5D**). Such α-helix to amyloid conformational transformations have been previously described in alanine-rich polypeptides as diverse as a fish antifreeze protein (Graether *et al.,* 2003) and a synthetic $(Ala)_{10}$-$(His)_6$ hexadecapeptide (Hamley *et al.,* 2014), and could form pathological amyloids in a number of reported diseases involving expansion of polyAla tracts (Albrecht & Mundlos 2005).

Residues 217-284 of CPEB3 have been reported to be key for interactions with actin (Stephan *et al.,* 2015). Considering this, the hypothetical associations among the α-helices formed by residues M_1-T_{12} and residues A_{202}-Q_{246} advanced in the preceding paragraphs could dispose the PPII helices formed by residues P_{86}-Q_{94} and P_{116}-A_{170} to bind profilin. This protein contains a second binding site specific for actin and promotes the formation of actin filament networks (Ferron *et al.,* 2007). Such actin networks are known to become more extensive and robust as a dendritic bud strengthens during memory consolidation (Basu & Lamprecht, 2018).

Acknowledgements: This study was supported by projects SAF2016-76678-C2-1-R (MC-V) and SAF2016-76678-C2-2-R (DVL) from the Spanish Ministry of Economy and Competitivity. NMR experiments were performed in the "Manuel Rico" NMR Laboratory (LMR) of the Spanish National Research Council (CSIC), a node of the Spanish Large-Scale National Facility (ICTS R-LRB). We are grateful to Dr. Miguel Mompeán and Dr. Javier Oroz for critical comments on the manuscript.

Author Contributions: 1. Funding: MC-V & DVL; 2. Research planning & design: DRdM, RH, DP-U, MC-V and DVL. 3. Cloning, sample production and purification: DRdM & RH; 4. Data acquisition: DP-U & DVL; 5. Data analysis: DP-U, DRdM & DVL; 6. Writing: DRdM & DVL. All authors carefully read the MS, provided corrections and approved the final version of the MS.

References

Albrecht, A., & Mundlos, S. (2005). The other trinucleotide repeat: polyalanine expansion disorders. *Current opinion in Genetics & Development*, *15*(3), 285–293. https://doi.org/10.1016/j.gde.2005.04.003

Alston, R. W., Urbanikova, L., Sevcik, J., Lasagna, M., Reinhart, G. D., Scholtz, J. M., & Pace, C. N. (2004). Contribution of single tryptophan residues to the fluorescence and stability of ribonuclease Sa. *Biophys. J.,* 87(6), 4036–4047. https://doi.org/10.1529/biophysj.104.050377

Aurora, R. & Rose, G. D. (1998). Helix capping. *Prot. Sci. 7*: 21-38.

Basu, S., & Lamprecht, R. (2018). The Role of Actin Cytoskeleton in Dendritic Spines in the Maintenance of Long-Term Memory. *Frontiers in Molecular Neuroscience*, *11*, 143. https://doi.org/10.3389/fnmol.2018.00143

Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J., & Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of Molecular Evolution*, *55*(1), 104–110. https://doi.org/10.1007/s00239-001-2309-6

Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., Creamer, T. P., & Wetzel, R. (2006). Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol., 355*(3), 524–535. https://doi.org/10.1016/j.jmb.2005.10.053

Bundi, A., & Wüthrich, K. (1979) ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solution of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers 18*(2), 285-297.

Cao Q, Boyer DR, Sawaya MR, Ge P, Eisenberg DS. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat Struct Mol Biol.* 2019 *26(7):*619-627. doi: 10.1038/s41594-019-0248-4.

Chakrabartty A, Baldwin RL. Stability of alpha-helices. *Adv Protein Chem.* 1995; *46*,141-76.

Chakrabartty, A., Kortemme, T., & Baldwin, R.L. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 1994 *3(5),* 843-52.

Chao HW, Lai YT, Lu YL, Lin CL, Mai W, & Huang YS. NMDAR signaling facilitates the IPO5-mediated nuclear import of CPEB3. *Nucleic Acids Res.* 2012 *40(17),* 8484-98.

Choi, Y. B., Kadakkuzha, B. M., Liu, X. A., Akhmedov, K., Kandel, E. R., & Puthanveettil, S. V. (2014). Huntingtin is critical both pre- and postsynaptically for long-term learning-related synaptic plasticity in Aplysia. *PLoS One, 9(7),* e103004. https://doi.org/10.1371/journal.pone.0103004

Chong, P. A., Vernon, R. M., & Forman-Kay, J. D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. Journal of Molecular Biology, *430(23),* 4650–4665. https://doi.org/10.1016/j.jmb.2018.06.014

Conicella, A. E., Zerze, G. H., Mittal, J., & Fawzi, N. L. (2016). ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure (London, England : 1993), 24(9), 1537–1549. https://doi.org/10.1016/j.str.2016.07.007

Corwin, T., Woodsmith, J., Apelt, F., Fontaine, J. F., Meierhofer, D., Helmuth, J., Grossmann, A., Andrade-Navarro, M. A., Ballif, B. A., & Stelzl, U. (2017). Defining human tyrosine kinase phosphorylation networks using yeast as an *in vivo* model substrate. *Cell Sstems, 5(2*), 128–139.e4. https://doi.org/10.1016/j.cels.2017.08.001

Cougot, N., Bhattacharyya, S. N., Tapia-Arancibia, L., Bordonné, R., Filipowicz, W., Bertrand, E., & Rage, F. (2008). Dendrites of mammalian neurons contain specialized P-body-like structures that respond to neuronal activation. *The Journal of 28*(51), 13793–13804. https://doi.org/10.1523/JNEUROSCI.4155-08.2008

Crick F. (1984). Memory and molecular turnover. *Nature, 312(5990),* 101. https://doi.org/10.1038/312101a0

de Araujo, E. D., Erdogan, F., Neubauer, H. A., Meneksedag-Erol, D., Manaswiyoungkul, P., Eram, M. S., Seo, H. S., Qadree, A. K., Israelian, J., Orlova, A., Suske, T., Pham, H., Boersma, A., Tangermann, S., Kenner, L., Rülicke, T., Dong, A., Ravichandran, M., Brown, P. J., Audette, G. F., … Gunning, P. T. (2019). Structural and functional consequences of the STAT5B^{N642H} driver mutation. *Nature Comm, 10(1)*, 2517. https://doi.org/10.1038/s41467-019-10422-7

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomol NMR*, *6*(3), 277–293. https://doi.org/10.1007/bf00197809

Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., & Rexach, M. (2003). Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proceedings of the National*

Academy of Sciences of the United States of America, 100(5), 2450–2455. https://doi.org/10.1073/pnas.0437902100

Diez-García, F., Gómez-Pinto, I., Chakrabartty, A., González, C., & Laurents, D. V. (2012). Conformation specificity and arene binding in a peptide composed only of Lys, Ile, Ala and Gly. *European Biophysics Journal 41(1*), 63–72. https://doi.org/10.1007/s00249-011-0758-4

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., & Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, *39*(Web Server issue), W13–W17. https://doi.org/10.1093/nar/gkr245

Drisaldi B, Colnaghi L, Fioriti L, Rao N, Myers C, Snyder AM, Metzger DJ,Tarasoff J, Konstantinov E, Fraser PE, Manley JL, Kandel ER. SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep.* 2015. *11(11),* 1694-702. doi: 10.1016/j.celrep.2015.04.061

Eftekharzadeh, B., Piai, A., Chiesa, G., Mungianu, D., García, J., Pierattelli, R., Felli, I. C., & Salvatella, X. (2016). Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract. *Biophysical J.110(11),* 2361–2366. https://doi.org/10.1016/j.bpj.2016.04.022

Escobedo A, Topal B, Kunze MBA, Aranda J, Chiesa G, Mungianu D, Bernardo-Seisdedos G, Eftekharzadeh B, Gairí M, Pierattelli R, Felli IC, Diercks T, Millet O, García J, Orozco M, Crehuet R, Lindorff-Larsen K, Salvatella X. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. *Nat Commun.* 2019, *10(1):* 2034. doi:10.1038/s41467-019-09923-2.

Ferron, F., Rebowski, G., Lee, S. H., & Dominguez, R. (2007). Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP. *The EMBO Journal*, *26*(21), 4597–4606. https://doi.org/10.1038/sj.emboj.7601874

Fioriti, L., Myers, C., Huang, Y. Y., Li, X., Stephan, J. S., Trifilieff, P., Colnaghi, L., Kosmidis, S., Drisaldi, B., Pavlopoulos, E., & Kandel, E. R. (2015). The persistence of hippocampal-based memory requires protein synthesis mediated by the prion-like protein CPEB3. *Neuron, 86(6),* 1433–1448. https://doi.org/10.1016/j.neuron.2015.05.021

Fiumara, F., Fioriti, L., Kandel, E. R., & Hendrickson, W. A. (2010). Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell*, *143*(7), 1121–1135. https://doi.org/10.1016/j.cell.2010.11.042

Fiumara, F., Rajasethupathy, P., Antonov, I., Kosmidis, S., Sossin, W. S., & Kandel, E. R. (2015). MicroRNA-22 gates long-term heterosynaptic plasticity in *Aplysia* through presynaptic regulation of CPEB and downstream targets. *Cell Reports, 11(12),* 1866–1875. https://doi.org/10.1016/j.celrep.2015.05.034

Ford, L., Ling, E., Kandel, E.R., & Fioriti, L. CPEB3 inhibits translation of mRNA targets by localizing them to P bodies. *Proc Natl Acad Sci USA.* 2019 *116(36),* 18078-18087. doi: 10.1073/pnas.1815275116),

Gasset-Rosa, F., Lu, S., Yu, H., Chen, C., Melamed, Z., Guo, L., Shorter, J., Da Cruz, S., & Cleveland, D. W. (2019). Cytoplasmic TDP-43 de-mixing independent of stress granules drives inhibition of nuclear import, Loss of nuclear TDP-43, and cell death. *Neuron, 102(2),* 339–357.e7. https://doi.org/10.1016/j.neuron.2019.02.038

Gill, J., Park, Y., McGinnis, J. P., Perez-Sanchez, C., Blanchette, M., & Si, K. (2017). Regulated intron removal integrates motivational state and experience. *Cell, 169(5),* 836–848.e15. https://doi.org/10.1016/j.cell.2017.05.006

Graether, S. P., Slupsky, C. M., & Sykes, B. D. (2003). Freezing of a fish antifreeze protein results in amyloid fibril formation. *Biophysical J., 84(1),* 552–557. https://doi.org/10.1016/S0006-3495(03)74874-7

Güntert, P., & Buchner, L. (2015) Combined automated NOE assignment and structure calculation with CYANA. J. *Biomol. NMR, 62*, 453-471.

Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Dall'Agnese, A., Hannett, N. M., Spille, J. H., Afeyan, L. K., Zamudio, A. V., Shrinivas, K., Abraham, B. J., Boija, A., Decker, T. M., Rimel, J. K., Fant, C. B., Lee, T. I., Cisse, I. I., Sharp, P. A., Taatjes, D. J., & Young, R. A. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*, *572*(7770), 543–548. https://doi.org/10.1038/s41586-019-1464-0

Hamley, I. W., Kirkham, S., Dehsorkhi, A., Castelletto, V., Adamcik, J., Mezzenga, R., Ruokolainen, J., Mazzuca, C., Gatto, E., Venanzi, M., Placidi, E., Bilalis, P., & Iatrou, H. (2014). Self-assembly of a model peptide incorporating a hexa-histidine sequence attached to an oligo-alanine sequence, and binding to gold NTA/nickel nanoparticles. *Biomacromolecules*, *15*(9), 3412–3420. https://doi.org/10.1021/bm500950c

Hervás, R., Li, L., Majumdar, A., Fernández-Ramírez, M., Unruh, J. R., Slaughter, B. D., Galera-Prat, A., Santana, E., Suzuki, M., Nagai, Y., Bruix, M., Casas-Tintó, S., Menéndez, M., Laurents, D. V., Si, K., & Carrión-Vázquez, M. (2016). Molecular basis of Orb2 amyloidogenesis and blockade of memory consolidation. PLoS biology, 14(1), e1002361. https://doi.org/10.1371/journal.pbio.1002361

Hervas, R., Rau, M. J., Park, Y., Zhang, W., Murzin, A. G., Fitzpatrick, J., Scheres, S., & Si, K. (2020). Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science 367(6483),* 1230–1234. https://doi.org/10.1126/science.aba3526

Hong JY, Wang DD, Xue W, Yue HW, Yang H, Jiang LL, Wang WN, Hu HY. Structural and dynamic studies reveal that the Ala-rich region of ataxin-7 initiates α-helix formation of the polyQ tract but suppresses its aggregation. *Sci Rep.* 2019; *9(1),* 7481. doi: 10.1038/s41598-019-43926-9.

Huang YS, Kan MC, Lin CL, Richter JD. CPEB3 and CPEB4 in neurons: analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J.* 2006 *25(20),* 4865-76.).

Huang, W. H., Chao, H. W., Tsai, L. Y., Chung, M. H., & Huang, Y. S. (2014). Elevated activation of CaMKIIα in the CPEB3-knockout hippocampus impairs a specific form of NMDAR-dependent synaptic depotentiation. *Frontiers in Cellular Neuroscience, 8*, 367. https://doi.org/10.3389/fncel.2014.00367

Hyman S. E. (2005). Addiction: a disease of learning and memory. *The American Journal of Psychiatry*, *162(8),* 1414–1422. https://doi.org/10.1176/appi.ajp.162.8.1414

Jayaraman, M., Kodali, R., Sahoo, B., Thakur, A. K., Mayasundari, A., Mishra, R., Peterson, C. B., & Wetzel, R. (2012). Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *Journal of Molecular Biology*, *415*(5), 881–899. https://doi.org/10.1016/j.jmb.2011.12.010

Kaczmarczyk, L., Labrie-Dion, É., Sehgal, K., Sylvester, M., Skubal, M., Josten, M., Steinhäuser, C., De Koninck, P., & Theis, M. (2016). New phosphospecific antibody reveals isoform-specific phosphorylation of CPEB3 protein. *PloS One, 11(2),* e0150000. https://doi.org/10.1371/journal.pone.0150000

Kandel ER, Dudai Y, Mayford MR (2014) The molecular and systems biology of memory. *Cell, 157:* 163-186

Karplus, M., (1963). Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc. 85(18),* 2870-2871. Doi. 10.1021/ja00901a059.

Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F., & Creamer, T. P. (2001). Host-guest study of left-handed polyproline II helix formation. *Biochemistry, 40(48),* 14376–14383. https://doi.org/10.1021/bi011043a

Khan, M. R., Li, L., Pérez-Sánchez, C., Saraf, A., Florens, L., Slaughter, B. D., Unruh, J. R., & Si, K. (2015). Amyloidogenic oligomerization transforms *Drosophila* Orb2 from a translation repressor to an activator. *Cell, 163(6),* 1468–1483. https://doi.org/10.1016/j.cell.2015.11.020

Kjaergaard, M. & Poulsen, F.M. (2011) Sequence correction of random coil chemical shifts: correlations between neighbor correction factors and changes in the Ramachandran distribution *J. Biomol. NMR 50(2)*, 157-165

Kjaergaard, M., Brander, S. & Poulsen, F.M. (2011) Random coil chemical shifts for intrinsically disordered proteins: Effects of temperature and pH. *J. Biomol. NMR* 49(2),139-49.

Li HR, Chiang WC, Chou PC, Wang WJ, Huang JR. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *J. Biol. Chem.* 2018; *293(16),* 6090–6098. doi:10.1074/jbc.AC117.001037

Lim, L., Wei, Y., Lu, Y., & Song, J. (2016). ALS-Causing Mutations Significantly Perturb the Self-Assembly and interaction with nucleic acid of the intrinsically disordered prion-like domain of TDP-43. *PLoS Biology, 14(1),* e1002338. https://doi.org/10.1371/journal.pbio.1002338

Liu, B. A., Jablonowski, K., Shah, E. E., Engelmann, B. W., Jones, R. B., & Nash, P. D. (2010). SH2 domains recognize contextual peptide sequence information to determine selectivity. *Molecular & Cellular Proteomics: 9(11)* 2391–2404. https://doi.org/10.1074/mcp.M110.001586

López-Alonso, J. P., Bruix, M., Font, J., Ribó, M., Vilanova, M., Jiménez, M. A., Santoro, J., González, C., & Laurents, D. V. (2010). NMR spectroscopy reveals that RNase A is chiefly denatured in 40% acetic acid: implications for oligomer formation by 3D domain swapping. *Journal of the American Chemical Society, 132(5),* 1621–1630. https://doi.org/10.1021/ja9081638

Mackenzie, I. R., Rademakers, R., & Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet. Neurology*, *9*(*10),* 995–1007. https://doi.org/10.1016/S1474-4422(10)70195-2

Majumdar, A., Cesario, W.C., White-Grindley, E., Jiang, H., Ren, R, Khan, M. R., Li. L., Choi, E. M. K., Kannan, K., Guo, F., Unruh, J., Slaughter, B., & Si, K. Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. *Cell. 2012;148(3),* 515–529. doi:10.1016/j.cell.2012.01.004

Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., & Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *European Journal of Biochemistry, 256(1)* , 1–15. https://doi.org/10.1046/j.1432-1327.1998.2560001.x

Marley, J., Lu, M., & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, *20*(1), 71–75. https://doi.org/10.1023/a:1011254402785

Metzler, W. J., Bell, A. J., Ernst, E., Lavoie, T. B., & Mueller, L. (1994). Identification of the poly-L-proline-binding site on human profilin. *The Journal of biological chemistry*, *269*(6), 4620–4625.

Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E., Partridge, L., & Isaacs, A. M. (2014). C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science 345(6201)*, 1192–1194. https://doi.org/10.1126/science.1256800

Morón-Oset, J., Supèr, T., Esser, J., Isaacs, A. M., Grönke, S., & Partridge, L. (2019). Glycine-alanine dipeptide repeats spread rapidly in a repeat length- and age-dependent manner in the fly brain. *Acta Neuropathologica Communications, 7(1)* , 209. https://doi.org/10.1186/s40478-019-0860-x

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., & Zweckstetter, M. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol. 2009 7(2):* e34. doi:10.1371/journal.pbio.1000034.

Muñoz, V., Serrano, L., Jiménez, M. A., & Rico, M. (1995). Structural analysis of peptides encompassing all alpha-helices of three alpha/beta parallel proteins: Che-Y, flavodoxin and P21-ras: implications for alpha-helix stability and the folding of alpha/beta parallel proteins. *Journal of Molecular Biology, 247(4)*, 648–669. https://doi.org/10.1016/s0022-2836(05)80145-7

Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T. D., & Waugh, D. S. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*. *Protein expression and purification, 38(1),* 108–115. https://doi.org/10.1016/j.pep.2004.08.016

Nelson, P. T., Dickson, D. W., Trojanowski, J. Q., Jack, C. R., Boyle, P. A., Arfanakis, K., Rademakers, R., Alafuzoff, I., Attems, J., Brayne, C., Coyle-Gilchrist, I., Chui, H. C., Fardo, D. W., Flanagan, M. E., Halliday, G., Hokkanen, S., Hunter, S., Jicha, G. A., Katsumata, Y., Kawas, C. H., … Schneider, J. A. (2019). Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. Brain : A *Journal of Neurology, 142(6),* 1503–1527. https://doi.org/10.1093/brain/awz099

Oakley, M.G., & Kim, P.S. A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry.* 1998, *37(36),* 12603–12610. doi:10.1021/bi981269m

Pantoja-Uceda, D., & Santoro, J. (2013). Direct correlation of consecutive C'-N groups in proteins: a method for the assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR*, *57(1),* 57–63. https://doi.org/10.1007/s10858-013-9765-3

Pantoja-Uceda, D., & Santoro, J. (2014). New ^{13}C-detected experiments for the assignment of intrinsically disordered proteins. *Journal of Biomolecular NMR, 59(1),* 43–50. https://doi.org/10.1007/s10858-014-9827-1

Pelassa, I., Corà, D., Cesano, F., Monje, F. J., Montarolo, P. G., & Fiumara, F. (2014). Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Human Molecular Genetics, 23(13),* 3402–3420. https://doi.org/10.1093/hmg/ddu049

Peng, S.C., Lai, Y.T., Huang, H.Y., Huang, H. D., & Huang, Y.S., A novel role of CPEB3 in regulating EGFR gene transcription via association with Stat5b in neurons. *Nucleic Acids Res.* 2010 *38(21):* 7446-57. doi: 10.1093/nar/gkq634.).

Pentelute, B. L., Gates, Z. P., Tereshko, V., Dashnau, J. L., Vanderkooi, J. M., Kossiakoff, A. A., & Kent, S. B. (2008). X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *Journal of the American Chemical Society*, *130*(30), 9695–9701. https://doi.org/10.1021/ja8013538

Polling, S., Ormsby, A. R., Wood, R. J., Lee, K., Shoubridge, C., Hughes, J. N., Thomas, P. Q., Griffin, M. D., Hill, A. F., Bowden, Q., Böcking, T., & Hatters, D. M. (2015). Polyalanine expansions drive a shift into α-helical clusters without amyloid-fibril formation. *Nature structural & molecular biology*, *22*(12), 1008–1015. https://doi.org/10.1038/nsmb.3127

Ramírez de Mingo D, López-García P, Laurents DV, Carrión-Vázquez M (2020) "Molecular forces governing hCPEB3 liquid demixing and amyloidogenesis". Submitted.

Raveendra, B. L., Siemer, A. B., Puthanveettil, S. V., Hendrickson, W. A., Kandel, E. R., & McDermott, A. E. (2013). Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. *Nature Structural & Molecular Biology*, *20*(4), 495–501. https://doi.org/10.1038/nsmb.2503

Salehi-Ashtiani, K., Lupták, A., Litovchick, A., & Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science 313(5794),* 1788–1792. https://doi.org/10.1126/science.1129308

Si, K., Lindquist, S., & Kandel, E.R. A neuronal isoform of the *Aplysia* CPEB has prion-like properties. *Cell.* 2003 *115(7),* 879-91

Si, K., & Kandel, E.R. The role of functional prion-like proteins in the persistence of memory. *Cold Spring Harb Perspect Biol.* 2016 *8(4),* a021774. doi: 10.1101/cshperspect.a021774

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein

multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Spera, S., & Bax, A. (1991) Empirical corrleation between protein backbone conformation and Calpha and Cbeta 13C NMR chemical shifts. *J. Am. Chem. Soc. 113(14):* 5490-5492. doi.10.1021/ja00014a071

Stephan, J.S., Fioriti, L., Lamba, N., Colnaghi, L., Karl, K., Derkatch, I.L., & Kandel, ER, The CPEB3 Protein Is a Functional Prion that Interacts with the Actin Cytoskeleton. *Cell Rep.* 2015, *11(11),* 1772-1785.

Sun, Z. Y., Frueh, D. P., Selenko, P., Hoch, J. C., & Wagner, G. (2005). Fast assignment of 15N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. *Journal of Biomolecular NMR*, *33*(1), 43–50. https://doi.org/10.1007/s10858-005-1284-4

Terlau, H., & Seifert, W. (1990). Fibroblast growth factor enhances long-term potentiation in the hippocampal slice. *The European Journal of Neuroscience*, 2(11), 973–977. https://doi.org/10.1111/j.1460-9568.1990.tb00009.x

Treviño, M. Á., Pantoja-Uceda, D., Menéndez, M., Gomez, M. V., Mompeán, M., & Laurents, D. V. (2018). The singular NMR fingerprint of a polyproline II helical bundle. *Journal of the American Chemical Society, 140(49),* 16988–17000. https://doi.org/10.1021/jacs.8b05261

Tsuda, K., Kuwasako, K., Nagata, T., Takahashi, M., Kigawa, T., Kobayashi, N., Güntert, P., Shirouzu, M., Yokoyama, S., Muto, Y. Novel RNA recognition motif domain in the cytoplasmic polyadenylation element binding protein 3. *Proteins* 2014 *82(10),* 2879-86. doi: 10.1002/prot.24651.

Vogler, C., Spalek, K., Aerni, A., Demougin, P., Müller, A., Huynh, K.D., Papassotiropoulos, A., de Quervain, D.J. CPEB3 is associated with human episodic memory. *Front Behav Neurosci.* 2009 *3:4.* doi: 10.3389/neuro.08.004.2009

Vuister, G. W., Bax, A., Quantitative J correlation: a new approach for measuring homonuclear J(HN-Ha) coupling constants in 15N-enriched proteins. *J. Am. Chem. Soc.* 1993, *115*, 7772-7777.

Wang, Y., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical shift data. *Prot. Sci. 11*: 842-861

Wishart, D. S., & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *Journal of Biomolecular NMR*, *4*(2), 171–180. https://doi.org/10.1007/bf00175245

Xu, D., Farmer, A., Collett, G., Grishin, N. V., & Chook, Y. M. (2012). Sequence and structural analyses of nuclear export signals in the NESdb database. *Molecular Biology of the Cell*, *23*(18), 3677–3693. https://doi.org/10.1091/mbc.E12-01-0046

Table 1: Sequence of the disordered N-terminal region of hCPEB3

```
1         10          20          30          40          50
MQDDLLMDKS KTQPQPQQQQ RQQQQPQPES SVSEAPSTPL SSETPKPEEN
          60          70          80          90         100
SAVPALSPAA APPAPNGPDK MQMESPLLPG LSFHQPPQQP PPPQEPAAPG
         110         120         130         140         150
ASLSPSFGST WSTGTTNAVE DSFFQGITPV NGTMLFQNFP HHVNPVFGGT
         160         170         180         190         200
FSPQIGLAQT QHHQQPPPPA PAPQPAQPAQ PPQAQPPQQR RSPASPSQAP
         210         220         230         240         250
YAQRSAAAAY GHQPIMTSKP SSSSAVAAAA AAAAASSASS SWNTHQSVNA
         260         270         280    *    29*        **300
AWSAPSNPWG GLQAGRDPRR AVGVGVGVGV GVPSPLNPIS PLKKPFSSNV
         ^310        320         330         340 ‡       350
IAPPKFPRAA PLTSKSWMED NAFRTDNGNN LLPFQDRSRP YDTFNLHSLE
#        360         370         380         390        400*
NSLMDMIRTD HEPLKGRMGI NFHHPGTDNI MALNSRSSLF PFEDAFLDDS
       **410**       420        §430         440         450
HGDQALSSGL SSPTRCQNGE RVERYSRKVF VGGLPPDIDE DEITASFRRF
```

Segment 1 spans hCPEB3 residues 1-100 and is colored purple residues;
segment 2, residues 51 – 150 (double underlined residues);
segment 3, residues 101 – 200 (blue residues);
segment 4, residues 151 – 250 (dashed underlined residues);
segment 5, residues 201 – 300 (green residues);
segment 6, residues 251 – 350, (wavy underlined residues);
segment 7, residues 302 – 400 (orange residues) and finally
segment 8, which is composed of residues 352 – 450 (single underlined residues).
Residues of structural or functional interest are in **bold font**.
^The putative dimethyl-Arg site (R308).
‡The putative phosphoTyr site, Y341 putatively recognized by STAT5B's SH2 domain.
#The nuclear export signal, L349-L353
*Putative phosphorylation sites: S284, S290, S297, S298, S400, S407, S408, S411 and S412.
§ Residues belonging to the first RRM1 domain are written in italics.

Table 2: Hydrophobic helices and Pro rich stretches modulate amyloid formation by Q-, Q/N-rich segments.

PROTEIN	SEQUENCE*
Huntingtin	M_{001}ATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQPP PPPPPPPPPQ LPQPPPQAQP
ANDROGEN RCPT	H_{041}PEAASAAPP GASLLLLQQQ QQQQQQQQQQ QQQQQQQQQQ ETSPRQQQQQ QGEDGSPQAH
TDP-43	P_{319}AMMAAAQAA LQSSWGMMGM LASQQNQSGP SGNNQNQGNM QREPNQAFGS FNNSYSSINI
Ataxin-7	A_{016}AAAAGGAAA AAARQQQQQQ QQQQPPPPQP QRQQHPPPPP RRTRPEDGGP GAASTSAAAM
hCPEB3	M_{001}QDDLLMDKS KTQPQPQQQQ QRQQQQPQPE SSVSEAPSTP LSSETPKPEE NSALPALSPA
Drosophila Orb2A	M_{001}YNKFVNFIC GGLPNLNLNK PPQLHQQQHQ QQHQQHQQHQ QQQQLHQHQQ QLSPNLSALH
Aplysia CPEB	F_{030}EIFRRINAL LDNSLEANNV SCSQSQSQQQ QQTQQQQQQQ QQQQQQQMLQ QVQQQQLLKQ

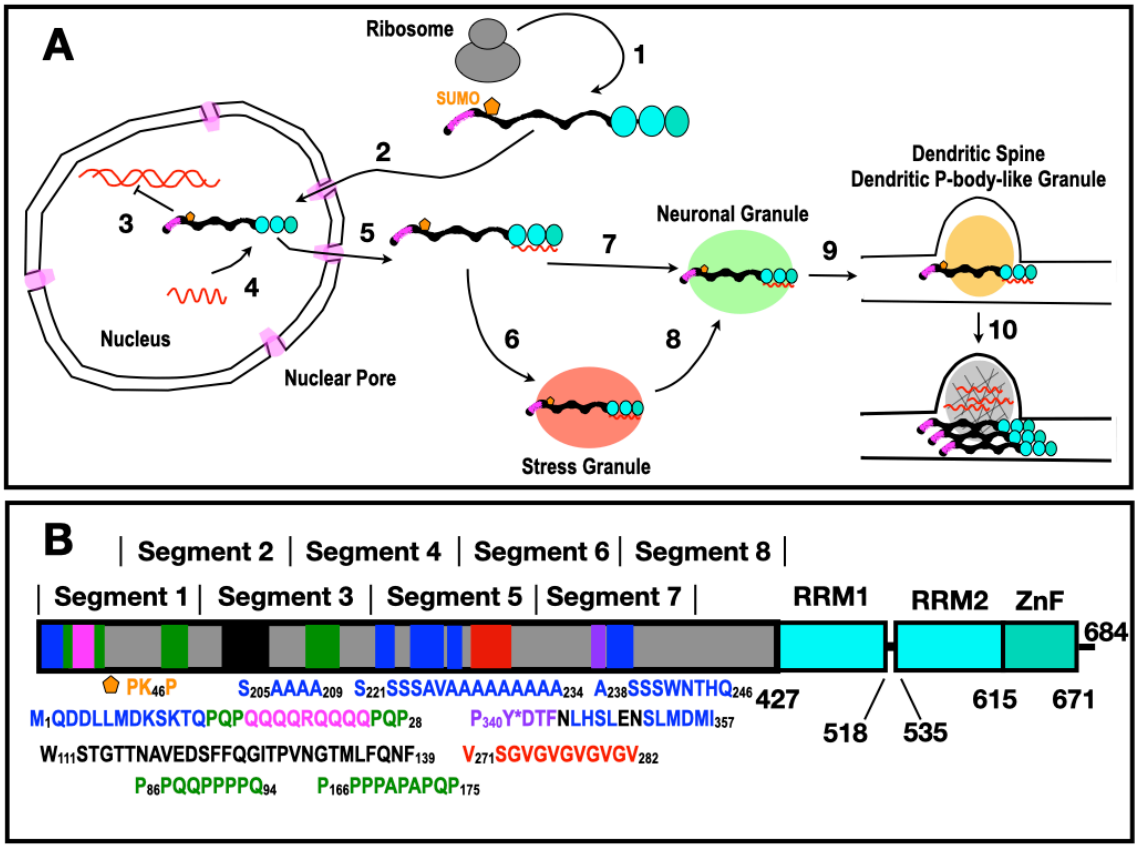
* In these sequences, nonpolar residues are shaded gray, Gln and Asn residues are shaded magenta and Prolines are shaded green.

Table 3:
Conformational Chemical Shifts for α-helices, β-strands and PPII helices.

	α-helix	β-strand	Gly-rich PPII helical bundle	Isolated Pro-rich PPII helix
$\delta\Delta\ ^{13}C\alpha$	3.1[a] / 2.9[b]	-1.5[a] / -1.8 [b]	-0.6[d]	+0.6[e]
$\delta\Delta\ ^{13}C\beta$	-0.4[a] / -0.8[b]	+2.2[a] / +2.7[b]	+0.3[d]	-1.0[e]
$\delta\Delta\ ^{13}CO$	2.2[c] / 2.2[b]	-2.2[c] / -2.1[b]	-0.2[d]	-0.3[e]

a. from Spera & Bax (1991)
b. from Wang and Jardetzky (2002) for alanine.
c. calculated from Wishart & Skyes (1994)
d. from Treviño *et al.* (2018)
e. This study.

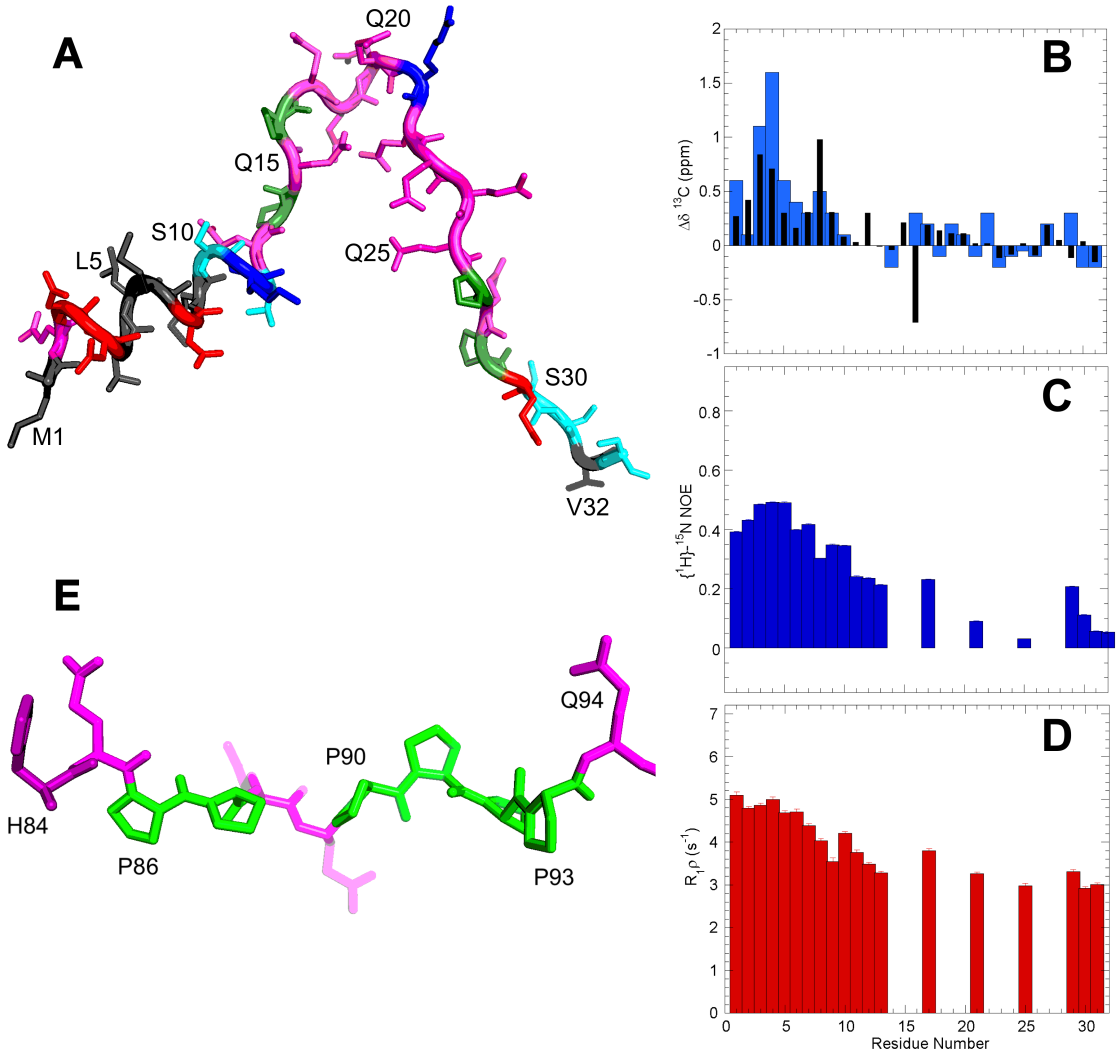
Figure 1. Domain Structure and Activities of CPEB3



A. **CPEB3 is present in multiple cellular compartments.** Dendritic stimulation leads to temporary, phosphorylation-mediated short term memory and **1**. increased synthesis of the protein CPEB3. Composed of an N-terminal disordered region (**black**) which includes a Q-rich segment necessary for functional aggregation (**magenta**), hCPEB3 also contains RRM domains (cyan) and a ZZ-Zinc finger domain (**turquoise**). **2**. Upon continued neuro-stimulation, CPEB3 enters the nucleus through the nuclear pore (light magenta), which is a macromolecular condensate. Once in the nucleus, CPEB3 indirectly regulates transcription through STAT5B **3,** and binds to certain mRNAs **4** (**red**). This binding suppresses translation. After exiting the nucleus through the nuclear pore, **6** CPEB3+mRNA may associate with a stress granule (**rose**) during moments of cellular stress. In the absence of stress (**7**) or its passing (**8**), CPEB3+mRNA will combine with another condensate called neuronal granules (**light green**) for transport to dendritic spines (**9**), where CPEB3+mRNA associate with still another class of condensate called a dendritic P-body-like structure (**golden**) (Cougot *et al.,* 2008). Further neuronal stimulation (**10**) causes synapse-specific deSUMOlyation, CPEB3 aggregation and translational activation of previously repressed mRNA, leading to morphological changes and fortification of the spine, which is proposed to be the basis of long-term memory. This is a simplified model based on (Kandel *et al.,* 2014).

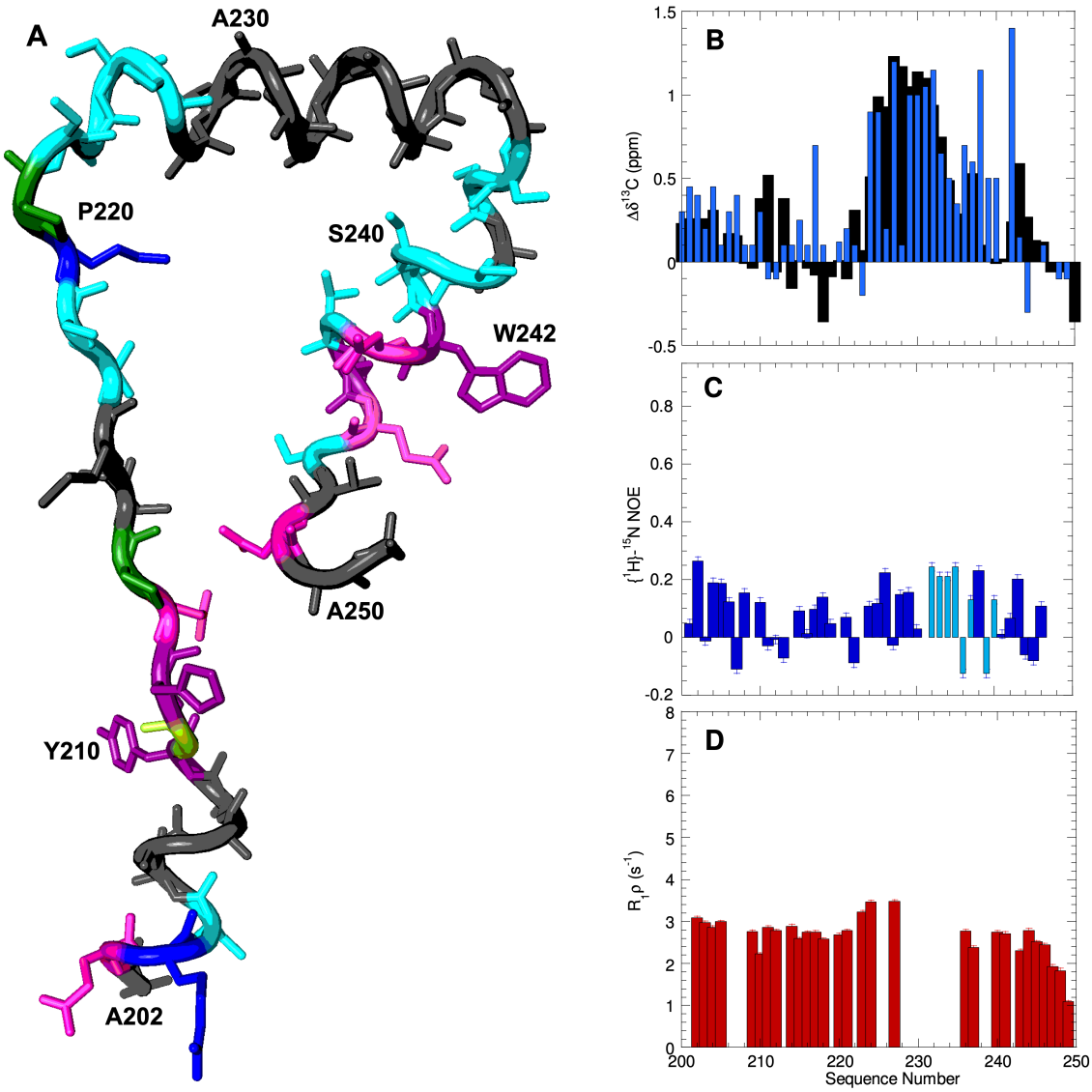
B. CPEB3 domain composition and its N-terminal intrinsically disordered domain (**gray**) contains key elements with preferred conformers colored **blue** for α-helix, **magenta** for polar amyloidogenic, **black** for hydrophobic amyloidogenic, **green** for PPII helix, **purple** for the putative phosphoTyr site, and **red** for highly disordered segments. The two RRM domain are colored cyan and the C-terminal Zinc Finger is shown in **turquoise**.

Figure 2: The N-terminal 25 residues of hCPEB3 adopt a hydrophobic α-helix followed by a disordered polyQ segment flanked by PQP mini-breaker motifs.



The N-terminus of hCPEB3 contain helix-forming and an amyloidogeneic Q_4RQ_4 segments, which are separated by PQP mini breaker motifs. A representative conformers (**A**) of the first 32 residues from a family of 20 structures calculated using the program CYANA based on TALOS+ backbone angle restrictions derived from chemical shift data. Nonpolar residues are colored **black**, aromatics=**purple**, Q/N=**magenta**, E/D=**red**, K/R=**blue**, G=**green**, C,S,T=**cyan**. Note that whereas this conformer shows an α-helix at the N-terminus, fully disordered species dominant the conformation ensemble, as the (**B**) ^{13}Cα (**blue**) and ^{13}CO (**black**) conformational chemical shifts indicate a 15-20% population of helix at 25 ºC. (**C**) {^{1}H}-^{15}N NOE and (**D**) $R_1\rho$ relaxation measurements indicate that this helical conformation is less mobile than the polyQ segment at ns/ps and μs/ms timescales, respectively at 25 ºC. Error bars are shown in panels (**C**) and (**D**) but are small as the estimated uncertainties are <0.01 for the hNOE and < 0.1 s^{-1} for $R_1\rho$. Missing values in panels **C** and **D** are due to overlap of $^1H^{15}N$ peaks or a lack of $^1H^{15}N$ signals in the case of proline residues. (**E**) Representative structure, from a family of 20 CYANA conformers, of the proline rich segment, H84-Q94, featuring a PPII helix that spans residues P86-Q94. Residues are colored as in panel **A**.

Figure 3. Residues 201 – 250 adopt three α-helices.



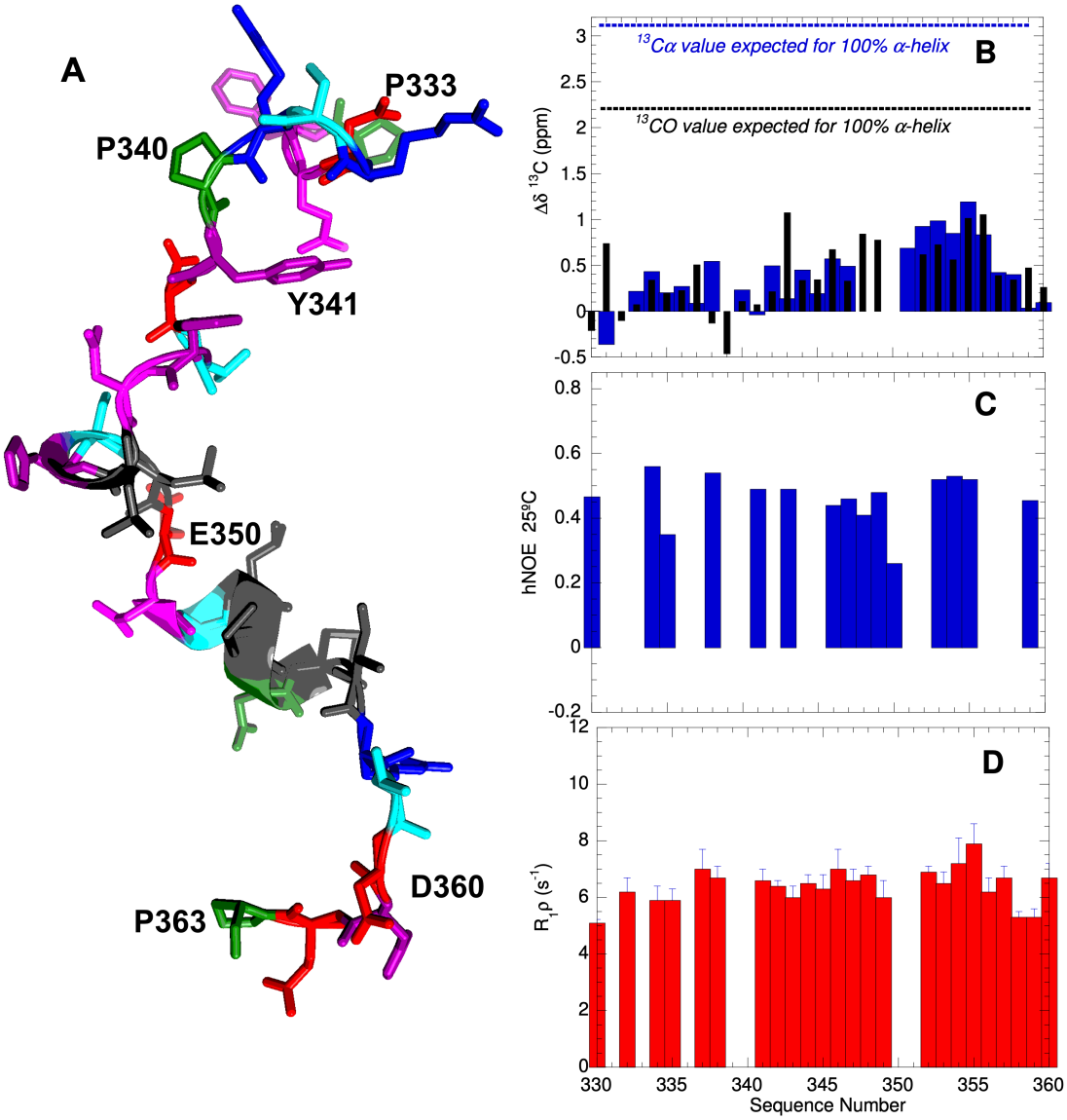
A. Representative conformer of residues 201-250, showing three alpha helices. As in **Figure 2**. Nonpolar residues are colored **black**, aromatics=**purple**, Q/N=magenta, E/D=red, K/R=blue, G=green, C,S,T=cyan.

B. ^{13}CO (black) and ^{13}Cα (blue) conformational chemical shifts of residues 201-250 at 25ºC. Note that the second α-helix which contains nine consecutive Ala residues has a relatively high helical population.

C. {^{1}H}-^{15}N NOE ratios. Values shown in dark blue are of individual ^{1}H^{15}N resonances; those in light blue correspond to overlapped peaks.

D. $R_{1\rho}$ values reveal the ps/ns and ms/ms time scales. Significantly higher {^{1}H}-^{15}N NOE ratios and $R_{1\rho}$ values are observed for these residues at 5 ºC (**Sup. Fig. 8**).

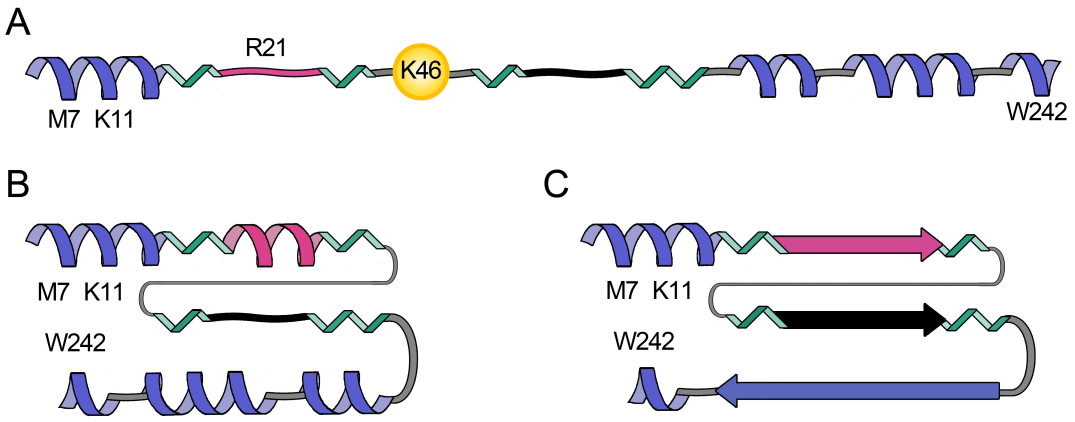
Figure 4. Conformation of the NES and nearby phospho-Tyr site.



A. Representative conformer of P_{333}FQDRSRP_{340}YDTFNLHSLE_{350}NSLMDMIRTD_{360}HEP_{363} shown with cationic residues (R & K) colored blue, aromatics (F, Y & H) = purple, anionic (E & D) = red, aliphatic (I, L, M) = dark gray, amyloidogenic (N & Q) = magenta, hydroxyl bearing (S & T) = cyan, and proline = green. Spiral ribbons mark the helical segments spanning residues 346 – 349 and 352 - 356. **B**. Conformational chemical shifts of ^{13}Cα (blue bars) and ^{13}CO (black narrow bars) afford detection of helical conformations. **C.** {^{1}H}-^{15}N NOE ratios of 0.85 and -0.20 are indicative of high rigidity and flexibility, respectively, on ps-ns time scales. **D**. higher R_2 rates are diagnostic of rigidity on µs-ms time scales.

Figure 5

Working Hypothesis for hCPEB3 Structural Changes During Memory Consolidation.



A. The first 250 residues of hCPEB3 contain 4 α-helices (blue spirals), the first and third α-helices are relatively stable. Proline-rich segments (green) and SUMOylation as putatively to occur at K46 by *in silicio* methods (Ramírez de Mingo *et al.,* 2020) prevent premature association and amyloid formation by the Q_4RQ_4 segment (magenta) and the hydrophobic motif (**black** squiggle).
B. Following deSUMOlyation and putatively phosphorylation, association between the fourth and first helices could occur, strengthened by hydrophobic and cation-π interactions. The Q_4RQ_4 segment may adopt an α-helix and associate with the Ala rich α-helices to form a coiled-coil. The structural transformations may well enhance intermolecular contacts within the dendritic P-body like granule, leading to gelification and eventually, amyloid formation.
C. The final amyloid could be composed of the Q_4RQ_4 segment and hydrophobic tract, and possibly the Ala-rich segments. The final configuration of the polyPro segments (**green**) would promote profilin binding and the initiation of a more robust actin filament network.

Sup. Table 1: NMR Spectral Parameters

Experiment	Number of Scans	Sweep Width (ppm)	Matrix
Segments 1, 3, 4, 5, 6, 7 & 8			
1D ^{1}H	8	10	32k
2D ^{1}H-^{15}N HSQC	2-8	10 ^{1}H x 20 ^{15}N	2k x 512
2D CON*	4–16	10 ^{13}CO x 35 ^{15}N	1k x 512
3D HNCO	4–8	10 ^{1}H x 20 ^{15}N x 10 ^{13}C	2k x 64 x 128
3D hacacoNcaNCO*	8	10 ^{13}CO x 35 ^{15}N x 20 ^{15}N	1k x 48 x 96
3D hacaCOncaNCO*	8	10 ^{13}CO x 35 ^{15}N x 20 ^{13}CO	1k x 48 x 96
3D CCCON*		10 ^{13}CO x 35 ^{15}N x 60 ^{13}C(aliphatic)	1k x 56 x 96
3D CBCACON	8	10 ^{13}CO x 35 ^{15}N x 60 ^{13}C(aliphatic)	1k x 64 x 80
3D HNHA§	8	11^{1}HN x 23 ^{15}N x 11 ^{1}Hα	2k x 96 x 48
Segments 3, 7 & 8†			
3D HNcaCO	16	10 ^{1}H x 20 ^{15}N x 10 ^{13}C	2k x 32 x 64
3D hNcocaNH	12	10 ^{1}H x 20 ^{15}N x 20 ^{15}N	2k x 64 x 90
3D HncocaNH	12	10 ^{13}CO x 20 ^{15}N x 10 ^{13}CO	2k x 96 x 48
3D CBCAcoNH	16	10 ^{1}H x 20 ^{15}N x 10 ^{13}C	2k x 32 x 64
Relaxation			
2D {^{1}H}-^{15}N NOE¶	8-16	10 ^{1}H x 20 ^{15}N	2k x 512
2D ^{1}H-^{15}N $T_1\rho$‡	8-16	10 ^{1}H x 20 ^{15}N	2k x 256
Peptides 1 & 2			
2D ^{1}H-^{1}H COSY	16	11 x 11	2k x 512
2D ^{1}H-^{1}H TOCSY	8	11 x 11, t_{mix} = 60 ms	2k x 512
2D ^{1}H-^{1}H NOESY	64	11 x 11, t_{mix}=150 ms	2k x 512
2D ^{1}H-^{13}C HSQC	144	10 x 80, recorded in 100% D_2O	1k x 128
2D ^{1}H-^{15}N HSQC	512	10 x 20	1k x 96

* ^{13}C detection; processed with in-phase anti-phase (IPAP) virtual decoupling.

§ Recorded for some segments to measure ^{1}HN-^{1}Hα coupling constants.

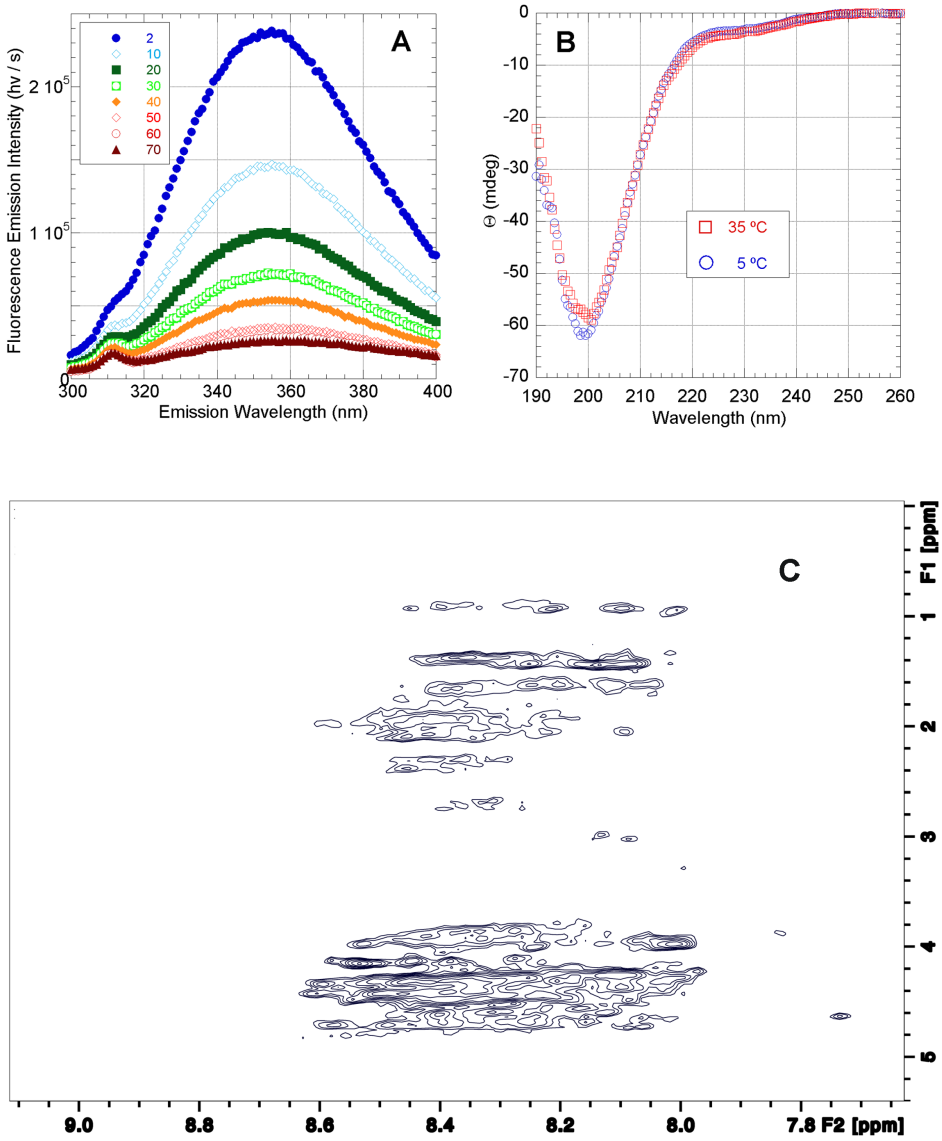
† Recorded for segments 3, 7 and 8, whose lower sample concentration required more sensitive ^{1}H-detected experiments.

¶ A delay of 10 seconds between pulses was used. Recorded in interleaved mode.

‡ Ten spectra with delays of 8, 300, 36, 76, 900, 100, 500, 156, 200 and 16 ms were recorded.

Sup. Fig. 1

Biophysical Characterization of the Complete hCPEB3 IDR



A. Fluorescence emission spectra of the complete IDR of hCPEB3 at pH 4.0 in 1.0 mM deuterated acetic acid buffer over temperatures ranging from 2 ºC to 70 ºC as indicated. The emission maximum over 350 nm is indicative of solvent exposed Trp side chains. This is consistent with the lack of a well packed hydrophobic core.

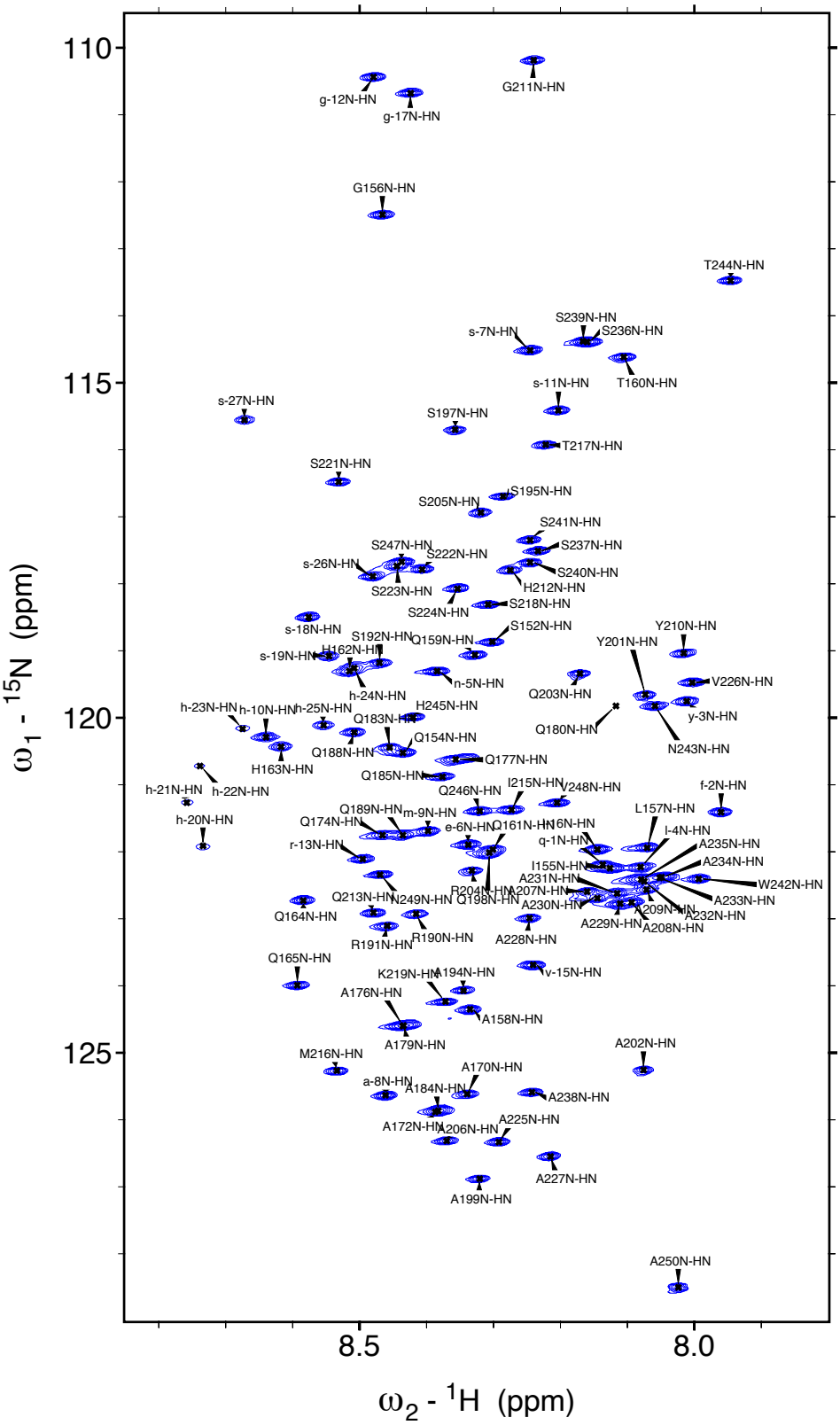
B. Far UV-CD spectra of the IDR of hCPEB3 at 5 ºC (blue open circles) and 35 ºC (red open squares). The observed minimum near 200 nm and the lack of minima near 208 nm, 218 nm and 222 nm and the lack of a maximum at 195 nm are all characteristic spectral features of a statistical coil.

C. 2D ^{1}H-^{1}H NOESY spectrum of the hCPEB3 IDR recorded at 25 ºC showing the ^{1}HN crosspeak region. The small chemical shift dispersion in ^{1}HN (8.6 to 7.9 ppm) is a typical feature of disordered and unfolded proteins (López-Alonso *et al.,* 2010).

Sup. Fig. 2: Sequence Alignments of CPEB3 from Representative Vertebrates

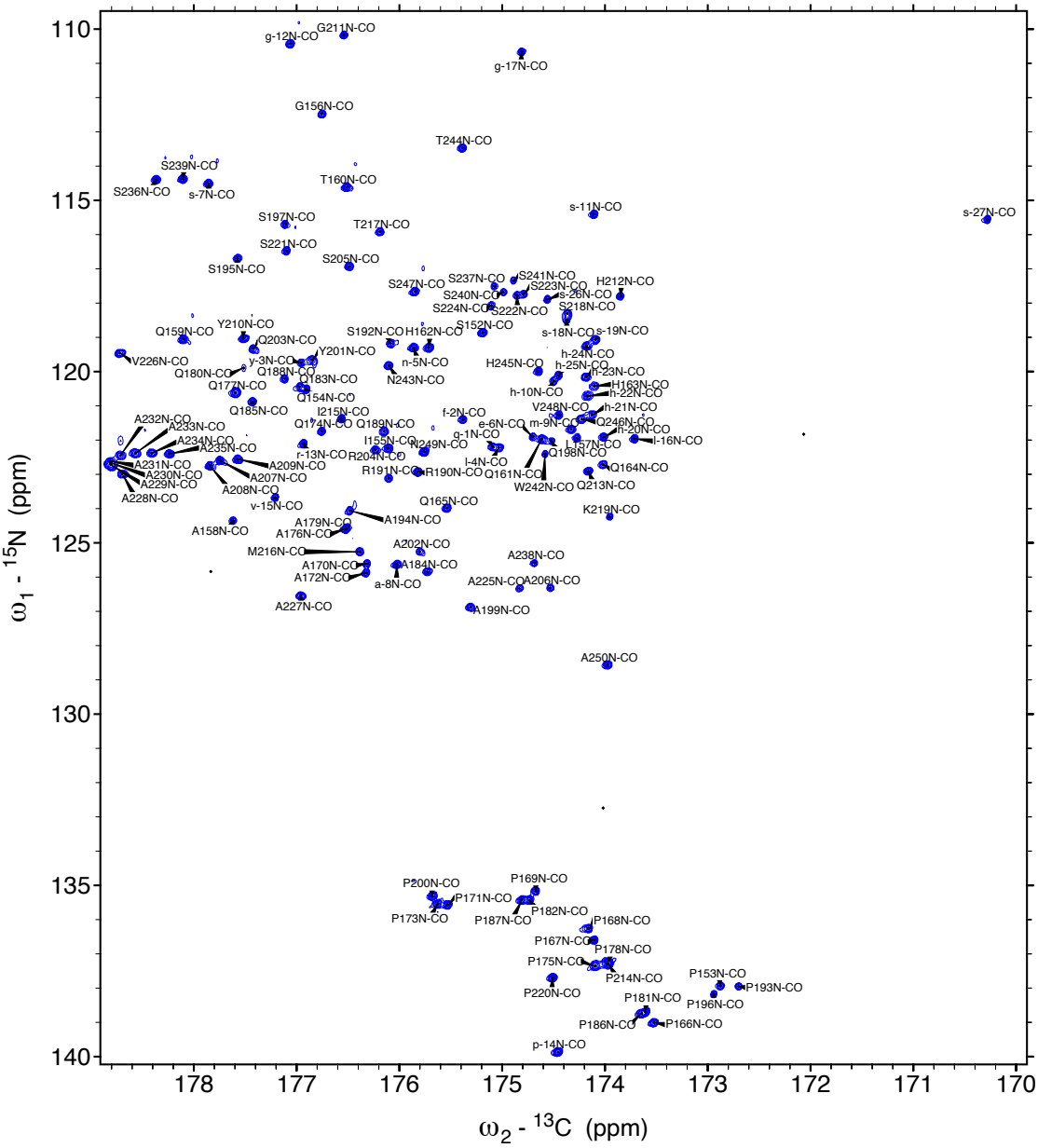
```
Human2    MODDLLMDKSKTOPOPOOOO----ROO----OOPOPESSVSEAPSTPLSSETPKPEENSAVPALSPAAAPPA
Human1    MODDLLMDKSKTOPOPOOOO----ROO----OOPOPESSVSEAPSTPLSSETPKPEENSAVPALSPAAAPPA
Mouse     MODDLLMDKSKTOPOSOOOO----ROQQQQQOOLQPEPGAAEAPSTPLSSEIPKPEDSSAVPALSPASAPPA
Chicken   MODDLLMDKSKAOPOOOROO----QQ-----OOO--DPNSAEAPSTPISSETPKPEDNSSVTSIGT--SAPS
Turtle    MODDLLMDKSKAOPOOOROO-----------OOO--DPNAAEPPSTPISSETSKPEDSSPVTSIGT--ATSV
Frog      MODDLLMDKSKAOOROOPQQPPSSQTQ----QQQKEAASVAEPPSSRESS---------------------P
Fish      MQDDLLMDKSKAQPHPHQ-----------------DPSNTEPPSTPTPSEPPILDEQHHV-----------

cons      ************:* : :  ▇  ▇  ▇  ▇  ▇  ▇  .. :*.**:  .*

Human2    PNGPDKMOMESPLLPGLSFHOPPOOPPPPOEPAAPGASLSPSFGSTWSTGTTN-AVEDSFFOGITPVNGTML
Human1    PNGPDKMOMESPLLPGLSFHOPPOOPPPPOEPAAPGASLSPSFGSTWSTGTTN-AVEDSFFOGITPVNGTML
Mouse     PNGPDKMOMESPLLPGLSFHOPPOOPPPPQEPTAPGASLSPSFGSTWSTGTTN-AVEDSFFOGITPVNGTML
Chicken   ONSKEKMOMESPILPGLSFHOPOOEP-------AAGTSLSPSFGSTWSTGTTN-TVDDSFFOGITPVNGTML
Turtle    ONSKEKMOMESPILPGLSFHOPOQEP-------TPGTSLSPSFGSTWSTGTTN-TVDDSFFOGITPVNGTML
Frog      PTHKDKMOMESPLLPGLSFOOEP----------PTTPSLSPSFGSTWSTGGSNSAVDDSFFPGITPVNGTML
Fish      -GHSERLAMESPVLPGLGFPQD--------------SSLSPSFGSTWST---N-TVDEGFFQGIPSVNGTML

cons        ::: ****:****.* *       ▇  ▇  .************  * :*::.** **..******

Human2    FONFPH--HVNPVFGGTFSPOIGLAOTOHHOOPPPPAPAPOPAOPAOPPOAOPPOORRSPASPSO-APYAOR
Human1    FONFPH--HVNPVFGGTFSPOIGLAOTOHHOOPPPPAPAPOPAOPAOPPOAOPPOORRSPASPSO-APYAOR
Mouse     FONFPH--HVNPVFGGTFSPOIGLAOTOHHOOPPPP--APQPPQPAOPPOAOPSOORRSPASPSO-APYAOR
Chicken   FONFPH--HVNPVFGGTFSPOIGLAOTOHHOH------------QQOOOOAOOOOORRSPVSPNO-APFAOR
Turtle    FONFPH--HVNPVFGGTFSPOIGLAOTOHHHH--------------QQQQAQOOOORRSPVSPNO-APFAOR
Frog      FONFPHHHHVNPVFGGTFSPOMGLAHQT-----------------------QQQQRRSPASPNNHTAYTQR
Fish      FQNFPH--HVNPVYGGNFSPQM--------------------------------RRSPLGVNQ------R

cons      ******  *****:**.****:  ▇  ▇  ▇  ▇  ▇  ▇
```

Sup. Fig. 3 2D ^{1}H-^{15}N HSQC NMR Spectrum of hCPEB3, Segment 4



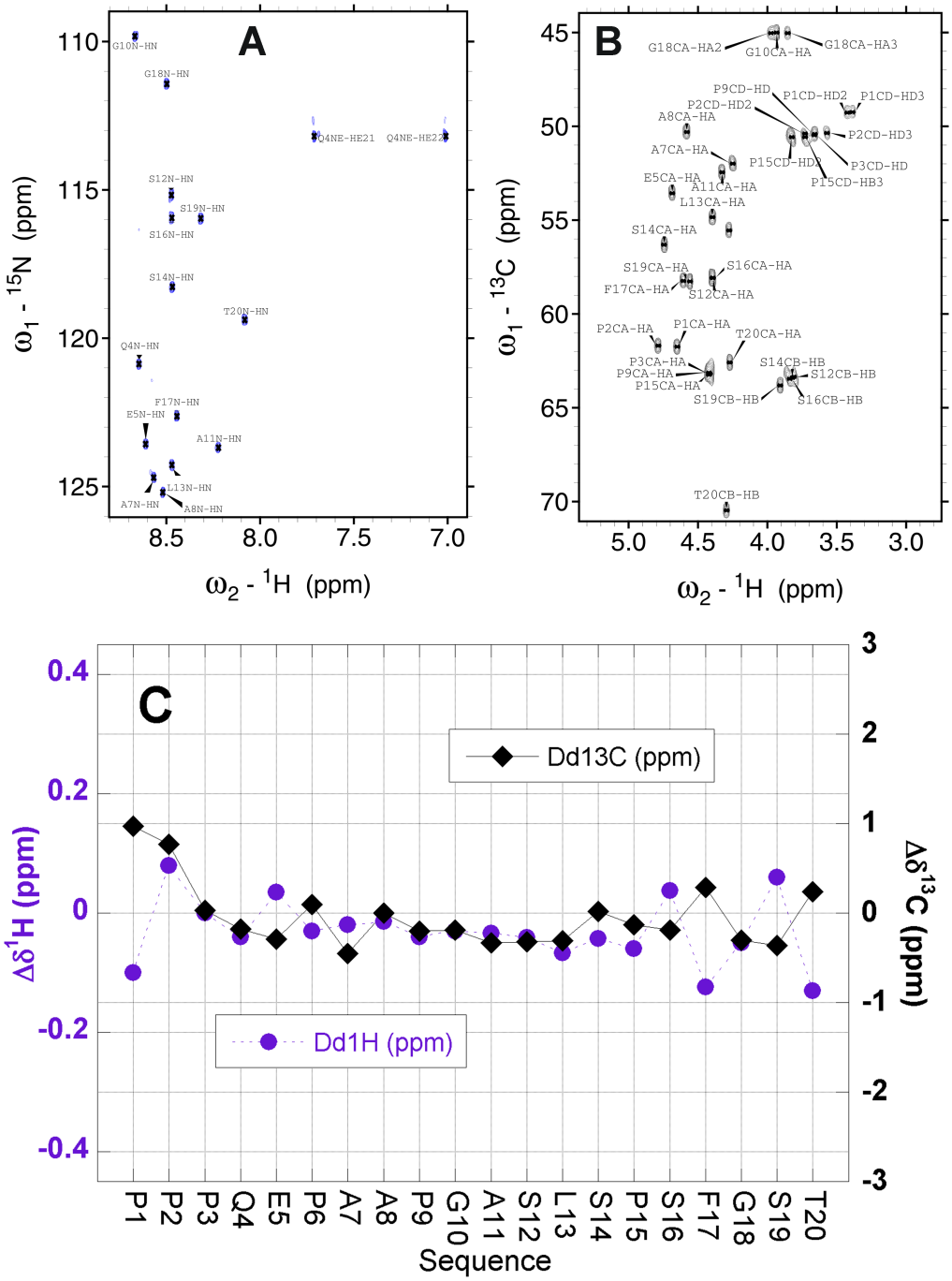
^{1}H-^{15}N correlations were recorded at 25ºC in 1 mM acetic acid, pH 4. Note the cluster of Ala signals in belonging to the polyA stretches near 8.05 ppm ^{1}H and 122.5 ppm ^{15}N. Their position contrasts with those of the isolated Ala residues, whose δ ^{15}N > 125 ppm. Signals labeled in lower case with negative numbers correspond the His/TEV tag: M_{-29}GSSHHHHHHSSGLVPRGSHMASENLYFQ$_{-1}$

Sup. Fig. 4 2D ^{13}CO-^{15}N NMR Spectrum of hCPEB3



^{13}CO-^{15}N correlations were recorded at 25ºC in 1 mM acetic acid, pH 4. Here, the signals of the long polyAla stretch, which appears around 122.5 ppm for ^{15}N x 178.5 ppm for ^{13}CO shows superior resolution relative to the ^{1}H-^{15}N HSQC spectrum. Their position at higher ^{13}CO and lower ^{15}N chemical shift values relative to isolated Ala and to coil values indicates an alpha helical conformation. For simplicity, the label refers to the ^{13}CO of the *i-1* residue. Signals labeled in lower case with negative numbers correspond the His/TEV tag: M_{-29}GSSHHHHHHSSGLVPRGSHMASENLYFQ$_{-1}$

Sup. Fig. 5. Corroboration of small to negligible populations of α-helix or β-strand conformations in residues 91 - 110 of hCPEB3.

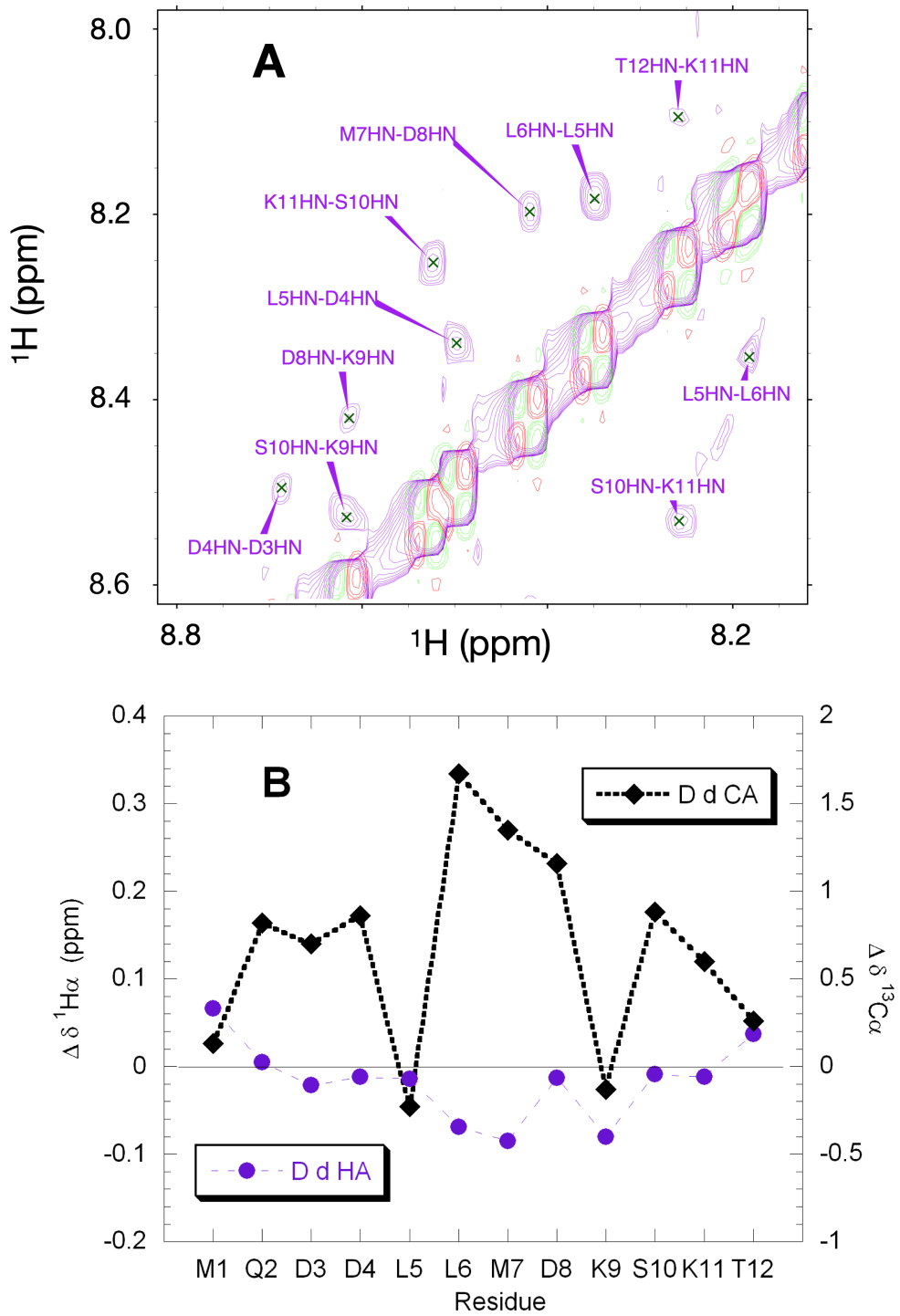


A. The assigned ^{1}H-^{15}N HSQC spectrum of a peptide corresponding to residues 91 - 110 of hCPEB3. The low chemical shift dispersion in the ^{1}HN dimension is a hallmark of a statistical coil. The residue number labels correspond to this peptide, to calculate the residue number in the full length protein, add 90.

B. The ^{1}H-^{13}C HSQC spectrum of the same peptide, showing the assigned ^{1}Hα-^{13}Cα, ^{1}Hδ-^{13}Cδ Pro and ^{1}Hβ-^{13}Cβ Ser & Thr correlations.

C. Conformational chemical shifts of ^{1}Hα (left y-axis, **purple circles**) and ^{13}Cα (right y-axis, **black diamonds**). Values of -0.41 ppm (^{1}H) and +3.1 ppm (^{13}C) correspond to 100% α-helix and values of +0.41 (^{1}H) and -3.1 ppm (^{13}C) correspond to 100% β-strand. Here residues show small conformation chemical shifts except for N- and C-termini, whose values are perturbed by end effects, and residues near Phe17 (Phe 107 in the full length protein) whose values are altered by ring current effects.

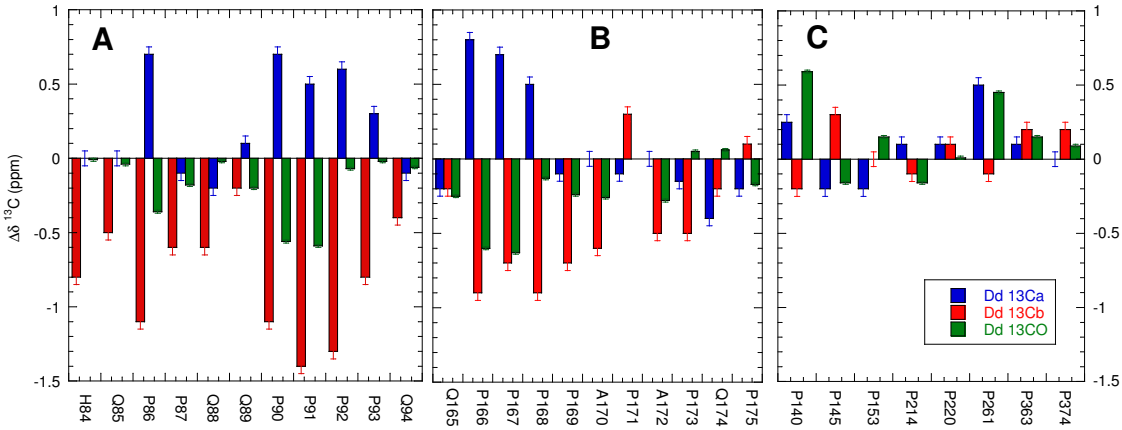
Sup. Fig. 6 Partial Formation of an α-Helix in the N-terminal Residues of hCPEB3



A. 2D ^{1}H-^{1}H NOESY (80 ms mixing time, **purple**) of a peptide corresponding to the initial residues of hCPEB3. Sequential NOE correlations, which are consistent with α-helical structure, are labeled. **Red** and **green** peaks along the diagonal are from the 2D ^{1}H-^{1}H COSY spectrum.

B. Conformational ^{1}H (**purple**) and ^{13}C (**black**) chemical shifts at 5 ºC for a peptide corresponding to the initial residues of hCPEB3 20% hexafluoroisopropanol (CF_3-CHOH-CF_3) The positive Δδ^{13}C and negative Δδ ^{1}H values are indicative of the formation of significant α-helix formation.

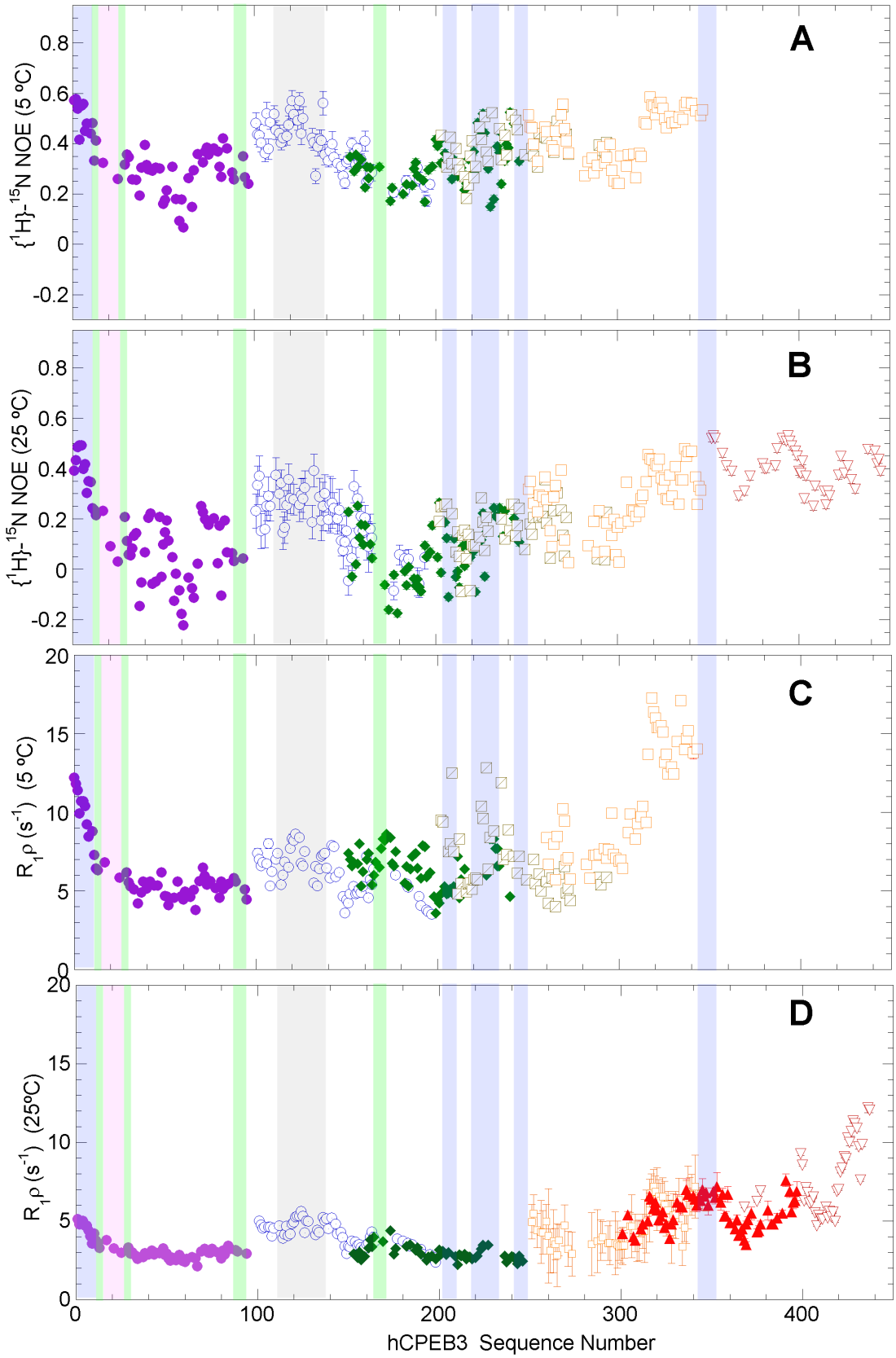
Sup. Fig. 7 The Consecutive Proline Residues of hCPEB3 show a Characteristic Pattern of Conformational Chemical Shifts.



A. and **B**. Conformational chemical shifts (Δδ) for ^{13}Cα (**blue** bars), ^{13}Cβ (**red**) and ^{13}CO (**green**) nuclei for residues in two Pro-rich stretches, H84-Q94 and Q165-175, each of which contains four consecutitive Proline residues which will be locked in the polyproline II helical conformation. The Δδ values of residues P90, P91, P92, P168, P167 and P168 were averaged to obtain the values given in **Table 3** on the main text.

C. Isolated proline residues, *i.e.* those which do not have another proline residue within two anterior or posterior positions along the sequence, do not show a pattern of significant Δδ deviations.

Sup. Fig. 8. Residue Level Dynamics of hCPEB3's Instrinsically Disordered Region

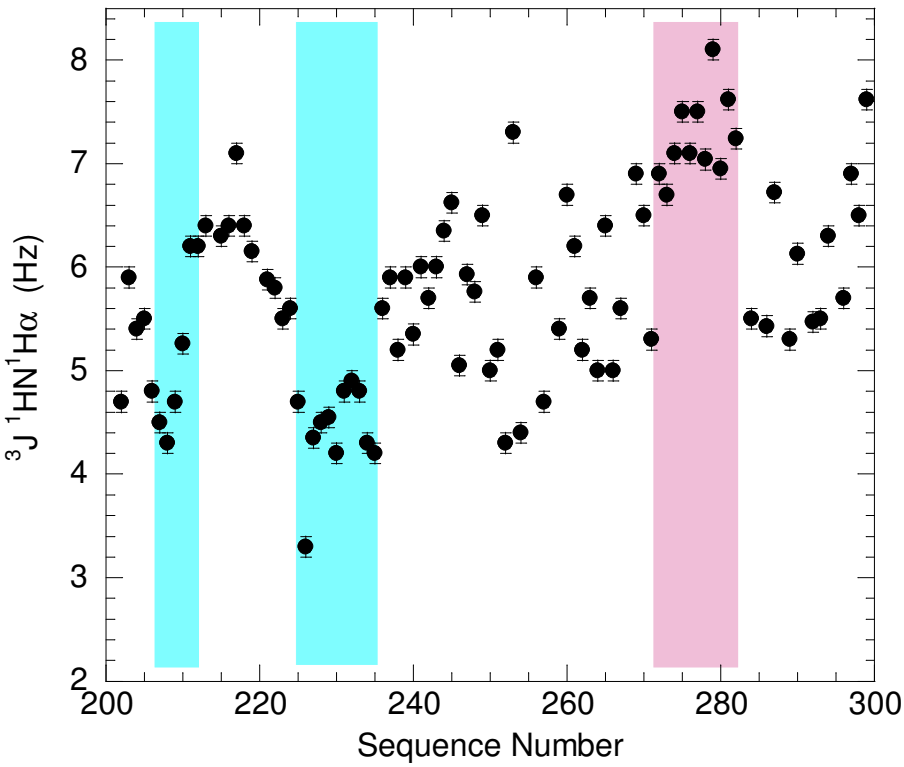


Residue-level dynamics on ns/ps time scales at 5ºC (**A**) and 25ºC (**B**) and on µs/ms at 5ºC (**C**) and 25 ºC (**D**). In panels **A** and **B**, a value of 0.86 is expected for completely rigid H-N groups; negative values are characteristic of a high flexibility. In panels **C** and **D**, while values less than 4 s^{-1} are hallmarks of

flexibility, higher values mean decreased mobility. Data from individual segments are colored differently: segment 1 = **purple filled circles**, segment 3 = **blue open circles**, segment 4 = **green diamonds**, segment 5 = **brown barred open squares**, segment 6 = **orange open squares**, segment 7 = **filled red triangles**, segment 8 = **open inverted maroon triangles**. No data are represented for segments 7 and 8 at 5ºC, segment 5 in panel **D**, and segment 7 in panel **B** as the spectra obtained were of insufficient quality. The lack of other values is due to $^{1}H^{15}N$ signal overlap or proline residues whose nitrogen lacks a hydrogen. Error bars represent uncertainties as estimated from peak intensity signal/noise for panels **A** and **B** and as obtained from the fit of a signal exponential decay function to the peak intensity versus delay time data for panels **C** and **D**. The error bar is frequently smaller than the data symbol.

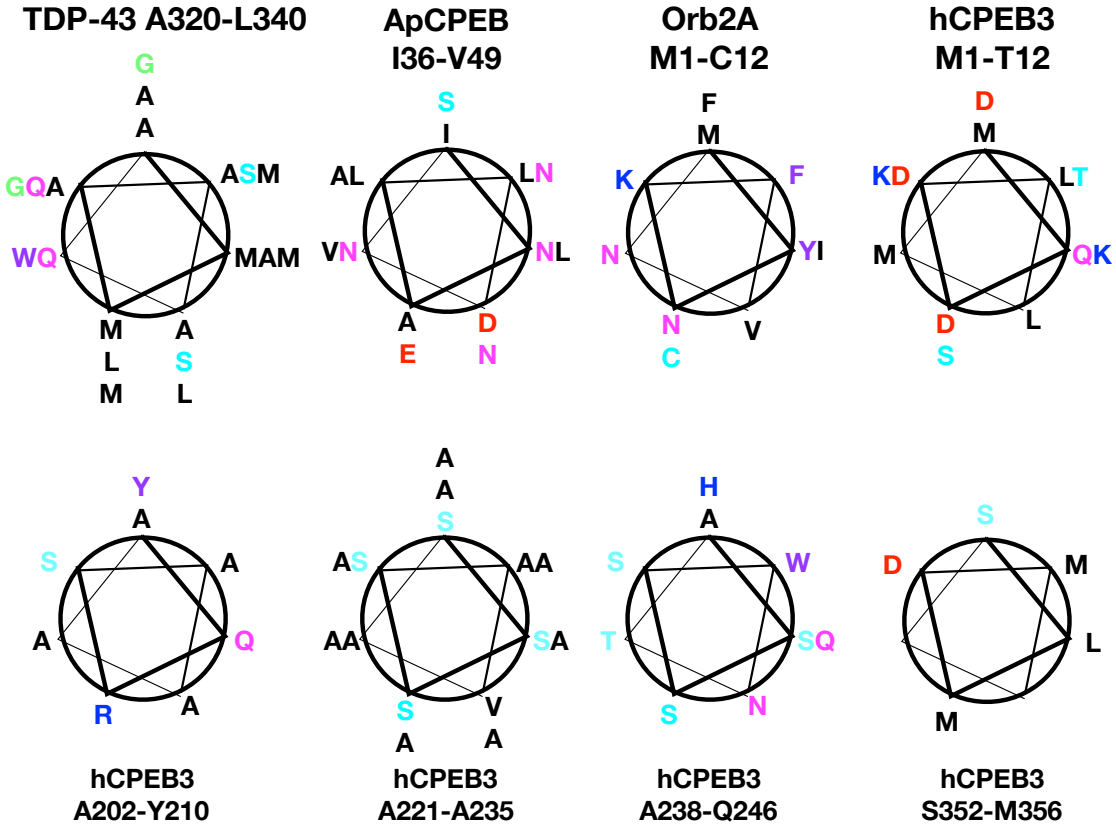
Elements of partial structure/interest are shaded **blue** for α-helices and **gray** for hydrophobic segments; these elements generally show decreased mobility. By contrast, the polyQ tract (shaded **magenta**) is more flexible. Whereas the dynamics of the PPII tracts (**green**) can now be assessed directly as these imine residues lack ^{1}HN, nearby residues do show relatively high mobility.

Sup. Fig. 9. ^{1}HN-^{1}Hα Coupling Constants for Segment 5 Confirm the Presence of α-Helices.



Intraresidual three bond ^{1}HN-^{1}Hα coupling constants ($^{3}J_{1HN1H\alpha}$) for segment 5 of hCPEB3. Regions with $^{3}J_{1HN1H\alpha} < 5$ (shaded **cyan**) correspond to two α-helices, in line with the results based on chemical shift deviations (*see **Fig. 3** in the Main Text*). A zone with $^{3}J_{1HN1H\alpha} > 7$ Hz (shaded **rose**) indicates extended conformations. This zone spans the $(VG)_5$ motif.

Sup. Fig. 10. Helices from Pathological and Functional Amyloids Are Stabilized by Distinct Interactions.



Helical wheel diagrams of TDP-43, ApCPEB, Orb2A and hCPEB3, nonpolar residues are colored **black**, aromatics=**purple**, Q/N=**magenta**, E/D=**red**, H/K/R=**blue**, G=**green**, C,S,T=**cyan**. The five hCPEB3 helices were identified here by NMR data. The TDP-43 helix was identified by NMR by Lim *et al.,* 2016. The ApCPEB and Orb2A helices are putative as they have not yet been confirmed experimentally.

Lim, L., Wei, Y., Lu, Y., & Song, J. (2016). ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43. PLoS biology, 14(1), e1002338. https://doi.org/10.1371/journal.pbio.1002338