Preferred conformations in the disordered region of human CPEB3, a functional amyloid linked to memory consolidation

D. Ramírez de Mingo,[1] D. Pantoja-Uceda,[2] R. Hervás,[3]
M. Carrión-Vázquez,[1*] D. V. Laurents, [2*]

[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]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, HKSAR, China.

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

Abstract

While implicated in neurodegenerative diseases, amyloids are also essential to some physiological processes, including memory consolidation by neuronal-specific isoforms of the Cytoplasmic Polyadenylation Element Binding (CPEB) protein family. CPEB mediates memory persistence by the formation of self-sustaining amyloid assemblies through its intrinsically disordered region (IDR). Here, we characterize the atomic level conformation and ps-ns dynamics of the 426-residue IDR of human CPEB3 (hCPEB3), which has been associated with episodic memory in humans, by NMR spectroscopy. We found that the first 29 residues: M_1QDDLLMDKSKTQPQPQQQQRQQQQPQP$_{29}$, adopt a helical+disordered motif. Residues 86-93: P_{83}QQPPPP$_{93}$, and 166-175: P_{166}PPPAPAPQP$_{175}$ form polyproline II (PPII) helices. 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 (NLS), form a modestly populated α-helix and border a phosphoTyr which may mediate STAT5B binding. These findings allow us to suggest a model for nascent hCPEB3 structural transitions at single residue resolution, advancing that amyloid breaker residues, like proline, are a key difference between

33 functional versus pathological amyloids. Besides revealing some aspects of the molecular basis of
34 memory, these findings could aid the future development of treatments for post-traumatic stress disorder.
35
36
37 **Areas:** Biophysics, Structural Biology, Biochemistry & Neurosciences.
38 **Keywords**: Memory Consolidation, Intrinsically disordered proteins, NMR spectroscopy.
39
40
41 **Introduction**
42 The molecular basis of long-term memory, which endures decades despite being built by
43 ephemeral biomolecules, has long fascinated biochemists [1]. Seminal findings by Si, Lindquist and
44 Kandel suggested that long-term changes in synaptic efficacy require a self-perpetuating amyloid state in
45 the *Aplysia* CPEB (*Ap*CPEB) [2][3] . This change in CPEB's conformation would lead to permanent
46 alterations at the synapse constituting a physical substrate of memory storage. To attain the aggregated
47 state necessary to stabilize memory, *Ap*CPEB contains a N-terminal IDR that is very rich in glutamine
48 (Q) residues which losses α-helix and gains coiled-coil and β-sheet structure during amyloid formation
49 *in* vitro [4][5].

50 The *Drosophila* homolog of CPEB, called Orb2, behaves in a similar fashion even though its N-
51 terminal IDR has a lower glutamine residue content and a more tightly regulated amyloid formation
52 [6][7][8][9]. Indeed, inhibition of Orb2 amyloid formation targeting the N-terminal IDR specifically
53 impairs memory consolidation, but not short-term memory in *Drosophila* [6][10]. Due to numerous
54 histidine (H) residues in the Orb2 Q/H-rich amyloid core comprised in the in the N-terminal IDR, pH
55 may regulate this structure's stability as suggested by CryoEM analysis [5] and characterization by NMR
56 spectroscopy [11] In mammals, the N-terminal IDR of the neuronal-specific isoform of CPEB3 is crucial
57 for amyloid formation and memory consolidation [12] [13]. The regulation of functional amyloid
58 formation in mammalian CPEB3 appears to be even more sophisticated due to multiple mechanisms
59 involving post-translational modifications [14], and feedback loops to maintain hCPEB3 expression levels
60 [13]. Compared to the *Aplysia* and *Drosophila* homologs, hCPEB3's content of glutamine residues in its

426-residues long IDR is lower and it contains diverse segments which are enriched for certain residues such as Ser, Ala, Pro, Gly+Val and hydrophobic residues (Table 1).

Mammalian CPEB3 travels to distinct neuronal regions to carry out multiple functions, where the 426-residue long IDR plays a key role (**Figure 1A**). Following its synthesis, CPEB3 is SUMOlyated, which has been reported to block CPEB3 aggregation [14]. Upon neuronal stimulation, CPEB3, which is mostly cytoplasmic, travels to the nucleus. This process is mediated by the karyopherin IPO5 through interactions with the NLS in the first RNA recognition motif (RRM) of hCPEB3 [15]. Inside the nucleus, CPEB3 interacts with STAT5B, which normally activates the transcription of genes such as *EGFR*, triggering signaling cascades thought to promote memory consolidation [16][17][17]. CPEB3-STAT5B binding, driven by interactions between the IDR of hCPEB3 and residues 639–700 of STAT5B, downregulates STAT5B-dependent transcription [17]. However, the details of this interaction have not yet been addressed. By contrast, the 3D structure of the first CPEB3 RRM domain has been elucidated and revealed a β-hairpin (W471-G485) proposed to play a key role in RNA recognition [18]. This domain, as well as the second RRM domain and Zinc Finger (ZnF) motif, were reported to bind specifically to the 3'UTR mRNA of the AMPA receptor subunit GluR2 [19]. Together, CPEB3 and its target mRNA eventually exit the nucleus and can join distinct biomolecular condensates such as stress granules and neuronal granules, which provide physiological transport to dendritic spines, or to dendritic P-body-like granules [20], where CPEB3 stores and downregulates GluR2 mRNA translation [19].

After synaptic activity in the hippocampus, SUMOylation of CPEB3 decreases [14] and CPEB3 converts, mediated by the IDR, from a translation repressor into a self-sustaining activator, promoting the translation of AMPA receptors [12]. This leads to structural modifications, including a more robust actin network, which fortify the spine and permanently enhance neurotransmission at a given particular synapse [13]. The hypothesis that CPEB functional amyloid formation is key for memory persistence in mammals is supported by the impairment in long-term memory and long-term potentiation in the CPEB3 conditional knockout mice [12]. In our species, the causal role of human CPEB3 (hCPEB3) 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 [21].

88 The 426-residue long IDR of hCPEB3 plays a key role mediating memory persistence through
89 this prion-like mechanism. It contains an amyloid-forming region spanning residues 1-200, and a
90 condensate-promoting region formed by residues 250-426 which are linked by an alanine rich segment
91 [22]. The full IDR is followed by two folded RRM which bind RNA and finally a ZZ-type ZnF domain
92 (**Figure 1B)**. Recent sequence and deletion mutational analyses of the IDR have begun to identify
93 subregions key for aggregation, such as the first 30 residues [13]. However for the hCPEB3 IDR,
94 programs to predict secondary structure tendencies give different outputs, Alpha Fold 2 structural
95 predictions [23] are marked as low to very low confidence (see https://alphafold.ebi.ac.uk/entry/Q8NE35)
96 and to date, no high-resolution experimental data on the partial structures or motions have been reported.
97 Here, motivated by the key roles of the IDR in CPEB3 functions in transcription via association with
98 STAT5B, as well as in CPEB prion-like aggregation required for memory persistence, we characterize
99 the atomic level conformation and dynamics of the complete IDR of hCPEB3 by NMR spectroscopy,
00 suggesting a speculative model for CPEB3 aggregation.
01

02

03

04 **Results**

05 **hCPEB3's IDR is chiefly disordered.** As a first step to experimentally characterize hCPEB3's IDR, we
06 probed the complete 426 residue IDR of hCPEB3 by biophysical techniques and homonuclear NMR. Its
07 fluorescence emission spectra, recorded at temperatures ranging from 2 to 70 ºC, show emission
08 maximum > 350 nm. This is consistent with its six Trp residues being solvent exposed and not buried in
09 the hydrophobic core of a folded domain (**Sup. Fig. 1A**) [24]. The far UV CD spectrum of the hCPEB3
10 IDR also shows the hallmarks of a disordered protein, namely a minimum near 200 nm [25]. No spectral
11 features indicative of α-helix and β-sheet; namely, minima at 208, 218 or 222 nm and no maximum at
12 195 nm, are evident (**Sup. Fig. 1B**). The 1D ^{1}H and 2D ^{1}H-^{1}H NOESY spectra show ^{1}H signals clustered
13 into a narrow bands near the values observed for short, unstructured peptides (**Sup Fig. 1C**) [26][27]. The
14 sequence alignment of several representative vertebrate CPEB3 proteins using the T-Coffee program is
15 shown in **Sup. Fig. 2**. Very similar results were obtained from the Omega Clustal program (not shown).
16 Whereas most IDPs show poor levels of sequence conservation, some stretches rich in hydrophobic
17 residues, such residues M1-T12, W111-F139 and Y341-I357, are highly conserved. By contrast,
18 glutamine rich, alanine rich and some proline rich segments are present only in mammals. Taking all these
19 data together, the presence of large, stably folded domains in the IDR can be ruled out, but short segments
20 with partly populated secondary structures could still be present.

21

22 **Atomic level characterization reveals partially structured elements in hCPEB's N-terminal**
23 **"disordered" region.** To discover and characterize possible segments with partial secondary structure,
24 we applied multidimensional heteronuclear NMR. As the full length IDR is too long to characterize by
25 this methodology, we have followed the "divide and conquer" approach implemented by Zweckstetter *et*
26 *al.* to characterize tau, a similarly sized IDP implicated in Alzheimer's disease and other tauopathies [28].
27 As described in the **Material & Methods** section, and shown in **Table 1**, eight overlapping segments of
28 100 residues were characterized.

Using our powerful ^{13}CO, ^{15}N, ^{1}HN based assignment strategy, over 99% of the main chain ^{13}CO, $^{13}C\alpha$, ^{15}N, ^{1}HN and the $^{13}C\beta$ resonances were assigned for residues 1 – 450 of hCPEB3. The chemical shifts of the complete IDR of hCPEB3 are reported in the **BMRB (**entry number **50256)**, and the original 2D and 3D spectral data have been deposited in the Mendeley data repository. 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 2D ^{1}H-^{15}N HSQC spectra of segments 1, 3, 4, 5, 6, 7 and 8 are shown in **Sup. Fig. 5**. The similar position of most signals of ^{1}H-^{15}N in neighboring segments additionally suggests a sparsity of long range interactions under these conditions. Likewise, the majority of the crosspeaks of the same residues in adjacent segments also overlap or are close together in the 2D $^{13}CO^{15}N$ spectra of segments 1, 3, 4, 5, 6 and 8 (**Sup. Fig. 6**).

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 91-110 located in 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 this region by NMR spectroscopy. No significant trends towards structure formation were detected (**Sup. Fig. 7**).

The $^{13}C\alpha$ and ^{13}CO conformational chemical shifts ($\Delta\delta$) of the hCPEB3 segments 1, and 3-8 are plotted in **Sup. Fig. 8**. These data show five segments, comprising residues 1-10, 202-210, 222-234, 238-246 and 346-356, with significantly high $\Delta\delta$ $^{13}C\alpha$ and $\Delta\delta$ ^{13}CO values. Such values are characteristic of partially populated α-helices and are examined in detail in the following paragraphs.

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 (20 %) α-helix in the first ten residues of the protein (**Figure 2 A,B**). Increased conformational chemical shifts are observed at 5 ºC, reflecting a higher amount of helical structure upon cooling. Standard ^{1}H-detected ^{15}N relaxation

55 measurements detect that these residues are the most rigid part of segment 1 (**Figure 2 C,D**). These
56 results are corroborated by ^{13}C-detected ^{15}N relaxation experiments (**Sup. Fig. 9**).

57 The α-helix detected for these first residues extends N-terminally into the His/Tev tag. To rule
58 out a possible structure-promoting effect on segment 1, we tried to remove it by proteolytic cleavage with
59 the TEV protease. Multiple attempts failed, which suggests that the helix spanning the last residues of
60 His/Tev tag and the first residues of the hCPEB3 IDR is present and impedes the proteolytic cleavage.
61 Therefore, we characterized a dodecamer peptide whose sequence corresponds to the first twelve residues:
62 M_1QDDLLMDKSKT_{12}, of the hCPEB3 IDR. The observation of a series of weak ^1HN_i - $^1HN_{i+1}$ nuclear
63 Overhauser enhancement (NOE) crosspeaks reveals that this peptide has a slight tendency to form α-helix
64 in aqueous buffer (**Sup. Fig. 10A**). Fluorinated alcohols like trifluoroethanol (TFE) and
65 hexafluoroisopropanol (HFIP) are known to increase the population of helical conformations in peptides
66 which have an α-helix forming tendency, but not in peptides which prefer to adopt β-strands or random
67 coil [29]. In the presence of 20% HFIP, the population of helix in this peptide increases strongly, based
68 on the observation of stronger and more numerous NOE crosspeaks as well as ^{1}Hα and ^{13}Cα
69 conformational chemical shifts (**Sup. Fig. 10B**). These findings evince that the first 12 residues of
70 hCPEB3 do tend to adopt an α-helix.

71 Interestingly enough, the polyQ segment, Q_{16}QQQRQQQQ_{24}, does not form an α-helix or a β-
72 strand and appears to be thoroughly disordered and flexible (**Figure 2A,B**). A construct spanning residues
73 1-200 of hCPEB3, which contains the Q_4RQ_4 motif, plays a role in hCPEB3 amyloid formation as
74 indirectly evidenced by the anti-amyloid action of the polyglutamine binding peptide 1 (QBP1) [22]. This
75 polyQ segment is preceded and followed by Pro-Gln-Pro residue triplets ($P_{13}QP_{15}$ and $P_{25}QP_{27}$).
76 Considering the inhibitory effect of proline residues previously observed for polyQ amyloid formation in
77 Huntingtin by Wetzel and co-workers [30], it is likely that these PQP mini-motifs check amyloidogenesis
78 by the polyQ segment. The first 100 residues also contain a predicted SUMOylation site [22] at Lys 47
79 and ends with a proline-rich segment P_{86}PQQPPPPQEPAAPG_{100}, which is associated with solubility.
80 Whereas recently reported NMR criteria [31], allow us to rule out that this stretch folds into a stable
81 polyproline II (PPII) helical bundle, the steric limitations of polyproline segments mean that residues 86-

82 93 adopt an isolated, partly populated PPII helix (**Figure 2E**). In fact, the consecutive proline residues
83 show a distinct pattern of conformational chemical shifts; namely +0.6 ppm, -1.0 and -0.3 for $^{13}C\alpha$, $^{13}C\beta$
84 and ^{13}CO, respectively (**Table 3**, **Sup. Fig. 11**). Not observed in isolated proline residues, we advance
85 that they are hallmarks of a PPII helical conformation.
86
87 **Residues 101 - 200 of hCPEB3 contains a rigid nonpolar segment and a PPII helix.** Regarding
88 residues 101 – 200, no strong trends to adopt α-helical or β-structures are detected. Nevertheless, the
89 stretch composed of residues, W_{111}STGTTNAVEDSFFQGITPVNGTMLFQNF_{139} which contain
90 numerous aliphatic and aromatic residues, shows relatively high rigidity, both on fast ns/ps as well as
91 slower μs/ms timescales (**Sup. Fig. 12)**. This finding is interesting considering that this relatively
92 hydrophobic segment also appears to be essential for hCPEB3 amyloid formation *in vitro* [22] and very
93 recently it has been reported to form amyloid in mouse CPEB3 [32]. In addition, the stretch of residues
94 161-190: Q_{161}HHQQPPPPA_{170}PAPQPAQPAQ_{180}PPQAQPPQQR_{190}, has a very high Q/P content and Pro
95 and Gln are the residues with the highest intrinsic tendencies to adopt PPII helices [33]. The consecutive
96 proline residues, P_{166}PPPAPAPQP_{175}, also display the characteristic PPII pattern of conformational
97 chemical shifts (**Sup. Fig. 11 ABC**) seen for residues 86-93. Although more weakly than long stretches
98 of pure polyproline [34][35], a synthetic peptide corresponding to residues P166-P175 of hCPEB3 binds
99 to human Profilin 1, a known mediator of interactions with actin (**Sup. Fig. 11 D**).
00
01 **Residues 201-300 contain three α-helical segments and a disordered $(VG)_5$ segment**. Significant $^{13}C\alpha$
02 and ^{13}CO chemical shift deviations with respect to values predicted for a statistical coil, for three residue
03 segments spanning **<u>A_{202}QRSAAAY</u>**$_{210}$ GHQPIMTSKP_{220}-**<u>S_{221}SSSAVAAAA$_{230}$AAAAA</u>**)
04 SS**<u>ASS$_{240}$SWNTHQ</u>** SVHAA_{250} (**Figure 3A**). These results indicate that the three segments of
05 underlined residues adopt partially populated α-helices. Based on the magnitude of the conformational
06 chemical shifts, the helical populations are different; being about 30% for the A_{202}-Y_{210} α-helix, 80% for
07 the S_{222}-A_{234} α-helix and 20% for the A_{238}-Q_{246} α-helix at 5 ºC; whereas these populations decrease at 25
08 ºC to approximately 10%, 40% and 15%, respectively, they are still significant (**Figure 3B**). The presence

09 of the first and second helices are confirmed by ^{1}HN-^{1}Hα coupling constants (**Sup. Fig. 13**). Moreover,
10 analysis with TALOS+, which predicts secondary structure taking into account ^{13}Cβ, ^{15}N and ^{1}Hα
11 chemical shifts in addition to ^{13}Cα and ^{13}CO, confirms the presence of these three helical segments and
12 structural calculations with CYANA suggest that the three helices do not tend to adopt a preferred
13 alignment relative to each other (data not shown). The helices are not especially rigid on fast ps-ns time
14 scales (**Figure 3C**) or the slower μs-ms time domain (**Figure 3D**) at 25ºC, but do show a heightened
15 stiffness at 5 ºC (**Sup. Fig. 12**). Helical wheel projections (**Sup. Fig. 14**) suggest that different interactions
16 contribute stability to these α-helices. Gly 211 and His 212 are positioned to stabilize the A_{202}-Y_{210} α-
17 helix by a C-capping motif [36]. Whereas Ala has a very high intrinsic helix forming propensity, the
18 propensity of Ser is low [37]. In this segment, however, the Ser residues are positioned at the N-terminus
19 of the α-helices, where adding negative charge via phosphorylation would increase the helical population,
20 considering the well-known stabilizing effects of charge / macrodipole interactions and N-capping H-
21 bonds [38]. Although this segment's insolubility thwarts attempts to test directly the impact of
22 phosphorylation using synthetic peptides, we note that these Ser residues are placed at the positions where
23 phosphorylation is expected to increase α-helix stability the most [39]. Moreover, at neutral pH, where
24 phosphoserine carries two negative charges, the stabilization is substantially greater than at pH 4, where
25 it carries one [39]. The last α-helix, A_{238}-Q_{246}, is less populated, but its stability might increasee if W242
26 were to engage in long-range interactions, such as with the hydrophobic or cationic residues of the first
27 α-helix, *i.e.* MQDDLLMDKSKT. To test this possibility, we studied two polypeptides containing the
28 M_1-T_{12} and A_{238}-Q_{246} helical segments with and without an N-terminal Dansyl group, connected by a
29 flexible $(Gly)_4$ linker. The results of FRET and 2D NMR spectroscopy evince that this polypeptide adopts
30 a conformational ensemble significantly more compact than a statistical coil, but that the helix contents
31 are not significantly altered (**Sup. Fig. 15).**

32 One of the most striking features in hCPEB3's sequence is a short dipeptide protein motif (Val-
33 $Gly)_5$ spanning residues 271 – 281, which is reminiscent of longer $(Ala\text{-}Gly)_N$ and $(Pro\text{-}Gly)_N$ and (Arg-
34 $Gly)_N$ dipeptide repeat proteins encoded by mutant C9orf72 which have been implicated in ALS [40];
35 [41]. Our *in silico* analysis identified this segment as having a high potential to form amyloid [22]. In

36 the context of hCPEB3, however, this segment is among the most disordered and flexible of all the zones
37 of the IDR (**Sup. Fig. 8, 12**). Just beyond the $(VG)_5$ segment, there is a stretch of 15 residues:
38 S_{284}PLNPISPLKKPFSS$_{298}$, whose NMR parameters indicate disorder and flexibility (**Sup. Fig. 8 & 12**).
39 Nevertheless, this stretch contains four Ser residues reported to phosphorylated by protein kinase A (PKA)
40 or calcium/calmodulin-dependent protein kinase II [42] (**Table 1**), and therefore might be important for
41 the transition between short- and long-term memory. Residues P_{303}-PKFP**R**AAP_{311} are proline rich.
42 Predictions suggest that Arg 308 can be methylated (**Table 1**). This modification, whose impact has not
43 been probed here, was reported to fortify cation–π interactions, reduce interactions with RNA and
44 destabilize condensates in other proteins [43].
45
46 **The residues forming the Nuclear Export Signal (NES) show a marked tendency to adopt α-helical**
47 **structures.** Significant conformational chemical shifts were also observed for residues L349-L353 which
48 form the NES (**Figure 4)** indicating the presence of helical structure. Using the ^{13}CO, ^{15}N, ^{1}HN, ^{13}Cα and
49 ^{13}Cβ chemical shift data as input, a family of conformers was calculated using the programs TALOS+
50 and CYANA for residues P333-P363. This 31-residue segment is rich in aromatic (five) and aliphatic
51 (six) residues, which is unusual for a disordered polypeptide. The resulting structures reveal that residues
52 L346-L349 adopt one turn of α-helix and residues S352-M356 form a short α-helix (**Figure 4A**). It is
53 notable that this conformer positions five nonpolar residues: L346, L349, L353, M354 and I357 on the
54 same face of the α-helices. Y341, the putative phosphorylation site, is in an extended portion of the
55 backbone and would be accessible for this post-translational modification (PTM). Whereas the
56 conformational ensemble will contain many other structures, based on conformational chemical shifts as
57 illustrated by the $\Delta\delta^{13}$Cα & $\Delta\delta^{13}$CO values shown in **Figure 4B** the α-helical population is about one
58 third. The presence of rigid conformers is corroborated by relatively high $\{^{1}H\}$-^{15}N NOE ratios (**Figure**
59 **4C**) and elevated transverse relaxation ratios (**Figure 4D**).
60 Beyond the NES α-helix, no segments with preferred secondary structure are detected. The last
61 residues of the segment 8 construct: S_{426}-RKVFVGGLPPDIDEDEITASFRRF$_{450}$, belong to the RRM1
62 domain. According to the 3D structure [18] residues K428-G432 adopt a β-strand and residues E440-

63 R449 form an α-helix in the context of the complete RRM1 domain. Here these segments appear to be
64 largely disordered. After a proline-rich zone ending around residue 380, the next fifty residues have a
65 higher content of nonpolar residues and tend to be more rigid (**Sup. Fig 12**). Residues 400 – 412:
66 SHGDQALSSGLSS contain five Ser residues reported to be phosphorylated [42], **Table 1**). Like the
67 four Ser residues of the 284-291 stretch mentioned above, this could be an element involved in the
68 structural basis of memory consolidation.
69
70 **Discussion**
71 Beyond the fundamental biological interest in comprehending the molecular basis of long-term human
72 memory, the atomic-level understanding of the structural transitions involved in CPEB aggregation could
73 also guide the development of treatments for Post-Traumatic Stress Disorder and other diseases associated
74 with the storage of "pathological memories" [44]. Furthermore, like its homologs in *Aplysia* and
75 *Drosophila*, hCPEB3 resembles the abundant and diverse superfamily of RNA-binding proteins that
76 contain RRM and/or ZnF domains as well as intrinsically disordered prion-like regions, such as Fused in
77 Sarcoma (FUS) or Transactive Response DNA Binding Protein of 43 kDa (TDP-43). FUS and TDP-43
78 are essential proteins, but their anomalous aggregation has been implicated in amyotrophic lateral
79 sclerosis (ALS) and frontotemporal dementia (FTD) [45]. In addition, TDP-43 has been linked to the
80 recently described Limbic-Predominant Age-Related TDP-43 Encephalopathy (LATE) [46]. Thus, the
81 comparison of the physiologically relevant amyloid formation, and the dynamics at the monomer level,
82 by hCPEB3 versus the aberrant amyloid formation by TDP-43 may reveal why the latter can become
83 pathological.
84
85 The biophysical analysis of the full length IDR of hCPEB3 shows that it lacks stable secondary
86 structure. Compared to well-ordered protein domains, disordered regions tend to change their amino
87 acid sequence much more rapidly over the course of evolution due to a lack of structural constraints for
88 folding [47]. In hCPEB3, the N-terminal α-helix and the 350's (NES) α-helix as well as its accompanying
89 putative phosphoTyr site, are well conserved throughout vertebrate CPEB3's, which suggests that they
90 are important for biological function (**Sup. Fig. 2**). Additional segments, like the stretch of hydrophobic

residues key for amyloid formation [22] and a cluster of three Trp residues (W242, W252 and W259) are also 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 poly-Q segment or substitute it for hydrophobic amyloidogenic segment highlights the cassette or modular nature of PLDs, which was previously established for the *Drosophila* CPEB homolog [5].

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 [48], the Androgen Receptor [49], Ataxin-7 [50] and TDP-43 [51] (**Table 2, Sup. Fig. 14**) 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 [52] [53] or by forming biomolecular condensates via liquid-liquid phase separation [54]. In the case of Huntingtin, the poly-Q segment is followed by a stretch of poly-P, which will be locked in a PPII helical conformation. This poly-P stretch has been shown to strongly slow Huntingtin amyloid formation *in vitro* [30]. However, poly-Q expansions in Huntingtin protein [48] and the Androgen Receptor [55] are known for their ability to form toxic aggregates in Huntington's disease and in spinobulbar muscular atrophy, respectively. Interestingly enough, Huntingtin homolog in *Aplysia* has been proposed to play a role in memory consolidation [56].

PPII helices may regulate amyloid formation and mediate interactions with actin. Whereas hCPEB3 and its homologs Orb2 and *Ap*CPEB have α-helix destabilizing residues between the N-terminal helix and the amyloidogenic Q-rich segment, such as proline in hCPEB3 and Orb2 or serine and valine in the case of *Ap*CPEB, no such α-helix busters are present in the analogous segments of Huntingtin, the Androgen Receptor or TDP-43 (**Table 2**). This allows the poly-Q segment to interact with the α-helix through the formation of sidechain to main chain hydrogen bonds in the case of the Androgen Receptor [55] and to form part of the amyloid structure as proposed recently for TDP-43 by Eisenberg and collaborators [57]

19 and present in ALS/FTLD patient brains [58]. Based on these observations, we suggest that the presence
20 of α-helix breakers between the α-helix and the poly-Q(/N) segments in functional amyloids and their
21 absence in pathological amyloids is a fundamental difference that could be a molecular basis of their
22 radically different toxicities.

23 PPII helices, well known in collagen and certain glycine-rich proteins [59], also play key roles
24 in mediating protein-protein interactions in biomolecular condensates [60]. Whereas conformation
25 chemical shifts have proven to be extremely useful tools to identify α-helices and β-strands in folded and
26 disordered proteins, conformational chemical shifts for PPII helices were less known. The first 450
27 residues of hCPEB3 include no less than 66 proline residues (15% of the total). We hypothesize that
28 segments such as G_{166}PPPPAPAPQP$_{175}$, may bind profilin, which has specific domains to interact with
29 proline-rich polypeptides [34] and to mediate interactions with the actin, whose levels rise at the synapse
30 following CPEB3 aggregation [13]. On the basis of the thorough set of chemical shifts obtained here,
31 including Pro ^{15}N assignments which are rare in the literature, we propose a pattern of conformational
32 chemical shifts that define PPII helices (**Table 3**). Combined with the recently published standards for
33 glycine-rich PPII helical bundles [31], these values should aid the detection of PPII helices in
34 biomolecular condensates and the elucidation of their roles in physiology and pathology.
35

36 **STAT5B binding may occlude hCPEB's NES leading to nuclear retention.** The calculated
37 conformers for the NES showed the preferred formation of α-helices which orient five hydrophobic
38 residues on the same side (**Fig. 4**); this is a common feature in nuclear export signals [61]. CPEB3
39 constantly shuttles in and out of the nucleus and its subcellular localization has important consequences
40 as it regulates gene transcription via association with STAT5B in the nuclei of neurons as well as mRNA
41 translation in dendritic spines. Neural stimulation leads to hCPEB3, which is normally predominately
42 cytoplasmic, to be concentrated in the nucleus [17]. That earlier study also revealed that the region
43 spanning residues 639 – 700 of STAT5B binds the N-terminal IDR of hCPEB3 [17]. To learn about the
44 nature of this interaction, we examined the structure of STAT5B (PDB: 6MBW) [62]. Remarkably, this
45 structure reveals that the key hCPEB3-interacting residues in STAT5B form an SH2 domain. Based on
46 this, we advance that hCPEB3 may contain a phospho-Tyr residue that is recognized and bound tightly

47 by STAT5B. Considering the results of previously reported motif analyses [63][64], the most plausible
48 site for phospho-Tyr modification in hCPEB3 is Tyr341, as this residue is bordered by SRPY$_{341}$DTF
49 which closely matches a consensus substrate "[E/V/D]-[G]-[I/P/L]-Y-E-X-[F/T/V]" of Scr, which is one
50 of the most relevant tyrosine kinases. Based on this proposed PTM site, and our finding that the preferred
51 conformation of nearby NES includes an α-helix spanning residues L349-L353 *(vide supra*), we can
52 advance a speculative working hypothesis to account for how STAT5B binding retains hCPEB3 in the
53 nucleus; namely binding of phospho-Tyr341 by the SH2 domain of STAT5B may occlude the neighboring
54 NES of hCPEB3. This might explain why hCPEB3 becomes concentrated in the nucleus under conditions
55 of neuronal stimulation [15]. Recent studies have shown that certain mutations in the nuclear localization
56 and export signals or defects in nucleocytoplasmic transport of TDP-43, another aggregation-prone, RNA-
57 binding protein, lead to an aberrant aggregation [65]. Although beyond the scope of this study, future
58 research could test whether mutations in the NES of hCPEB3 affect transport and lead to long-term
59 memory defects.
60
61 **Towards an atomic-level description of the first steps in hCPEB3 IDR self-association.** We suggest
62 a speculative working hypothesis for hCPEB3 conformational changes during memory consolidation
63 (**Figure 5**). Initially, the first 200 residues of hCPEB3, which are necessary and sufficient for aggregation
64 and amyloid formation *in vitro* [12][22], are mostly disordered except for the relatively stable α-helices
65 formed by residues 222-234 and modestly populated α-helices at the N-terminus and spanning residues
66 202-212 and 237-245 (**Figure 5A**). The Q_4RQ_4 motif and hydrophobic segment **F$_{123}$FQGIT**-PVNGT-
67 **MLFQNF$_{139}$** are initially disordered and premature amyloidogenesis is discouraged by SUMOlyation
68 [14] and proline breaker motifs.
69 Following deSUMOlyation and other putative PTMs such as phosphorylation, association
70 among the α-helices into more compact ensembles, as suggested by FRET results (**Sup. Fig. 15**),
71 would become possible, and might be strengthened by hydrophobic interactions between Met 7 and
72 Trp 242 or cation-π interactions between Lys 11 and Trp 242 (**Figure 5B**), analogous to the long
73 range contacts seen in a TDP-43 amyloid [66]. The increased production of hCPEB3 upon neuronal

74 stimulation [12] as well as the proximity of the serine and alanine rich α-helix (residues 222-234)
75 may promote α-helix formation by the QQQQRQQQQ motif. These could then associate to form a
76 coiled-coil, as has been demonstrated in model polypeptides [67] [52] [68]. Moreover, completely
77 helical polyalanine peptides, some of which come from proteins linked to polyalanine expansion
78 diseases, were reported to promote coiled-coil mediated aggregation although in some case no
79 conversion into β-sheet structures was observed [52] [69]. Although future studies are necessary to
80 provide additional evidence and test plausible hybrid aggregation mechanisms with several elements
81 of structure acting in parallel, this model is already supported by analogous results on *Ap*CPEB [4]
82 [53] and polyalanine expansions [69], which led to the proposal of similar mechanisms.

83 These events would reinforce intermolecular contacts within the dendritic P-body like granule
84 [20] leading to gelification and, eventually, amyloid formation (**Figure 5C**). The final amyloid structure
85 could be comprised of the Q_4RQ_4 motif and the hydrophobic segment as evidenced by hCPEB3 fibril
86 formation kinetics. A construct spanning residues 1-200 of hCPEB3, which contains the Q_4RQ_4 motif,
87 plays a role in hCPEB3 amyloid formation as indirectly evidenced by the anti-amyloid action of the
88 polyglutamine binding peptide 1 (QBP1) [22] . In the case of the hydrophobic segment, its role in mouse
89 CPEB3 amyloidogenesis has recently been thoroughly corroborated [32]. In addition, the poly-A segment
90 of helices 202-212 and 222-234 may also transform into amyloid (**Figure 5C**). Such α-helix to amyloid
91 conformational transformations have been previously described in alanine-rich polypeptides as diverse as
92 a fish antifreeze protein [70] and a synthetic $(Ala)_{10}$-$(His)_6$ hexadecapeptide [71], and could form
93 pathological amyloids in a number of reported diseases involving expansion of poly-A tracts [72].

94 Residues 217-284 of CPEB3 have been reported to be key for interactions with actin [13].
95 Considering this, the hypothetical associations among the α-helices formed by residues M_1-T_{12} and
96 residues A_{202}-Q_{246} suggested in the preceding paragraphs could dispose the PPII helices formed by
97 residues P_{86}-Q_{94} and P_{166}-P_{175} (**Sup. Fig. 10**) to bind profilin. This protein contains a second binding site
98 specific for actin and promotes the formation of actin filament networks [73]. Such actin networks are
99 known to become more extensive and robust as a dendritic bud strengthens during memory consolidation
00 [74].

01

02 **Materials and Methods**

03 **1. Materials**: $^{15}NH_4Cl$ and ^{13}C-glucose were purchased Tracertec (Madrid, Spain), D_2O, was a product
04 of Euroisotop, deuterated acetic acid was from Sigma/Aldrich and 4,4-dimethyl-4-silapentane-1-sulfonic
05 acid (DSS) as the internal chemical shift reference, was from Stolher Isotopes Chemical Company.
06 A twelve residue peptide, called hCPEBpep1, whose sequence corresponds to the protein's first 12
07 residues (M_1QDDLLMDKSKT$_{12}$) and a twenty residue peptide, called hCPEBpep2 with the sequence of
08 residues P_{91}PPQEPAAPGASLSPSFGST$_{110}$ in hCPEB3 were purchased from Genscript. hCPEBpep2's
09 sequence overlaps with the C-terminal of Segment 1 and the N-terminus of Segment 3. A proline-rich
10 peptide acGPPPPAPAPQPam, with acetylated and amidated termini and whose sequence covers residues
11 166-175 in hCPEB3, two polypeptides with the sequence MQDDLLMDKSKTGGGGASSSWNTHQ
12 (with and without an N-terminal Dansyl group) were also obtained from Genscript. All the peptides were
13 over 95% pure, as assessed by HPLC, and their identities were confirmed by mass spectrometry and NMR
14 spectroscopy. Recombinant Human Profilin 1, produced recombinantly in *E. coli* was obtained from
15 Abcam (reference number ab87760). It is over 95% pure as judged by SDS PAGE.

16

17 **2. Sample Production:** Coding mRNAs of hCPEB3 vary in length due to an embedded human delta
18 virus-like ribozyme which slowly splices out introns, leading to the generation of multiple isoforms [75].
19 It is also noteworthy that Orb2A's pre-mRNA contains an intron with multiple stop codons which is only
20 spliced out when certain "memorable" stimuli are experienced [9]. Here, the hCPEB3 isoform 2, Uniprot
21 Q8NE35-2 / Genebank CAI14105.1, is studied.

22

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

28 Each segment was cloned into the pET-28a(+) by PCR using the full length human CPEB3-2 in
29 pLL3.7 plasmid as the template kindly provided by Dr. Yi-Shuian Huang [76]. The DNA amplified
30 fragments were digested with XhoI and NheI. Expression of the resulting clones led to fusion proteins
31 containing a His_6 tag and a TEV NIα protease cleavage site. Thus, each segment studied had the sequence:
32 MGSSHHHHHHSSGLVPRGSHMASENLYFQ, at its N-terminus.

33 All overlapping segments were expressed in the *E. coli* BL21 Star (DE3) strain using the T7
34 expression system (Novagene). $^{13}C/^{15}N$ isotopic labelling of each segment was done by using a previously
35 published protocol [77]. Briefly, cells were grown in 1 L of LB at 37 ºC by shaking at 280 rpm upon
36 reaching optical cell densities at 595 nm (OD_{595}) ~ 0.6-0.7. Cells were pelleted by a 30 min centrifugation
37 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
38 for 1 L of 5 × M9 salts) excluding nitrogen and carbon sources. Cell pellets were resuspended in 250 mL
39 of isotopically labelled minimal media M9 salt solution supplemented with ^{13}C D-glucose 4.0 g/L and
40 $^{15}NH_4Cl$ 1.0 g/L (Cambridge Isotope Laboratories, Inc.), then incubated to allow the recovery of growth
41 and clearance of unlabeled metabolites. Protein expression was induced after 1 h by addition of IPTG to
42 a concentration of 1 mM. After a 4.5 h incubation period, the cells were harvested.

43 Cell pellets were lysed with buffer with the following composition: 50 mM NaH_2PO_4/Na_2HPO_4,
44 500 mM NaCl, 50 mM imidazole, 6 M guanindium chloride (GdmCl), pH 7.4 and then sonicated. Each
45 recombinant segment was purified by Ni^{2+}-affinity chromatography using HisTrap HP purification
46 columns with a FPLC system (ÄKTA Purifier, GE Healthcare) with elution buffer consisting of 50 mM
47 NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, 6 M GdmCl, pH 7.4.

48 If necessary, the pure segments were then incubated with TEV NIα protease O/N at 4ºC [78].
49 The cleaved protein was then subjected to dialysis and recovered. Protein samples were stored at -80ºC
50 until use. Eventually, they were desalted to 1 mM DAc, pH 4.0 by gel filtration chromatography using a
51 PD-10 column (GE Healthcare), and concentrated to a final protein concentration of 1.0 – 1.5 mM in 250
52 μL using a Vivaspin microfiltration device and placed in a 5 mm Shigemi reduced volume NMR tube for
53 measurements. These low pH, low ionic strength conditions have been reported [79] and corroborated
54 [80] to maximize solubility by increasing charge-charge repulsion between protein molecules. α-helix

55 stability depends on pH due to charge-charge interactions, charge-helix macrodipole interactions and
56 intrinsic helix propensities [38] . A consideration of these factors suggests that the effect of increasing the
57 pH from 4 to 7 would be mildly stabilizing for the first α-helix and insignificant for the four remaining α-
58 helices detected in the hCPEB3 PLD.

59 Non-labeled full length hCPEB3-IDR expression and purification was carried out essentially as
60 described in [22]. Briefly, cells were grown in 1 L of LB medium at 37 ºC until reaching an OD_{600} = 0.6
61 – 0.7 and protein expression was induced for 4 hours by adding IPTG at 1 mM final concentration. Cells
62 were harvested and following sonication, then lysed with buffer 50 mM Na_2HPO_4, 500 mM NaCl, 50 mM
63 imidazole, 6 M GdmCl pH 7.4. After centrifugation at 18000 rpm for 45 min, supernatants were purified
64 with Ni^{++} affinity chromatography and elution was performed in buffer containing 50 mM $NaPO_4$, 500
65 mM NaCl, 500 mM imidazole, 3 M GdmCl, pH 7.4. Purified CPEB3-IDD was diluted to PBS, 1 M
66 GdmCl pH 7.4, dialysed against PBS pH 7.4 at 4ºC and finally concentrated by using Amicon Ultra-15
67 centrifugal filters 10 kDa cut-off.
68
69 **3. Sequence alignment:** The conservation of vertebrate CPEB3 protein sequences was assessed using
70 the programs T-coffee [81] and Clustal Omega [82] using the default settings. The sequences chosen as
71 representative are human isoform 1 NP_001171608.1, human isoform 2 CAI14105.1, mouse
72 NP_001277755.1, chicken XP_105144323.1, turtle (*Chrysemys picta belli*) XP_005301348.1, frog
73 NP.001015925.1 (*Xenopus tropicalis*) and fish (*Danio rerio*) XP_009305819.1.
74
75 **4. Fluorescence and Circular Dichroism Spectroscopies**. Fluorescence spectra on the full length,
76 unlabeled hCPEB3-IDR at ca. 5 μM in 1.0 mM deuterated acetic acid at pH 4.0 were recorded on a Horiba
77 FluorMax 4 instrument equipped with a Peltier temperature control device using a 0.2 $sec \cdot nm^{-1}$ scan speed
78 and three nm excitation and emission slit widths. The excitation wavelength was 280 nm and the emission
79 was scanned over 300 – 400 nm. A series of spectra were recorded at 2, 10, 20, 30, 40, 50, 60 and 70 ºC.
80 To test for binding of a segment of hCPEB3 rich in proline residues to Profilin, fluorescence
81 spectra were recorded at 20 ºC on 10 μM human Profilin 1 using two nm excitation and emission slits, a

82 2 nm·s^{-1} scan speed, an excitation wavelength of 295 nm and emission was recorded from 300 to 400 nm
83 in the absence and presence 4.5 mM of the peptide acGPPPPAPAPQPnm, which corresponds to residues
84 P166-Q175 of hCPEB3, with an N-terminal acetyl and glycine residue and a C-terminal amide group
85 added to avoid end charges. This assay is based on blue shift and enhanced emission of Trp3 and Trp31
86 of human Profilin 1 as the environment surrounding their indole moieties becomes less solvent exposed
87 upon polyproline ligand binding [34].

88 Fluorescence spectra on Dansyl-MQDDLLMDKSKTGGGGASSSWNTHQ and the control
89 polypeptide without Dansy, MQDDLLMDKSKTGGGGASSSWNTHQ, were recorded using an
90 excitation wavelength of 295 nm, which is selective for the donor Trp, and scanning the emission over
91 300 – 580 nm, using a 2 nm·s^{-1} scan speed and slit widths of 2 nm. The final concentration of both
92 polypeptides was matched at 10.0 μM. Spectra were recorded in 100 mM KCl, 20 mM K_2HPO_4 / KH_2PO_4
93 (pH 7) with or without the denaturant GdmCl present at a final concentration of 7.4 M. Assuming
94 randomized orientations of the donor and acceptor groups and a Förster distance (R_0) of 23.6 angstroms
95 for the Dansyl / Trp pair, their average distance, <r>, can be estimated using the equation <r> =
96 $\sqrt[6]{(R_0^6 - ER_0^6) \,/\, E}$, where E is the fraction of energy transferred from donor to acceptor [83] .

97 A Jasco 810 spectropolarimeter fitted with a Peltier temperature control unit was used to record
98 far UV-CD spectra at 5 and 35 ºC on the complete hCPEP3 IDD at ca. 5 μM in 1 mM deuterated acetic
99 acid (pH 4.0) using a 1.2 nm bandwidth scanning from 260 to 190 nm in a 0.1 cm quartz cuvette at 50
00 nm·min^{-1}. Eight scans were recorded and averaged for each spectrum.
01

02 **5. NMR Spectroscopy: Instrumentation**: All spectra for the hCPEB3 segments were recorded on a
03 Bruker 800 MHz (^{1}H) Avance spectrometer fitted with a triple resonance TCI cryoprobe and Z-gradients.
04 The ^{1}H chemical shift was referenced to 50 μM DSS measured in the same buffer and at the same
05 temperatures as those used for the CPEB3 segments. Since DSS can sometimes bind to intrinsically
06 disordered proteins) [84], the DSS signal was recorded in an independent reference tube containing the
07 same buffer at the same temperature. The ^{13}C and ^{15}N chemical shift references values were calculated
08 by multiplying by their respective gyromagnetic ratios with ^{1}H; that is Ξ ^{13}C / ^{1}H = 0.251449530 and Ξ

09 ^{15}N / ^{1}H = 0.101329118 [85]. NMR spectra were recorded and transformed using TOPSPIN (versions 2.1)
10 (Bruker Biospin).
11
12 **^{13}C-Detection Assignment Strategy**: Among the NMR approaches for studying disordered proteins
13 (reviewed by [86], ^{13}C-detection has been gaining in popularity as it affords the characterization of proline
14 residues [87][88] and offers superior signal dispersion [89]. Here, to speed and improve the assignment
15 of the backbone, we used a "proton-less" NMR approach for segments 1, 4, 5, 6 and 8 based on 2D CON
16 spectra in which successive ^{15}N-^{13}CO nuclei correlations are obtained in two 3D spectra called
17 hacacoNcaNCO and hacaCOncaNCO [90]. For segments 7 and 8, which tend to form condensates [22]
18 and seem to be more rigid, this strategy afforded less intense spectra and in particular about 35 CON
19 crosspeaks were missing. Therefore, an additional strategy based on ^{13}CO connectivities from 3D HNCO
20 and HNcaCO spectra, as well as ^{1}HN and ^{15}N connectivities of consecutive residues from 3D HncocaHN
21 and hNcocaNH spectra [91][92] was utilized to check and complete the backbone assignments. The
22 success of this approach seems to be due to a very slow equilibrium between aggregated protein molecules
23 and those remaining in solution, which are detectable by NMR. The latter strategy was also employed for
24 segment 3, which was less soluble. For all segments, further corroboration was obtained by conventional
25 2D ^{1}H-^{15}N HSQC and 3D HNCO spectra as well as 3D CCCON to confirm the residue identity and obtain
26 the chemical shift values of ^{13}C nuclei of the side chains.
27 Of the eight segments, only segment 2 failed to yield a soluble sample. Whereas the sequence
28 assignments are complete thanks to the analysis of segments 1 and 3, to test for possible end effects, a 20
29 residue peptide, hCPEBpep2, corresponding to residues 91-110 of the hCPEB3 sequence was assigned
30 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
31 NMR spectra at 5.0 ºC on a Bruker 600 MHz spectrometer fitted with a cryoprobe and Z-gradients. A 2D
32 ^{1}H-^{15}N HSQC was also recorded on the 800 MHz Bruker spectrometer. Both the ^{1}H-^{15}N HSQC and the
33 ^{1}H-^{13}C HSQC spectra were recorded at the natural abundance of ^{15}N and ^{13}C. The program NMRFAM-
34 Sparky [93] was used to facilitate manual spectral assignment. The NMR spectral parameters are
35 summarized in **Sup. Table 1**.

36 Theoretical chemical shift values (δ_{coil}) for statistical coil ensembles were calculated using the
37 parameters tabulated by [94] and [95], as implemented on the server at the Bax laboratory. These values
38 were used to calculate conformational chemical shifts ($\Delta\delta$), as the experimentally measured chemical shift
39 (δ_{exp}) minus the calculated chemical shift (δ_{coil}). Segments of five or more residues with $\Delta\delta^{13}C\alpha > 0.3$
40 ppm and $\Delta\delta^{13}CO > 0.3$ were considered to have a significant preference to form an α-helical segment.
41 When appropriate, families of representative preferred conformers were obtained using the program
42 CYANA 3.98 [96]using the chemical shift data to delimit helical segments. The conformers with the
43 lowest energy functions were chosen to be represented in the figures.
44
45 **Coupling Constants**: For segment 5, as an additional, independent test, a 3D HNHA spectrum were
46 recorded and the ratio of the $^{1}H\alpha$-^{1}HN crosspeak to the ^{1}HN-^{1}HN diagonal peak intensities in the 3D
47 HNHA spectrum was utilized to calculate the $^{3}J_{HNCH\alpha}$ coupling constants following the procedure of [97].
48 $^{3}J_{HNCH\alpha}$ coupling constants were also measured for hCPEBpep2 using the 2D ^{1}H-^{1}H COSY spectrum.
49 Utilizing the Karplus equation [98], these $^{3}J_{HNCH\alpha}$ coupling constants can be related at the backbone ϕ
50 angle, which is different for α-helical, statistical coil and β-strands.
51
52 **Relaxation**: To assess the dynamics on the ps–ns time scales, the heteronuclear $^{15}N\{^{1}H\}$ NOE (hNOE)
53 of backbone amide groups was registered as the ratio of spectra recorded with and without saturation in
54 an interleaved mode. Long recycling delays of 13 s were used. Two sets of experiments, one at 25 ºC
55 and one at 5 ºC were recorded at 800 MHz. Uncertainties in peak integrals were determined from the
56 standard deviation of intensities from spectral regions devoid of signal which contain only noise.
57 In addition, $R_1\rho$ relaxation rates, which are sensitive to the presence of preferred, rigid conformers
58 on slower μs-ms timescales were measured by recording two sets of ten ^{1}H-^{15}N correlation spectra with
59 relaxation delays at 8, 300, 36, 76, 900, 100, 500, 156, 200 and 700 ms. One set of experiments was
60 recorded at 25 ºC and the second was recorded at 5 ºC. The relaxation rates were calculated by least-
61 squares fitting of an exponential decay function to the data using NMRPipe [99]. As an additional check,
62 the data were also analyzed independently by using the program DynamicsCenter 2.5.2 (Bruker Biospin).

^{13}C detected relaxation experiments were measured for segment 1 to confirm the presence of a rigid N-terminus and to determine proline residue imino ^{15}N relaxation rates. Transverse relaxation rates (R_2) of hCPEB3 segment 1 imino ^{15}N nuclei were measured by a ^{13}C-detected c_hcacon_nt2_ia3d pulse sequence [87] as a pseudo 3D experiment time composed of nine 2D experiments with relaxation delays of 15.9, 79.2, 158.4, 269.3, 396.0, 554.4, 712.8, 871.2 and 1030 milliseconds over a ^{15}N chemical shift range that is selective for the Pro ^{15}N chemical shifts (132 – 140 ppm). This experiment, and a similar one with a wider sweep width to also measure the ^{15}N T_2 relaxation of all 20 imin/amino-acid residues, were recorded at 25 ºC, without non-uniform sampling or linear prediction. Following Fourier transformation, IPAP virtual decoupling and baseline correction, the peaks were integrated with TOPSPIN 4.0.8 or alternatively NMRPipe by a different operator for comparison. A single exponential decay curve was then fit to the peak integral versus time data to calculate R_2 rates for each position.

Acknowledgments

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. Albert Galera-Prat for help in the initial stages of the study, and to Dr. Miguel Mompeán and Dr. Javier Oroz for critical comments on the manuscript. 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 and PID 2019-109306RB-I00/AEI/10.13039/501100011033 from the Spanish Ministry of Science and Innovation (DVL). The authors declare no competing interests. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

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, RH & DVL. All authors carefully read the MS, provided corrections and approved the final version of the MS.

91

92 **References**

93

94 1. Crick F. Memory and molecular turnover. Nature. 1984;312:101.

95 2. Si K, Giustetto M, Etkin A, Hsu R, Janisiewicz AM, Miniaci MC, et al. A neuronal isoform of
96 CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in
97 aplysia. Cell. 2003;115:893–904.

98 3. Si K, Choi Y-B, White-Grindley E, Majumdar A, Kandel ER. Aplysia CPEB can form prion-like
99 multimers in sensory neurons that contribute to long-term facilitation. Cell. 2010;140:421–35.

00 4. Raveendra BL, Siemer AB, Puthanveettil SV, Hendrickson WA, Kandel ER, McDermott AE.
01 Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB.
02 Nat Struct Mol Biol. 2013;20:495–501.

03 5. Hervas R, Rau MJ, Park Y, Zhang W, Murzin AG, Fitzpatrick JAJ, et al. Cryo-EM structure of
04 a neuronal functional amyloid implicated in memory persistence in Drosophila. Science.
05 2020;367:1230–4.

06 6. Majumdar A, Cesario WC, White-Grindley E, Jiang H, Ren F, Khan MR, et al. Critical role of
07 amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. Cell. 2012;148:515–29.

08 7. White-Grindley E, Li L, Mohammad Khan R, Ren F, Saraf A, Florens L, et al. Contribution of
09 Orb2A stability in regulated amyloid-like oligomerization of Drosophila Orb2. PLoS Biol.
10 2014;12:e1001786.

11 8. Khan MR, Li L, Pérez-Sánchez C, Saraf A, Florens L, Slaughter BD, et al. Amyloidogenic
12 Oligomerization Transforms Drosophila Orb2 from a Translation Repressor to an Activator. Cell.
13 2015;163:1468–83.

14 9. Gill J, Park Y, McGinnis JP, Perez-Sanchez C, Blanchette M, Si K. Regulated Intron Removal
15 Integrates Motivational State and Experience. Cell. 2017;169:836-848.e15.

16 10. Hervás R, Li L, Majumdar A, Fernández-Ramírez MDC, Unruh JR, Slaughter BD, et al.

17 Molecular Basis of Orb2 Amyloidogenesis and Blockade of Memory Consolidation. PLoS Biol.
18 2016;14:e1002361.
19 11. Oroz J, Félix SS, Cabrita EJ, Laurents DV. Structural transitions in Orb2 prion-like domain
20 relevant for functional aggregation in memory consolidation. J Biol Chem. 2020;295:18122–33.
21 12. Fioriti L, Myers C, Huang Y-Y, Li X, Stephan JS, Trifilieff P, et al. The Persistence of
22 Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein
23 CPEB3. Neuron. 2015;86:1433–48.
24 13. Stephan JS, Fioriti L, Lamba N, Colnaghi L, Karl K, Derkatch IL, et al. The CPEB3 Protein Is
25 a Functional Prion that Interacts with the Actin Cytoskeleton. Cell Rep. 2015;11:1772–85.
26 14. Drisaldi B, Colnaghi L, Fioriti L, Rao N, Myers C, Snyder AM, et al. SUMOylation Is an
27 Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. Cell Rep.
28 2015;11:1694–702.
29 15. Chao H-W, Lai Y-T, Lu Y-L, Lin C, Mai W, Huang Y-S. NMDAR signaling facilitates the
30 IPO5-mediated nuclear import of CPEB3. Nucleic Acids Res. 2012;40:8484–98.
31 16. Terlau H, Seifert W. Fibroblast Growth Factor Enhances Long-term Potentiation in the
32 Hippocampal Slice. Eur J Neurosci. 1990;2:973–7.
33 17. Peng S-C, Lai Y-T, Huang H-Y, Huang H-D, Huang Y-S. A novel role of CPEB3 in regulating
34 EGFR gene transcription via association with Stat5b in neurons. Nucleic Acids Res. 2010;38:7446–
35 57.
36 18. Tsuda K, Kuwasako K, Nagata T, Takahashi M, Kigawa T, Kobayashi N, et al. Novel RNA
37 recognition motif domain in the cytoplasmic polyadenylation element binding protein 3. Proteins.
38 2014;82:2879–86.
39 19. Huang Y-S, Kan M-C, Lin C-L, Richter JD. CPEB3 and CPEB4 in neurons: analysis of RNA-
40 binding specificity and translational control of AMPA receptor GluR2 mRNA. EMBO J.
41 2006;25:4865–76.

42 20. Ford L, Ling E, Kandel ER, Fioriti L. CPEB3 inhibits translation of mRNA targets by localizing
43 them to P bodies. Proc Natl Acad Sci U S A. 2019;116:18078–87.
44 21. Vogler C, Spalek K, Aerni A, Demougin P, Müller A, Huynh K-D, et al. CPEB3 is associated
45 with human episodic memory. Front Behav Neurosci. 2009;3:4.
46 22. de Mingo DR, López-García P, Hervás R, Laurents DV, Carrión-Vázquez M. Molecular
47 Determinants of Liquid Demixing and Amyloidogenesis in Human CPEB3. bioRxiv.
48 2020;:2020.06.02.129783.
49 23. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate
50 protein structure prediction with AlphaFold. Nature. 2021;596:583–9.
51 24. Alston RW, Urbanikova L, Sevcik J, Lasagna M, Reinhart GD, Scholtz JM, et al. Contribution
52 of single tryptophan residues to the fluorescence and stability of ribonuclease Sa. Biophys J.
53 2004;87:4036–47.
54 25. Denning DP, Patel SS, Uversky V, Fink AL, Rexach M. Disorder in the nuclear pore complex:
55 the FG repeat regions of nucleoporins are natively unfolded. Proc Natl Acad Sci U S A.
56 2003;100:2450–5.
57 26. Bundi A, Wüthrich K. 1H-NMR parameters of the common amino acid residues measured in
58 aqueous solution of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers. 1979;18:285–
59 97.
60 27. López-Alonso JP, Bruix M, Font J, Ribó M, Vilanova M, Jiménez MA, et al. NMR spectroscopy
61 reveals that RNase A is chiefly denatured in 40% acetic acid: implications for oligomer formation
62 by 3D domain swapping. J Am Chem Soc. 2010;132:1621–30.
63 28. Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, et al. Structural
64 polymorphism of 441-residue tau at single residue resolution. PLoS Biol. 2009;7:e34.
65 29. Muñoz V, Serrano L, Jiménez MA, Rico M. Structural analysis of peptides encompassing all
66 alpha-helices of three alpha/beta parallel proteins: Che-Y, flavodoxin and P21-ras: implications for

67 alpha-helix stability and the folding of alpha/beta parallel proteins. J Mol Biol. 1995;247:648–69.
68 30. Bhattacharyya A, Thakur AK, Chellgren VM, Thiagarajan G, Williams AD, Chellgren BW, et
69 al. Oligoproline effects on polyglutamine conformation and aggregation. J Mol Biol.
70 2006;355:524–35.
71 31. Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M, Laurents DV. The
72 Singular NMR Fingerprint of a Polyproline II Helical Bundle. J Am Chem Soc. 2018;140:16988–
73 7000.
74 32. Reselammal DS, Pinhero F, Sharma R, Oliyantakath Hassan MS, Srinivasula SM, Vijayan V.
75 Mapping the Fibril Core of the Prion Subdomain of the Mammalian CPEB3 that is Involved in
76 Long Term Memory Retention. J Mol Biol. 2021;433:167084.
77 33. Kelly MA, Chellgren BW, Rucker AL, Troutman JM, Fried MG, Miller AF, et al. Host-guest
78 study of left-handed polyproline II helix formation. Biochemistry. 2001;40:14376–83.
79 34. Metzler WJ, Bell AJ, Ernst E, Lavoie TB, Mueller L. Identification of the poly-L-proline-
80 binding site on human profilin. J Biol Chem. 1994;269:4620–5.
81 35. Petrella EC, Machesky LM, Kaiser DA, Pollard TD. Structural requirements and
82 thermodynamics of the interaction of proline peptides with profilin. Biochemistry. 1996;35:16535–
83 43.
84 36. Aurora R, Rose GD. Helix capping. Protein Sci. 1998;7:21–38.
85 37. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in
86 alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci. 1994;3:843–
87 52.
88 38. Chakrabartty A, Baldwin RL. Stability of alpha-helices. Adv Protein Chem. 1995;46:141–76.
89 39. Andrew CD, Warwicker J, Jones GR, Doig AJ. Effect of phosphorylation on alpha-helical
90 stability as a function of position. Biochemistry. 2002;41:1897–905.
91 40. Morón-Oset J, Supèr T, Esser J, Isaacs AM, Grönke S, Partridge L. Glycine-alanine dipeptide

92 repeats spread rapidly in a repeat length- and age-dependent manner in the fly brain. Acta
93 Neuropathol Commun. 2019;7:209.
94 41. Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, et al. C9orf72 repeat
95 expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science.
96 2014;345:1192–4.
97 42. Kaczmarczyk L, Labrie-Dion É, Sehgal K, Sylvester M, Skubal M, Josten M, et al. New
98 Phosphospecific Antibody Reveals Isoform-Specific Phosphorylation of CPEB3 Protein. PLoS
99 One. 2016;11:e0150000.
00 43. Chong PA, Vernon RM, Forman-Kay JD. RGG/RG Motif Regions in RNA Binding and Phase
01 Separation. J Mol Biol. 2018;430:4650–65.
02 44. Hyman SE. Addiction: a disease of learning and memory. Am J Psychiatry. 2005;162:1414–22.
03 45. Mackenzie IR, Rademakers R, Neumann M. TDP-43 and FUS in amyotrophic lateral sclerosis
04 and frontotemporal dementia. Lancet Neurol. 2010;9:995–1007.
05 46. Nelson PT, Dickson DW, Trojanowski JQ, Jack CR, Boyle PA, Arfanakis K, et al. Limbic-
06 predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. Brain.
07 2019;142:1503–27.
08 47. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary
09 rate heterogeneity in proteins with long disordered regions. J Mol Evol. 2002;55:104–10.
10 48. Jayaraman M, Kodali R, Sahoo B, Thakur AK, Mayasundari A, Mishra R, et al. Slow amyloid
11 nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin
12 fragments. J Mol Biol. 2012;415:881–99.
13 49. Eftekharzadeh B, Piai A, Chiesa G, Mungianu D, García J, Pierattelli R, et al. Sequence Context
14 Influences the Structure and Aggregation Behavior of a PolyQ Tract. Biophys J. 2016;110:2361–
15 6.
16 50. Hong J-Y, Wang D-D, Xue W, Yue H-W, Yang H, Jiang L-L, et al. Structural and dynamic

17 studies reveal that the Ala-rich region of ataxin-7 initiates α-helix formation of the polyQ tract but
18 suppresses its aggregation. Sci Rep. 2019;9:7481.
19 51. Lim L, Wei Y, Lu Y, Song J. ALS-Causing Mutations Significantly Perturb the Self-Assembly
20 and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43.
21 PLoS Biol. 2016;14:e1002338.
22 52. Pelassa I, Corà D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. Association of polyalanine
23 and polyglutamine coiled coils mediates expansion disease-related protein aggregation and
24 dysfunction. Hum Mol Genet. 2014;23:3402–20.
25 53. Hervás R, Del Carmen Fernández-Ramírez M, Galera-Prat A, Suzuki M, Nagai Y, Bruix M, et
26 al. Divergent CPEB prion-like domains reveal different assembly mechanisms for a generic
27 amyloid-like fold. BMC Biol. 2021;19:43.
28 54. Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation
29 Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure.
30 2016;24:1537–49.
31 55. Escobedo A, Topal B, Kunze MBA, Aranda J, Chiesa G, Mungianu D, et al. Side chain to main
32 chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun.
33 2019;10:2034.
34 56. Choi Y-B, Kadakkuzha BM, Liu X-A, Akhmedov K, Kandel ER, Puthanveettil SV. Huntingtin
35 is critical both pre- and postsynaptically for long-term learning-related synaptic plasticity in
36 Aplysia. PLoS One. 2014;9:e103004.
37 57. Cao Q, Boyer DR, Sawaya MR, Ge P, Eisenberg DS. Cryo-EM structures of four polymorphic
38 TDP-43 amyloid cores. Nat Struct Mol Biol. 2019;26:619–27.
39 58. Arseni D, Hasegawa M, Murzin AG, Kametani F, Arai M, Yoshida M, et al. Structure of
40 pathological TDP-43 filaments from ALS with FTLD. Nature. 2021.
41 https://doi.org/10.1038/s41586-021-04199-3.

42 59. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, et al. X-ray
43 structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein
44 enantiomers. J Am Chem Soc. 2008;130:9695–701.
45 60. Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, et al. Pol II
46 phosphorylation regulates a switch between transcriptional and splicing condensates. Nature.
47 2019;572:543–8.
48 61. Xu D, Farmer A, Collett G, Grishin NV, Chook YM. Sequence and structural analyses of
49 nuclear export signals in the NESdb database. Mol Biol Cell. 2012;23:3677–93.
50 62. de Araujo ED, Erdogan F, Neubauer HA, Meneksedag-Erol D, Manaswiyoungkul P, Eram MS,
51 et al. Structural and functional consequences of the STAT5B(N642H) driver mutation. 2019.
52 63. Corwin T, Woodsmith J, Apelt F, Fontaine J-F, Meierhofer D, Helmuth J, et al. Defining Human
53 Tyrosine Kinase Phosphorylation Networks Using Yeast as an In Vivo Model Substrate. Cell Syst.
54 2017;5:128-139.e4.
55 64. Liu BA, Jablonowski K, Shah EE, Engelmann BW, Jones RB, Nash PD. SH2 domains
56 recognize contextual peptide sequence information to determine selectivity. Mol Cell Proteomics.
57 2010;9:2391–404.
58 65. Gasset-Rosa F, Lu S, Yu H, Chen C, Melamed Z, Guo L, et al. Cytoplasmic TDP-43 De-mixing
59 Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and
60 Cell Death. Neuron. 2019;102:339-357.e7.
61 66. Li H-R, Chiang W-C, Chou P-C, Wang W-J, Huang J-R. TAR DNA-binding protein 43 (TDP-
62 43) liquid-liquid phase separation is mediated by just a few aromatic residues. J Biol Chem.
63 2018;293:6090–8.
64 67. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA. Essential role of coiled coils for aggregation
65 and activity of Q/N-rich prions and PolyQ proteins. Cell. 2010;143:1121–35.
66 68. Lilliu E, Villeri V, Pelassa I, Cesano F, Scarano D, Fiumara F. Polyserine repeats promote

67 coiled coil-mediated fibril formation and length-dependent protein aggregation. J Struct Biol.
68 2018;204:572–84.
69 69. Polling S, Ormsby AR, Wood RJ, Lee K, Shoubridge C, Hughes JN, et al. Polyalanine
70 expansions drive a shift into α-helical clusters without amyloid-fibril formation. Nat Struct Mol
71 Biol. 2015;22:1008–15.
72 70. Graether SP, Slupsky CM, Sykes BD. Freezing of a fish antifreeze protein results in amyloid
73 fibril formation. Biophys J. 2003;84:552–7.
74 71. Hamley IW, Kirkham S, Dehsorkhi A, Castelletto V, Adamcik J, Mezzenga R, et al. Self-
75 assembly of a model peptide incorporating a hexa-histidine sequence attached to an oligo-alanine
76 sequence, and binding to gold NTA/nickel nanoparticles. Biomacromolecules. 2014;15:3412–20.
77 72. Albrecht A, Mundlos S. The other trinucleotide repeat: polyalanine expansion disorders. Curr
78 Opin Genet Dev. 2005;15:285–93.
79 73. Ferron F, Rebowski G, Lee SH, Dominguez R. Structural basis for the recruitment of profilin-
80 actin complexes during filament elongation by Ena/VASP. EMBO J. 2007;26:4597–606.
81 74. Basu S, Lamprecht R. The Role of Actin Cytoskeleton in Dendritic Spines in the Maintenance
82 of Long-Term Memory. Front Mol Neurosci. 2018;11:143.
83 75. Salehi-Ashtiani K, Lupták A, Litovchick A, Szostak JW. A genomewide search for ribozymes
84 reveals an HDV-like sequence in the human CPEB3 gene. Science. 2006;313:1788–92.
85 76. Huang W-H, Chao H-W, Tsai L-Y, Chung M-H, Huang Y-S. Elevated activation of CaMKIIα
86 in the CPEB3-knockout hippocampus impairs a specific form of NMDAR-dependent synaptic
87 depotentiation. Front Cell Neurosci. 2014;8:367.
88 77. Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. J
89 Biomol NMR. 2001;20:71–5.
90 78. Nallamsetty S, Kapust RB, Tözsér J, Cherry S, Tropea JE, Copeland TD, et al. Efficient site-
91 specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro.

92 Protein Expr Purif. 2004;38:108–15.
93 79. Li M, Liu J, Ran X, Fang M, Shi J, Qin H, et al. Resurrecting abandoned proteins with pure
94 water: CD and NMR studies of protein fragments solubilized in salt-free water. Biophys J.
95 2006;91:4201–9.
96 80. Mompeán M, Romano V, Pantoja-Uceda D, Stuani C, Baralle FE, Buratti E, et al. The TDP-43
97 N-terminal domain structure at high resolution. FEBS J. 2016;283:1242–60.
98 81. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, et al. T-Coffee:
99 a web server for the multiple sequence alignment of protein and RNA sequences using structural
00 information and homology extension. Nucleic Acids Res. 2011;39 Web Server issue:W13-17.
01 82. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of
02 high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.
03 2011;7:539.
04 83. Lakowicz JR, Gryczynski I, Wiczk W, Laczko G, Prendergast FC, Johnson ML.
05 Conformational distributions of melittin in water/methanol mixtures from frequency-domain
06 measurements of nonradiative energy transfer. Biophys Chem. 1990;36:99–115.
07 84. Diez-García F, Gómez-Pinto I, Chakrabartty A, González C, Laurents DV. Conformation
08 specificity and arene binding in a peptide composed only of Lys, Ile, Ala and Gly. Eur Biophys J.
09 2012;41:63–72.
10 85. Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, et al. Recommendations for
11 the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-
12 Union Task Group on the standardization of data bases of protein and nucleic acid structures
13 determined by NMR spectroscopy. Eur J Biochem. 1998;256:1–15.
14 86. Gibbs EB, Cook EC, Showalter SA. Application of NMR to studies of intrinsically disordered
15 proteins. Arch Biochem Biophys. 2017;628:57–70.
16 87. Murrali MG, Piai A, Bermel W, Felli IC, Pierattelli R. Proline Fingerprint in Intrinsically

17 Disordered Proteins. Chembiochem. 2018;19:1625–9.
18 88. Mateos B, Conrad-Billroth C, Schiavina M, Beier A, Kontaxis G, Konrat R, et al. The
19 Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder
20 Promoters to Compaction Facilitators. J Mol Biol. 2020;432:3093–111.
21 89. Cook EC, Usher GA, Showalter SA. The Use of (13)C Direct-Detect NMR to Characterize
22 Flexible and Disordered Proteins. Methods Enzymol. 2018;611:81–100.
23 90. Pantoja-Uceda D, Santoro J. New 13C-detected experiments for the assignment of intrinsically
24 disordered proteins. J Biomol NMR. 2014;59:43–50.
25 91. Sun Z-YJ, Frueh DP, Selenko P, Hoch JC, Wagner G. Fast assignment of 15N-HSQC peaks
26 using high-resolution 3D HNcocaNH experiments with non-uniform sampling. J Biomol NMR.
27 2005;33:43–50.
28 92. Pantoja-Uceda D, Santoro J. Aliasing in reduced dimensionality NMR spectra: (3,2)D HNHA
29 and (4,2)D HN(COCA)NH experiments as examples. J Biomol NMR. 2009;45:351–6.
30 93. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular
31 NMR spectroscopy. Bioinformatics. 2015;31:1325–7.
32 94. Kjaergaard M, Poulsen FM. Sequence correction of random coil chemical shifts: correlation
33 between neighbor correction factors and changes in the Ramachandran distribution. J Biomol NMR.
34 2011;50:157–65.
35 95. Kjaergaard M, Brander S, Poulsen FM. Random coil chemical shift for intrinsically disordered
36 proteins: effects of temperature and pH. J Biomol NMR. 2011;49:139–49.
37 96. Güntert P, Buchner L. Combined automated NOE assignment and structure calculation with
38 CYANA. J Biomol NMR. 2015;62:453–71.
39 97. Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear
40 J(HN-Ha) coupling constants in 15N-enriched proteins. J Am Chem Soc. 115:7772–7.
41 98. Karplus M. Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc.

42 1963;85:2870–1.

43 99. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional
44 spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277–93.

45 100. Spera S, Bax A. Empirical correlation between protein backbone conformation and Calpha
46 and Cbeta 13C NMR chemical shifts. J Am Chem Soc. 1991;113:5490–2.

47 101. Wang Y, Jardetzky O. Probability-based protein secondary structure identification using
48 combined NMR chemical-shift data. Protein Sci. 2002;11:852–61.

49 102. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification
50 of protein secondary structure using 13C chemical-shift data. J Biomol NMR. 1994;4:171–80.

51 103. Cougot N, Bhattacharyya SN, Tapia-Arancibia L, Bordonné R, Filipowicz W, Bertrand E, et
52 al. Dendrites of mammalian neurons contain specialized P-body-like structures that respond to
53 neuronal activation. J Neurosci. 2008;28:13793–804.

54 104. Kandel ER, Dudai Y, Mayford MR. The molecular and systems biology of memory. Cell.
55 2014;157:163–86.

56

57

Figures and Tables

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

```
     1        10         20         30         40         50
S1  MQDDLLMDKS KTQPQPQQQQ RQQQQPQPES SVSEAPSTPL SSETPKPEEN
            60         70         80         90        100
    SAVPALSPAA APPAPNGPDK MQMESPLLPG LSFHQPPQQP PPPQEPAAPG  S2
           110        120        130        140        150
S3  ASLSPSFGST WSTGTTNAVE DSFFQGITPV NGTMLFQNFP HHVNPVFGGT
           160        170        180        190        200
    FSPQIGLAQT QHHQQPPPPA PAPQPAQPAQ PPQAQPPQQR RSPASPSQAP  S4
           210        220        230        240        250
S5  YAQRSAAAAY GHQPIMTSKP SSSSAVAAAA AAAAASSASS SWNTHQSVNA
           260        270        280        290        300
    AWSAPSNPWG GLQAGRDPRR AVGVGVGVGV GVPSPLNPIS PLKKPFSSNV  S6
           310        320        330        340        350
S7  IAPPKFPRAA PLTSKSWMED NAFRTDNGNN LLPFQDRSRP YDTFNLHSLE
           360        370        380        390        400
    NSLMDMIRTD HEPLKGRMGI NFHHPGTDNI MALNSRSSLF PFEDAFLDDS  S8
           410        420        430        440        450
    HGDQAISSGL SSPTRCQNGE RVERYSRKVF VGGLPPDIDE DEITASFRRF
```

Segment 1 spans hCPEB3 residues 1-100; segment 2, residues 51 – 150; segment 3, residues 101 – 200; segment 4, residues 151 – 250; segment 5, residues 201 – 300 (green residues); segment 6, residues 251 – 350; segment 7, residues 302 – 400 and finally segment 8, which is composed of residues 352 – 450. Residues of structural or functional interest are in **bold font**. The putative dimethyl-Arg site (R308) is highlighted in green. The putative phosphoTyr site, Y341 plausible recognized by STAT5B's SH2 domain is highlighted in blue. The predicted nuclear export signal, L349-L353, is highlighted in yellow. Putative phosphorylation sites: S284, S290, S297, S298, S400, S407, S408, S411 and S412 are highlighted in red. Finally, 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) [100].

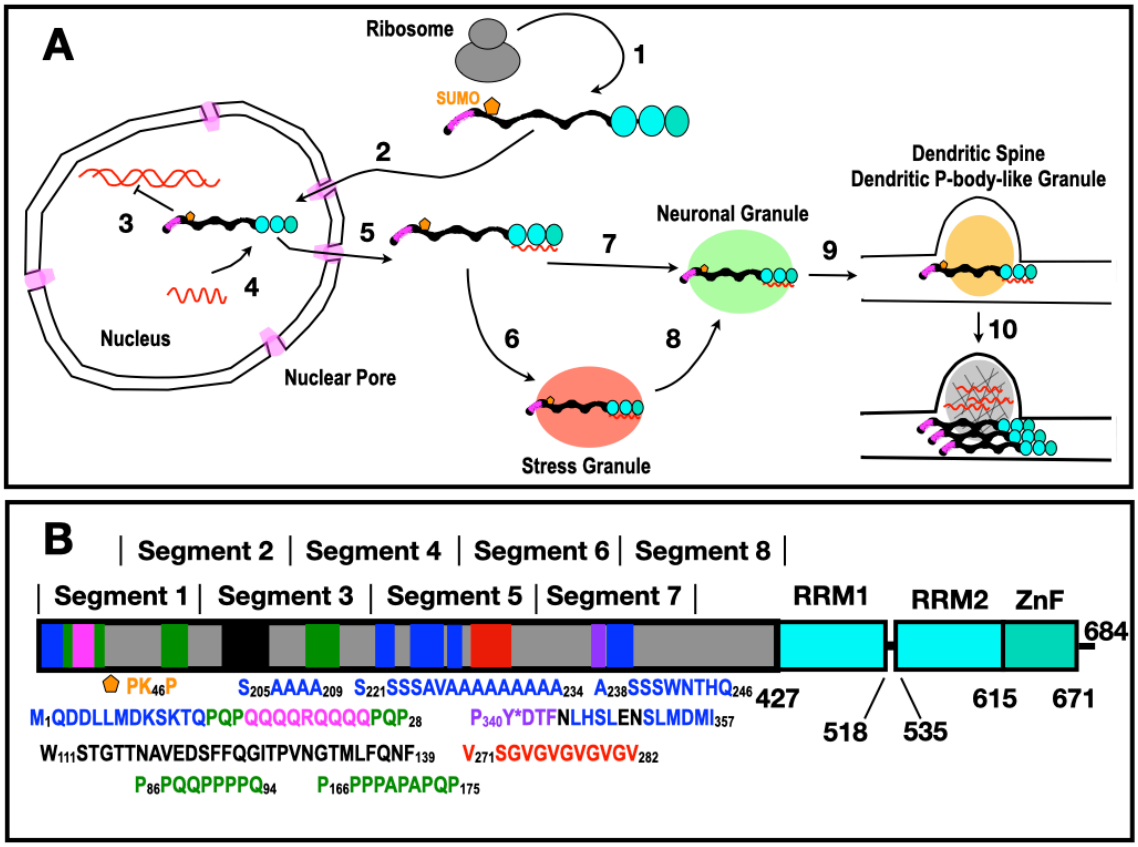
b. from Wang and Jardetzky (2002) [101] for alanine.

c. calculated from Wishart & Skyes (1994) [102].

d. from Treviño *et al.* (2018) [31].

e. This study.

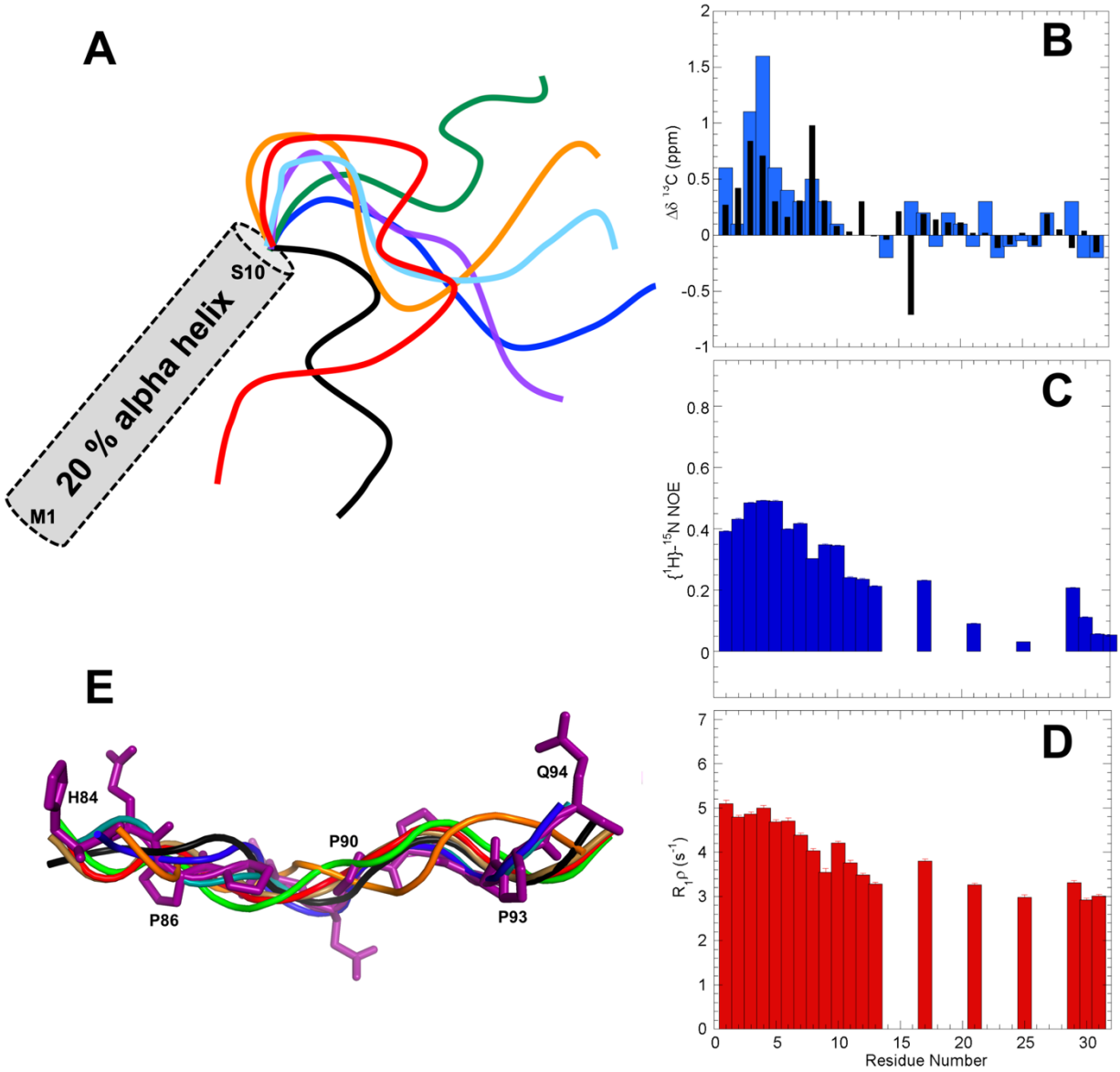
Figure 1. Domain Structure and Activities of hCPEB3



A. **hCPEB3 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 aiding 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) [103]. 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) [104].

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**.

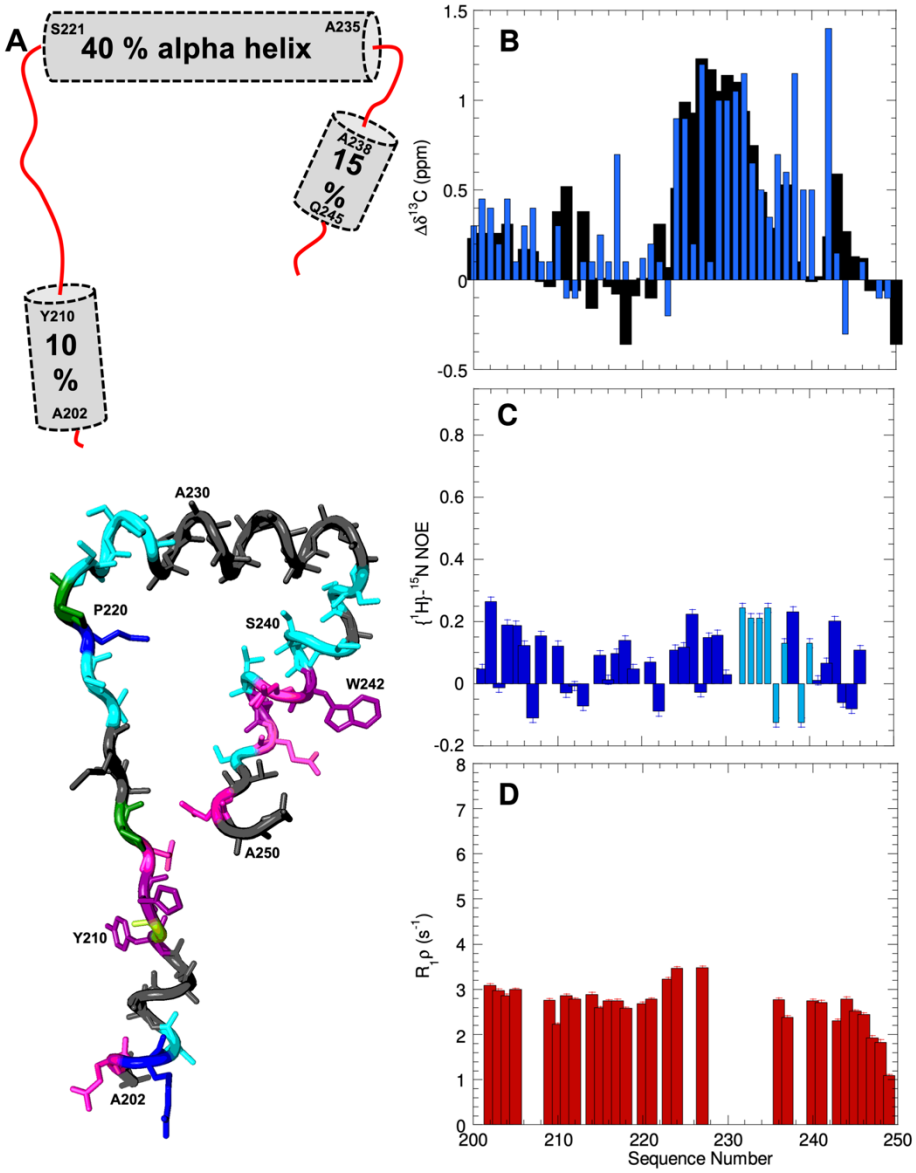
15
16 **Figure 2**: The N-terminal 25 residues of hCPEB3 adopt a hydrophobic α-helix followed by a disordered
17 polyQ segment flanked by PQP mini-breaker motifs.
18



19
20
21 The N-terminus of hCPEB3 contain an α-helix-forming and a disordered amyloidogeneic Q_4RQ_4 segments,
22 which are separated by PQP mini breaker motifs.
23 (**A**) Schematic representation as a gray cylinder of partial (20%) α-helix formation by the first ten residues
24 of hCPEB3. The disordered conformational ensemble of residues 11 -32 is represented curved lines colored
25 purple, blue, cyan, green, orange, red and black . (**B**) ^{13}Cα (**blue**) and ^{13}CO (**black**) conformational chemical
26 shifts indicate a 20% population of helix at 25 ºC. Uncertaintes in the conformational chemical shifts (Δδ)
27 are 0.02 and 0.10 ppm for ^{13}CO and ^{13}Cα, respectively. (**C**) {^{1}H}-^{15}N NOE and (**D**) $R_1\rho$ relaxation

28 measurements indicate that this helical conformation is less mobile than the polyQ segment at ns/ps and
29 μs/ms timescales, respectively at 25 ºC. Error bars are shown in panels (**C**) and (**D**) but are small as the
30 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
31 due to overlap of $^1H^{15}N$ peaks or a lack of $^1H^{15}N$ signals in the case of proline residues. (**E**) Eight
32 representative backbone conformers, colored purple, blue, teal, green, amber, orange, red and black, of the
33 proline rich segment, H84-Q94, featuring a PPII helix that spans residues P86-Q94. All heavy atoms are
34 shown for the purple conformer.
35

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



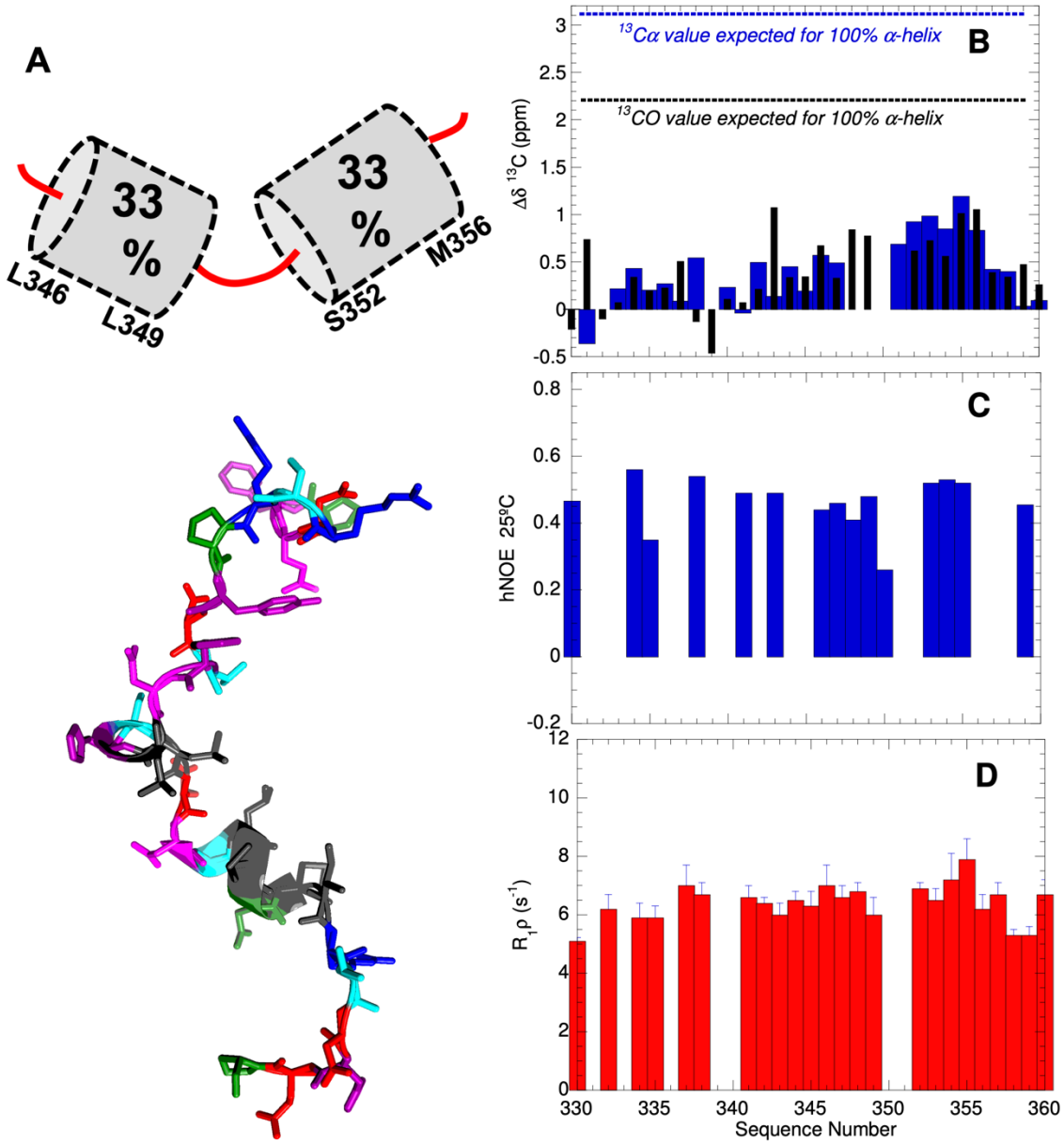
A. (*Top*) Schematic representation as gray cylinders of the three parcially populated helices present in residues 200 - 250. (*Bottom*) One conformer with all three α-helices is shown; residues are colored: cationic residues (R & K) = **blue**, aromatics (F, Y, W & H) = **purple**, anionic (E & D) = **red**, aliphatic (A, I, L, M) = **dark gray**, amyloidogenic (N & Q) = **magenta**, hydroxyl bearing (S & T) = **cyan**, and proline = **green**.

B. ^{13}CO (**black**) and ^{13}Cα (**blue**) conformational chemical shifts (Δδ) of residues 201-250 at 25ºC. Note that the second α-helix which contains nine consecutive Ala residues has a relatively high helical population. Uncertainites in the conformational chemical shifts (Δδ) are 0.02 and 0.10 ppm for ^{13}CO and ^{13}Cα, respectively.

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. 10**).

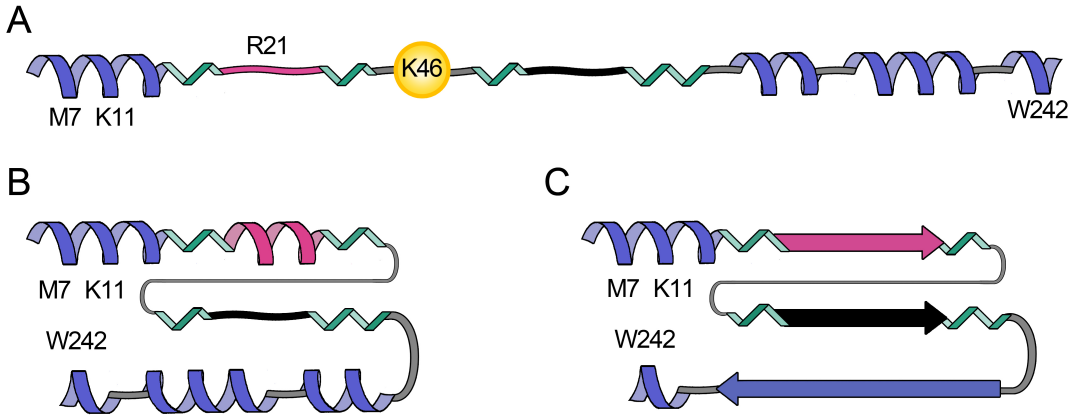
54 **Figure 4**. Conformation of the NES and nearby phospho-Tyr site.



55
56
57 **A**. (*Top*) Residues L346-L349 and S352-M356 adopt two short, partially populated α-helices. (*Bottom*).
58 Representative conformer of shown with cationic residues (R & K) colored **blue**, aromatics (F, Y & H) =
59 **purple**, anionic (E & D) = **red**, aliphatic (I, L, M) = **dark gray**, amyloidogenic (N & Q) = **magenta**,
60 hydroxyl bearing (S & T) = **cyan**, and proline = **green**. Spiral ribbons mark the helical segments spanning
61 residues 346 – 349 and 352 - 356. **B**. Conformational chemical shifts of ^{13}Cα (**blue** bars) and ^{13}CO (**black**
62 narrow bars) afford detection of helical conformations. Uncertaintes in the conformational chemical shifts
63 (Δδ) are 0.02 and 0.10 ppm for ^{13}CO and ^{13}Cα, respectively. **C.** {^{1}H}-^{15}N NOE ratios of 0.85 and -0.20 are
64 indicative of high rigidity and flexibility, respectively, on ps-ns time scales. **D**. Higher $R_{1\rho}$ rates are
65 diagnostic of rigidity on μs-ms time scales.
66

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 silico* 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**) may promote profilin binding and the initiation of a more robust actin filament network.

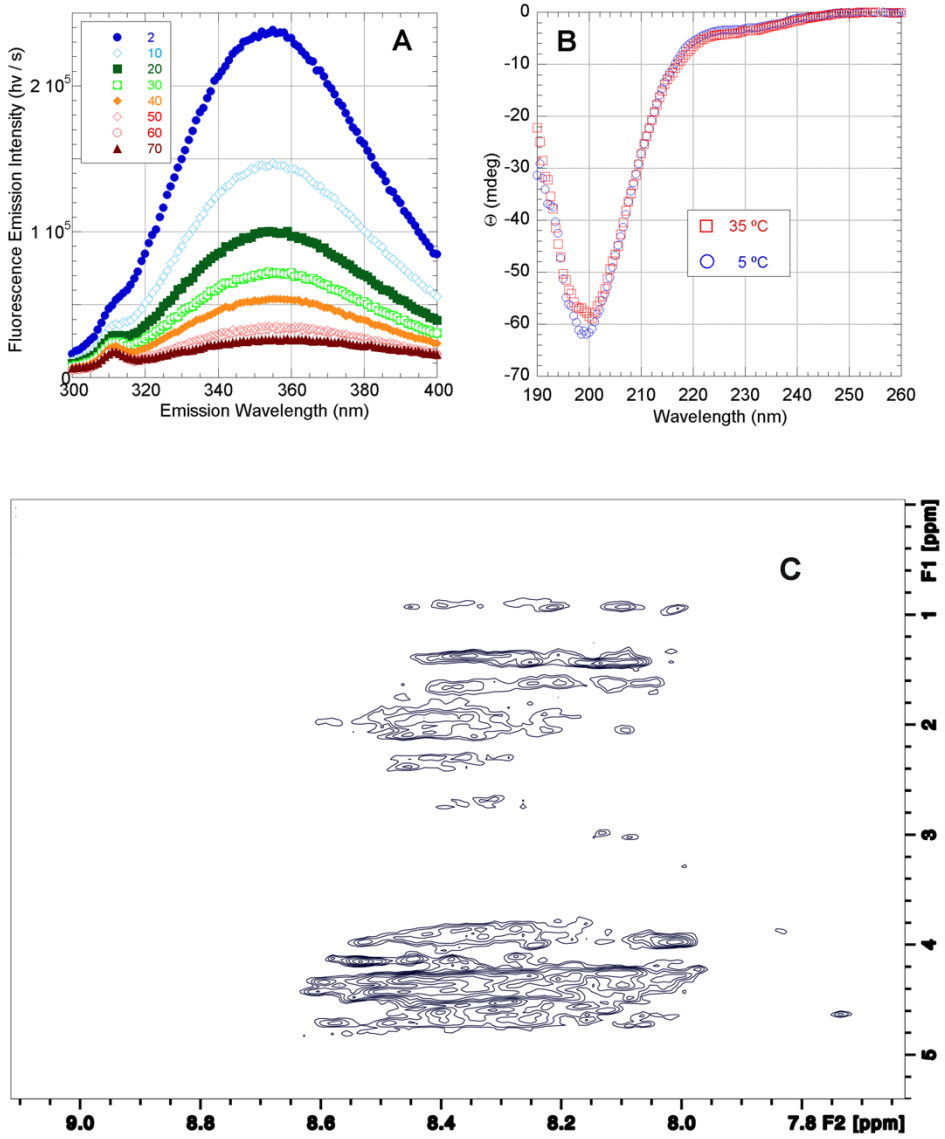
Sup. Table 1: NMR Spectral Parameters

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 ^{1}H x 20 ^{15}N x 10 ^{1}H	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
2D c_hcacon_nt2_ia3d	4	10 ^{13}CO x 33 ^{15}N	700 x 256
2D c_hcacon_nt2_ia3d	8	10 ^{13}CO x 6 ^{15}N (Pro focussed)	1k x 48
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.

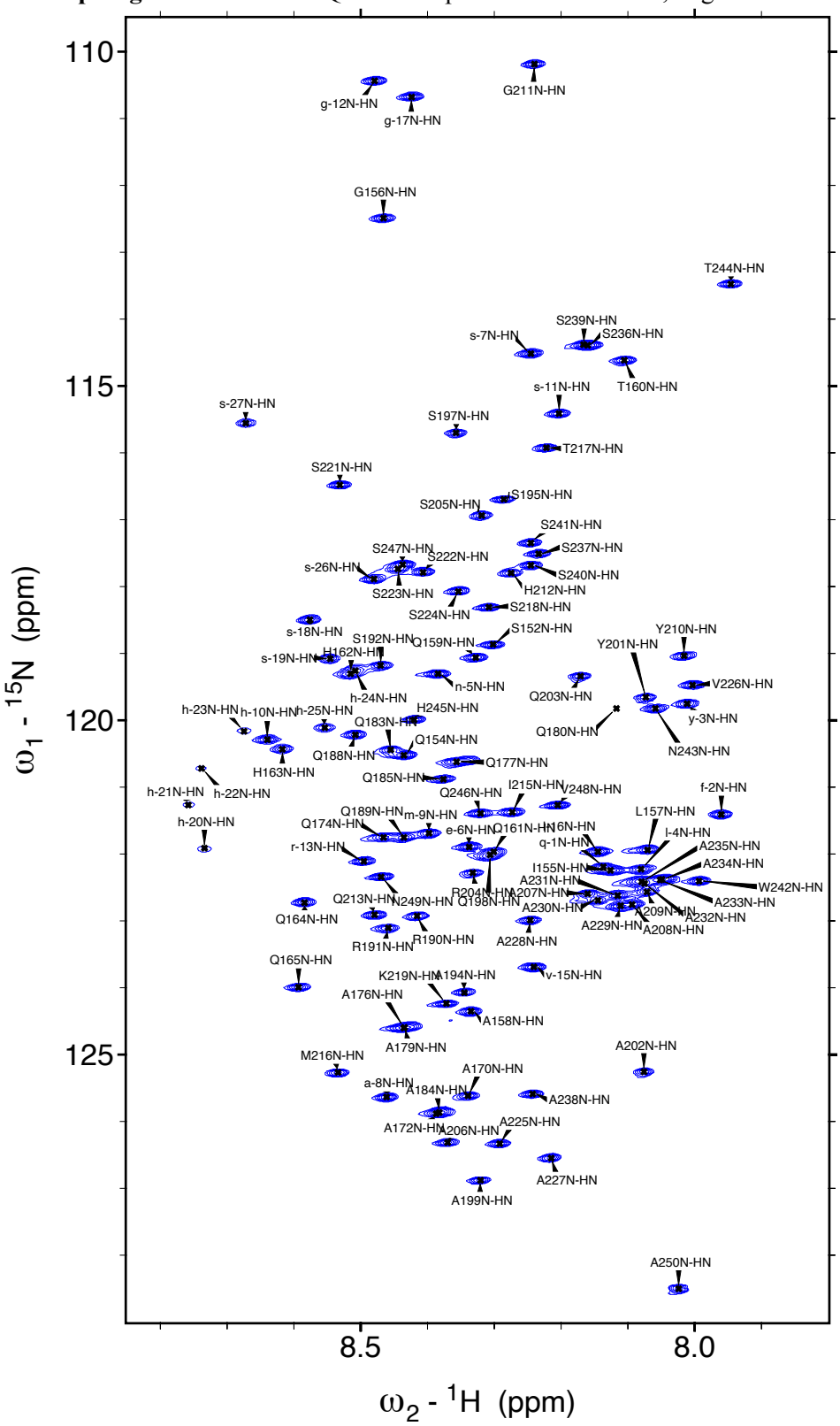
B. Far UV-CD spectra of the IDR of hCPEB3 at 5 ºC (blue open circles) and 35 ºC (red open squares). The observed minimum near 200 nm and the lack of minima near 208 nm, 218 nm and 222 nm and the lack of a maximum at 195 nm are all characteristic spectral features of a statistical coil.

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) [27].

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

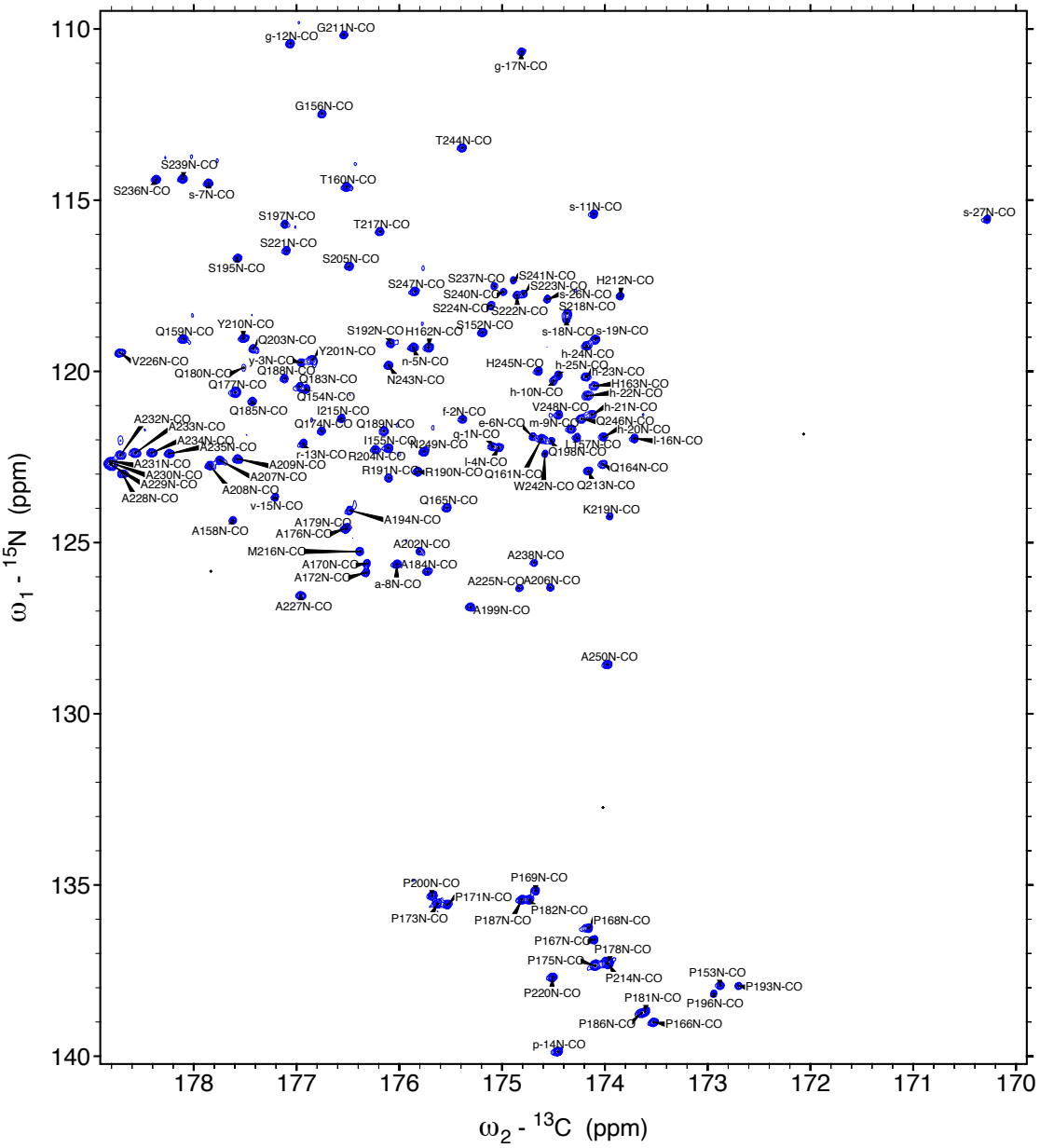
```
Human2    MODDLLMDKSKTOPOPOOOO----ROO----OOPOPESSVSEAPSTPLSSETPKPEENSAVPALSPAAAPPA
Human1    MODDLLMDKSKTOPOPOOOO----ROO----OOPOPESSVSEAPSTPLSSETPKPEENSAVPALSPAAAPPA
Mouse     MODDLLMDKSKTOPOSOOOO----ROQQQQQOOLQPEPGAAEAPSTPLSSEIPKPEDSSAVPALSPASAPPA
Chicken   MODDLLMDKSKAOPOOOROO----QQ-----OOO--DPNSAEAPSTPISSETPKPEDNSSVTSIGT--SAPS
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



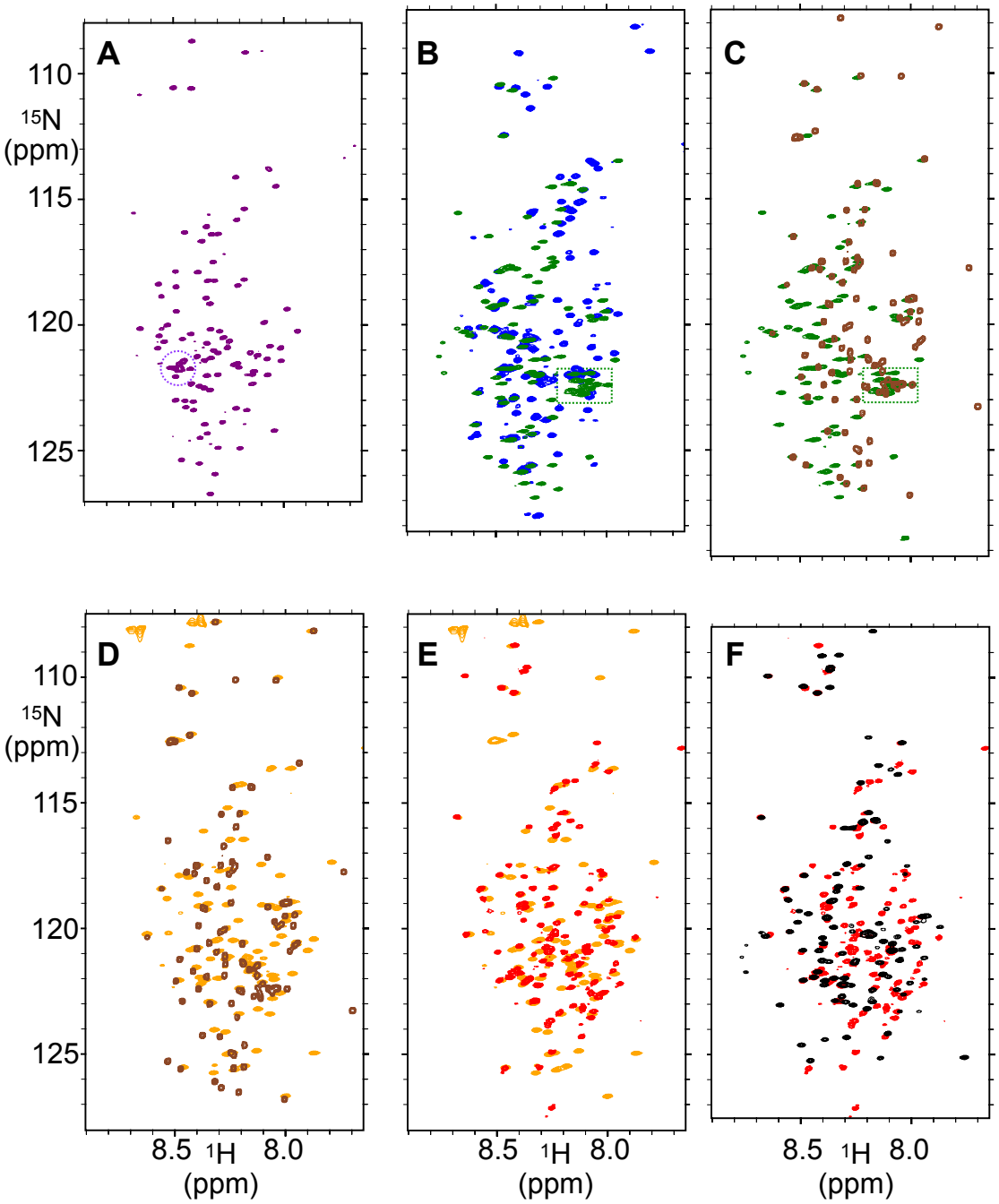
^{1}H-^{15}N correlations were recorded at 25ºC in 1 mM acetic acid, pH 4. Note the cluster of Ala signals in belonging to the polyA stretches near 8.05 ppm ^{1}H and 122.5 ppm ^{15}N. Their position contrasts with those of the isolated Ala residues, whose δ ^{15}N > 125 ppm. Signals labeled in lower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

42 **Sup. Fig. 5** ^{1}H-^{15}N HSQC spectra of hCPEB3 IDR Segments



43

44 2D ^{1}H-^{15}N HSQC spectra of hCPEB3 IDR recorded at pH 4, 25ºC on:

45 A. Segment 1 (**purple**), signals from the polyQ rich segment are circled.

46 B. Segments 3 (**blue**) and 4 (**green**). Signals from the polyA rich segment are boxed.

47 C. Segments 4 (**green**) and 5 (**brown**). Signals from the polyA rich segment are boxed.

48 D. Segments 5 (**brown**) and 6 (**orange**).

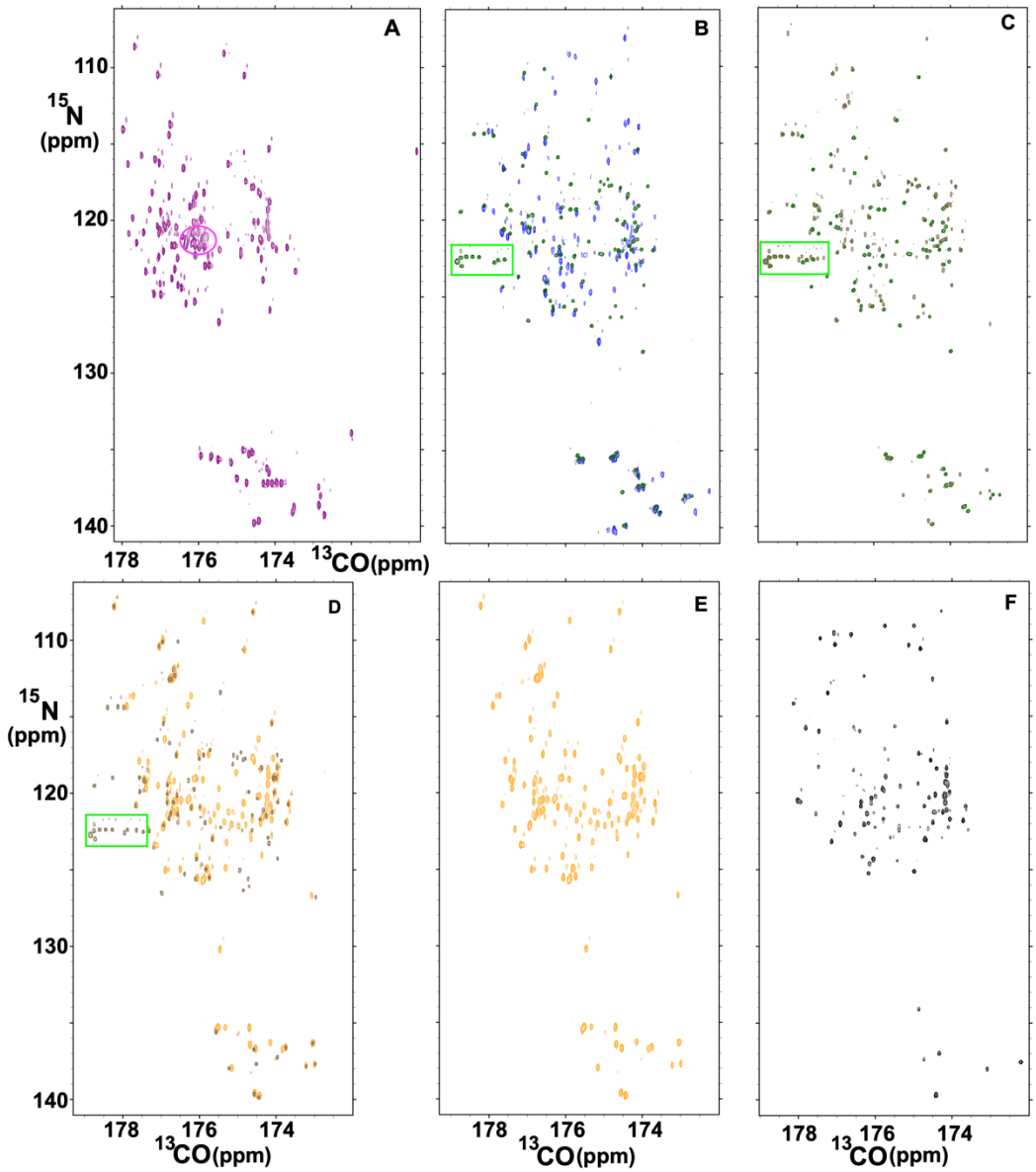
49 E. Segments 6 (**orange**) and 7 (**red**).

50 F. Segments 7 (**red**) and 8 (**black**)

51

52 **Sup. Fig. 6** 2D ^{13}CO-^{15}N spectra of hCPEB3 IDR Segments

53



54

55 2D ^{13}CO-^{15}N HSQC spectra of hCPEB3 IDR recorded at pH 4, 25ºC on:

56 A. Segment 1 (**purple**), signals from the polyQ rich segment are circled in magenta.

57 B. Segments 3 (**blue**) and 4 (**green**). Signals from the polyA rich segment are boxed in **light green**.

58 C. Segments 4 (**green**) and 5 (**brown**). Signals from the polyA rich segment are boxed in **light green**.

59 D. Segments 5 (**brown**) and 6 (**orange**).

60 E. Segment 6 (**orange**).

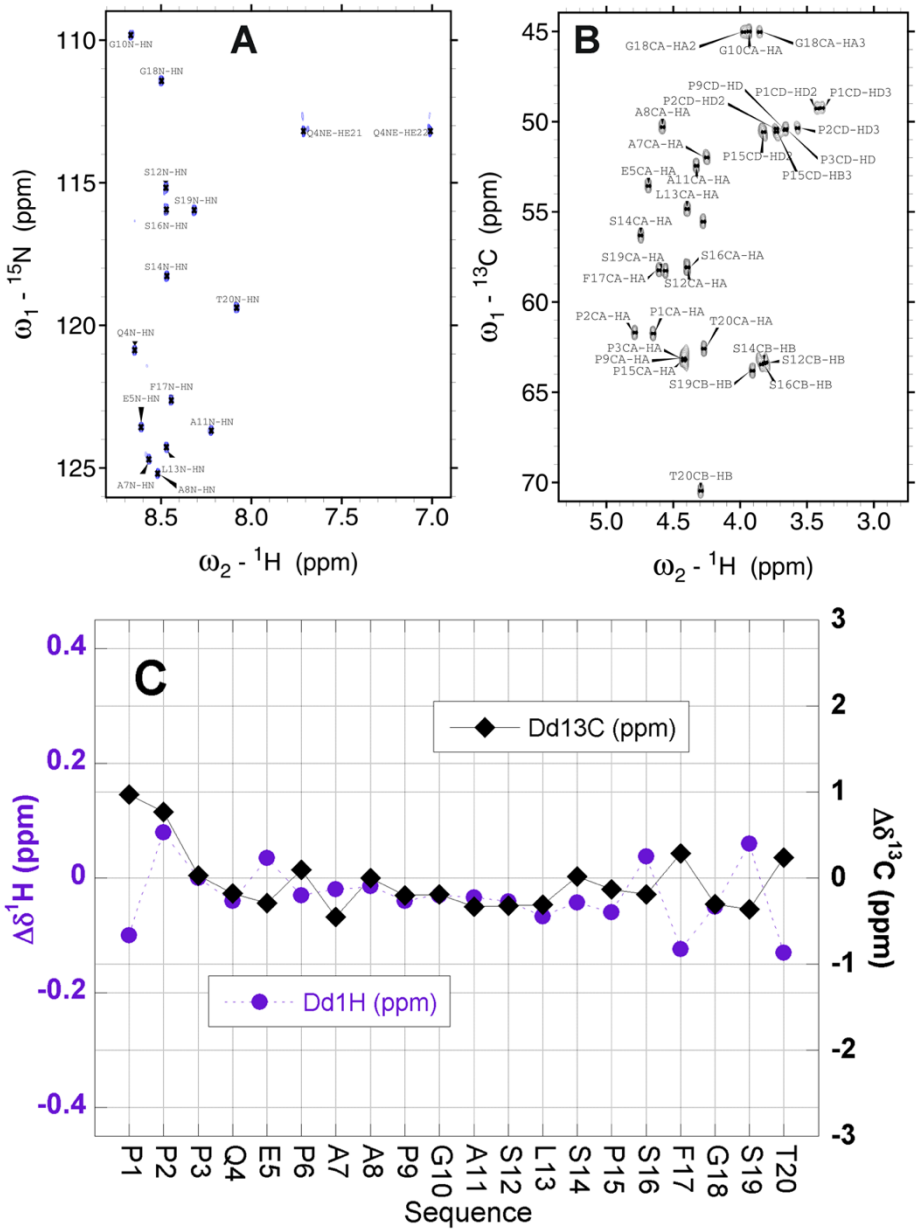
61 F. Segment 8 (**black**)

62 The spectrum of segment 1 is displaced because its optimal spectral range was distinct relative to the other

63 spectra. The weak satellite peaks appearing to the right and above the main peaks arise from deuteration.

64

Sup. Fig. 7. Corroboration of small to negligible populations of α-helix or β-strand conformations in residues 91 - 110 of hCPEB3.

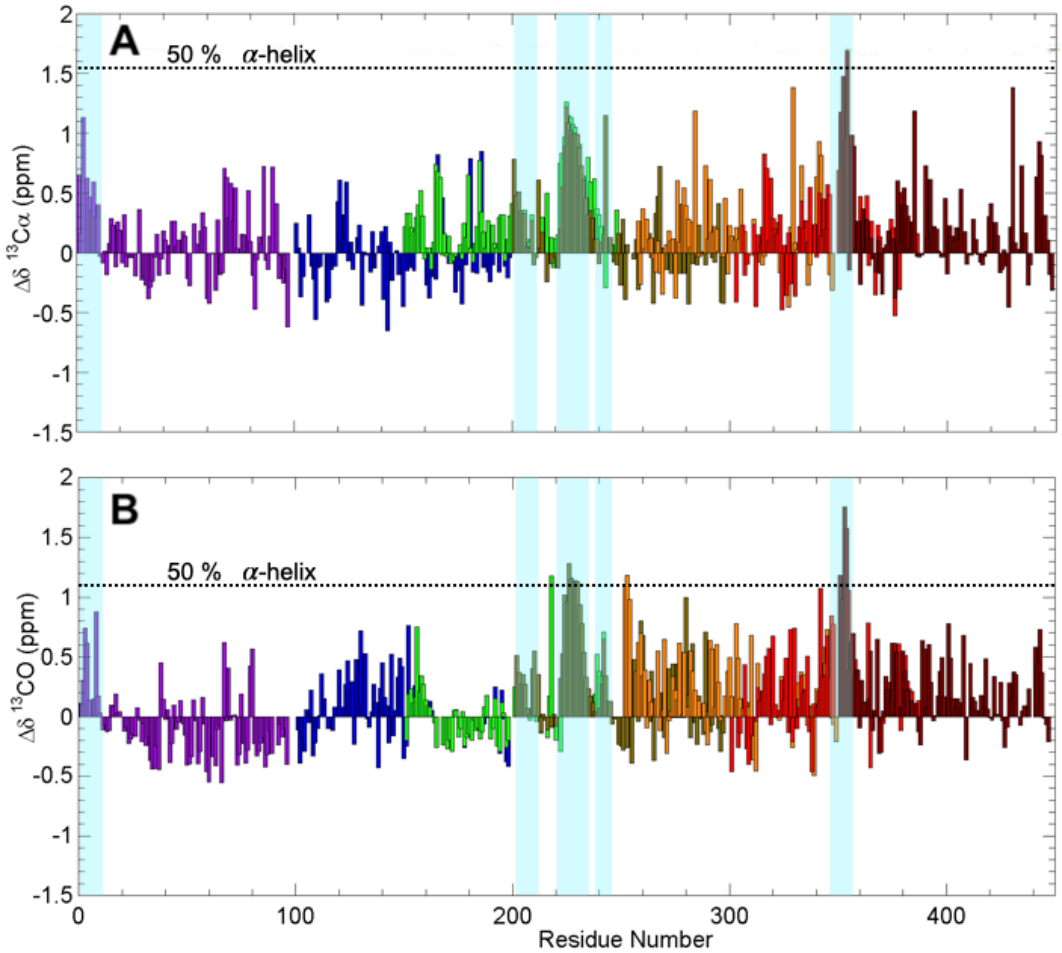


A. The assigned ^{1}H-^{15}N HSQC spectrum of a peptide corresponding to residues 91 - 110 of hCPEB3. The low chemical shift dispersion in the ^{1}HN dimension is a hallmark of a statistical coil. The residue number labels correspond to this peptide, to calculate the residue number in the full length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

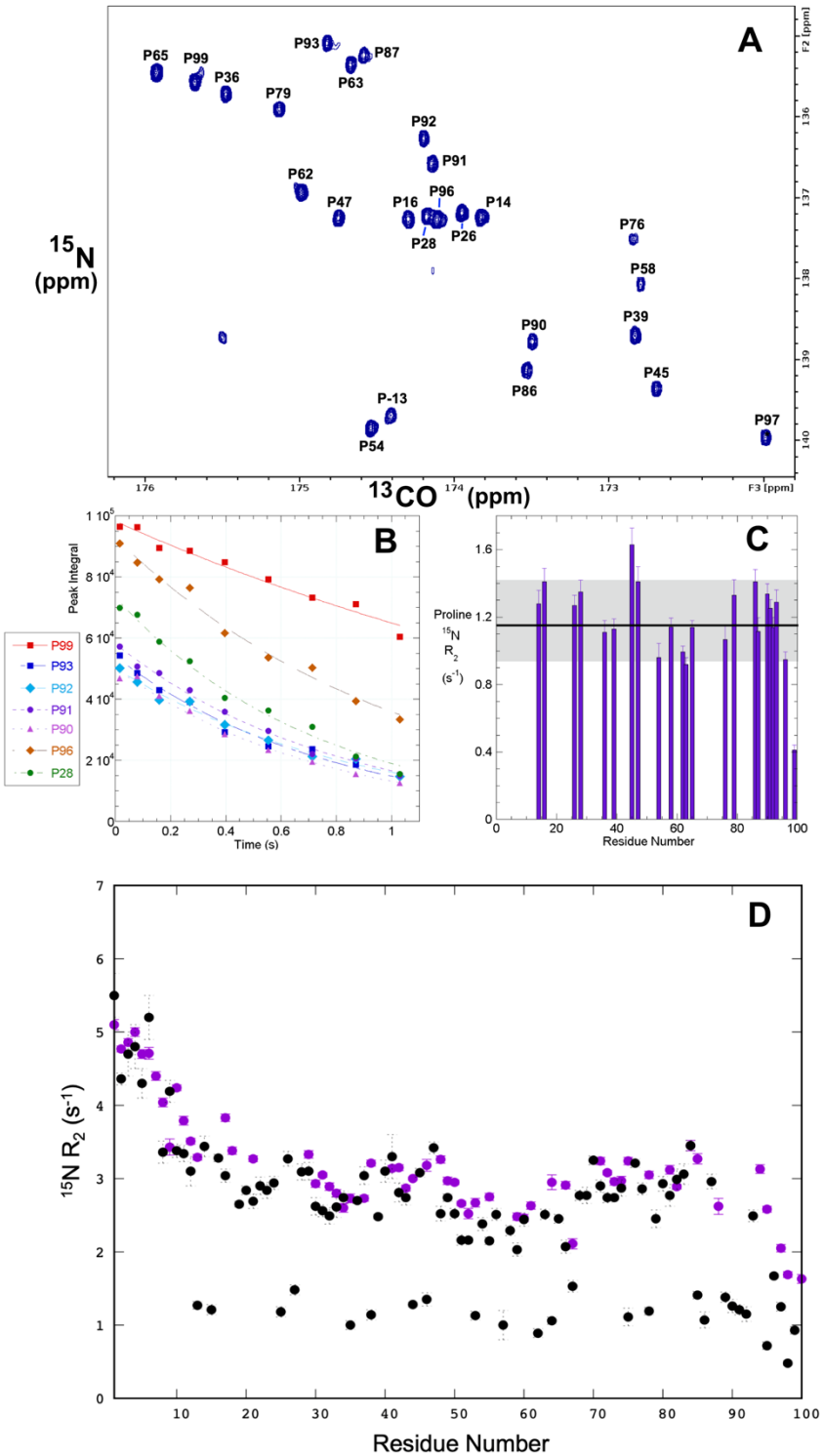
Sup. Fig. 8. Conformational chemical shifts reveal the presence of partially populated secondary structure in the hCPEB3 IDR.



Conformational chemical shifts (Δδ) at 25ºC, calculated as the experimentally measured chemical shift (δ_{exp}) minus the chemical shift expected for statistical coil (δ_{coil}) for ^{13}Cα (*top panel*) and ^{13}CO (*bottom panel*) under these conditions. Data are colored according to segment: 1=**purple**, 3=**blue**, 4=**green**, 5=**brown**, 6=**orange**, 7=**red**, 8=**maroon**. Negative values Δδ values of ^{13}Cα and ^{13}CO are characteristic of extended conformations like β-strands. Values of +1.55 ppm for Δδ ^{13}Cα and +1.1 ppm Δδ ^{13}CO are expected for 50% α-helix; these values represented as dotted lines. Stretches of positive Δδ for ^{13}Cα and Δδ ^{13}CO values which evince partial α-helix formation span residues 1-10, 202-210, 222-234, 238-246 and 346-356 and are shaded cyan.

96 **Sup. Fig. 9** ^{13}C-detected ^{15}N Relaxation Experiments of hCPEB3 segment 1.

97



98

99 **A**. The ^{13}C-detected ^{15}N relaxation experiment obtained using a 15.9 ms delay. Proline ^{15}N atoms

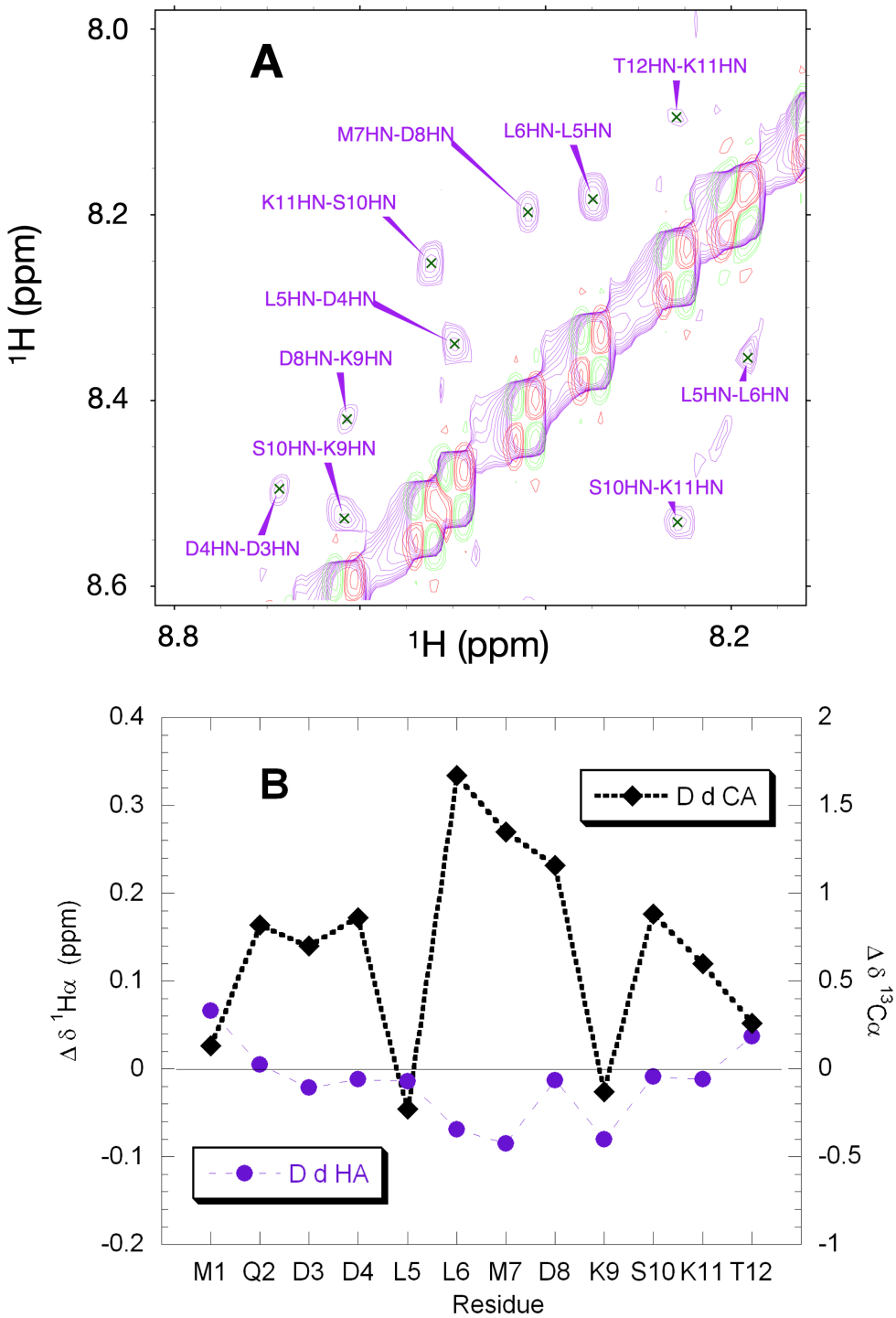
00 are labeled. **B**. Fits of some represantative proline ^{15}N nuclei are shown Peaks "P-13", which comes

01 from the His tag, and P97, whose resonance is folded, were not included in the analysis. The R_2

02 rates and their one standard deviation uncertainties of their fits (error bars) from the Topspin analysis

are shown in **panel C**. NMRPipe analysis yielded similar results (data not shown). The black line marks the mean R_2 rate (1.18 s^{-1}) and the gray shaded area is the one standard deviation (+/- 0.24 s^{-1}) from the mean of all the proline R_2 rates. As previously observed in two recent studies (Murrali *et al.,* (2018); Mateos *et al.,* 2020), the R_2 rates for the proline imine ^{15}N are consistent with *trans* Xaa-Pro peptide bond (Mateos *et al.,* 2020) and are significantly slower relative those for the 19 amino acid residues. The residues near the C-terminus, especially P99, shows slower relaxation which is indicative of high flexibility (**panel C**). In contrast, the prolines of the two "PQP" mini-motifs (P14 & P16) and (P26 & P28) flanking the Q_4RQ_4 motif show slightly higher R_2 rates as do P45 & P47 (which are found within a cluster of charged residues) and P90, P91, P92 & P93 (which adopt a short PPII helix). This is indicative of slightly less flexibility. Similar trends were observed recently in isolated Pro residues and a run of four consecuetive proline residues in the ID4 linker domain of the CREB binding protein (Murrali *et al.,* (2018). Finally the ^{13}C-detected relaxation experiment was repeated for all hCPEB3 segment 1 residues using a broad sweep width and the results are shown in **panel D** as black circles along the results obtained using a ^{15}N detected experiment, which are colored purple. The good agreement is a further confirmation of the presence of a rigid, partially populated α-helix at the N-terminus of Segment 1 (**Figure 2, Sup. Fig. 10 & 12**).

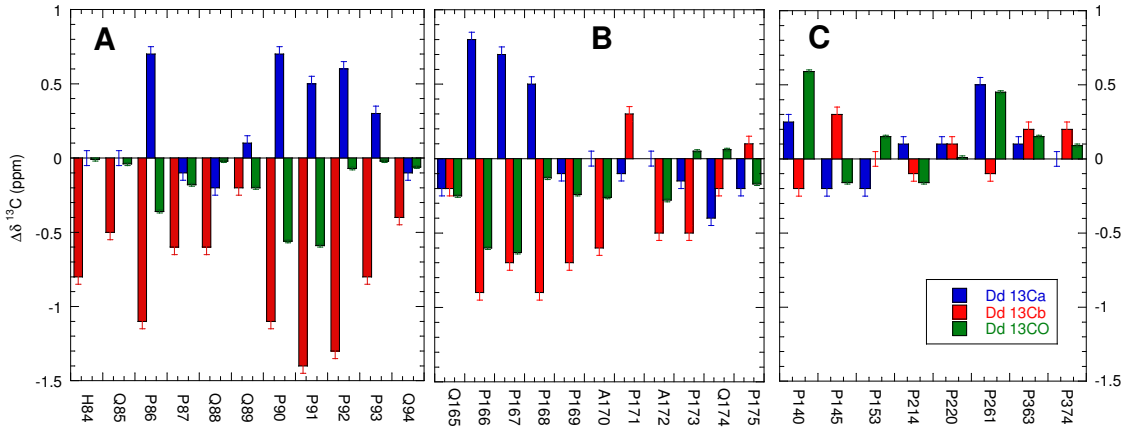
27 **Sup. Fig. 10** Partial Formation of an α-Helix in the N-terminal Residues of hCPEB3

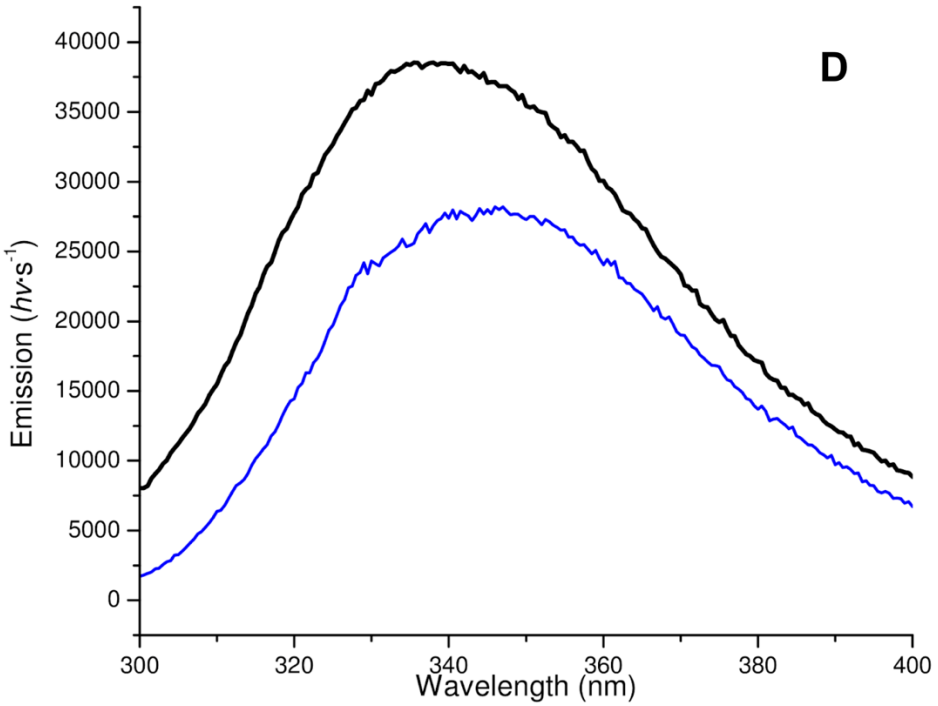


28

29 **A**. 2D ^{1}H-^{1}H NOESY (80 ms mixing time, **purple**) of a peptide corresponding to the initial residues of
30 hCPEB3. Sequential NOE correlations, which are consistent with α-helical structure, are labeled. **Red** and
31 **green** peaks along the diagonal are from the 2D ^{1}H-^{1}H COSY spectrum.
32 **B**. Conformational ^{1}H (**purple**) and ^{13}C (**black**) chemical shifts at 5 ºC for a peptide corresponding to the
33 initial residues of hCPEB3 20% hexafluoroisopropanol (CF_3-CHOH-CF_3) The positive $\Delta\delta^{13}$C and negative
34 $\Delta\delta$ ^{1}H values are indicative of the formation of significant α-helix formation.
35

Sup. Fig. 11 The Consecutive Proline Residues of hCPEB3 Show a Characteristic Pattern of Conformational Chemical Shifts and Weak Binding to Profilin.

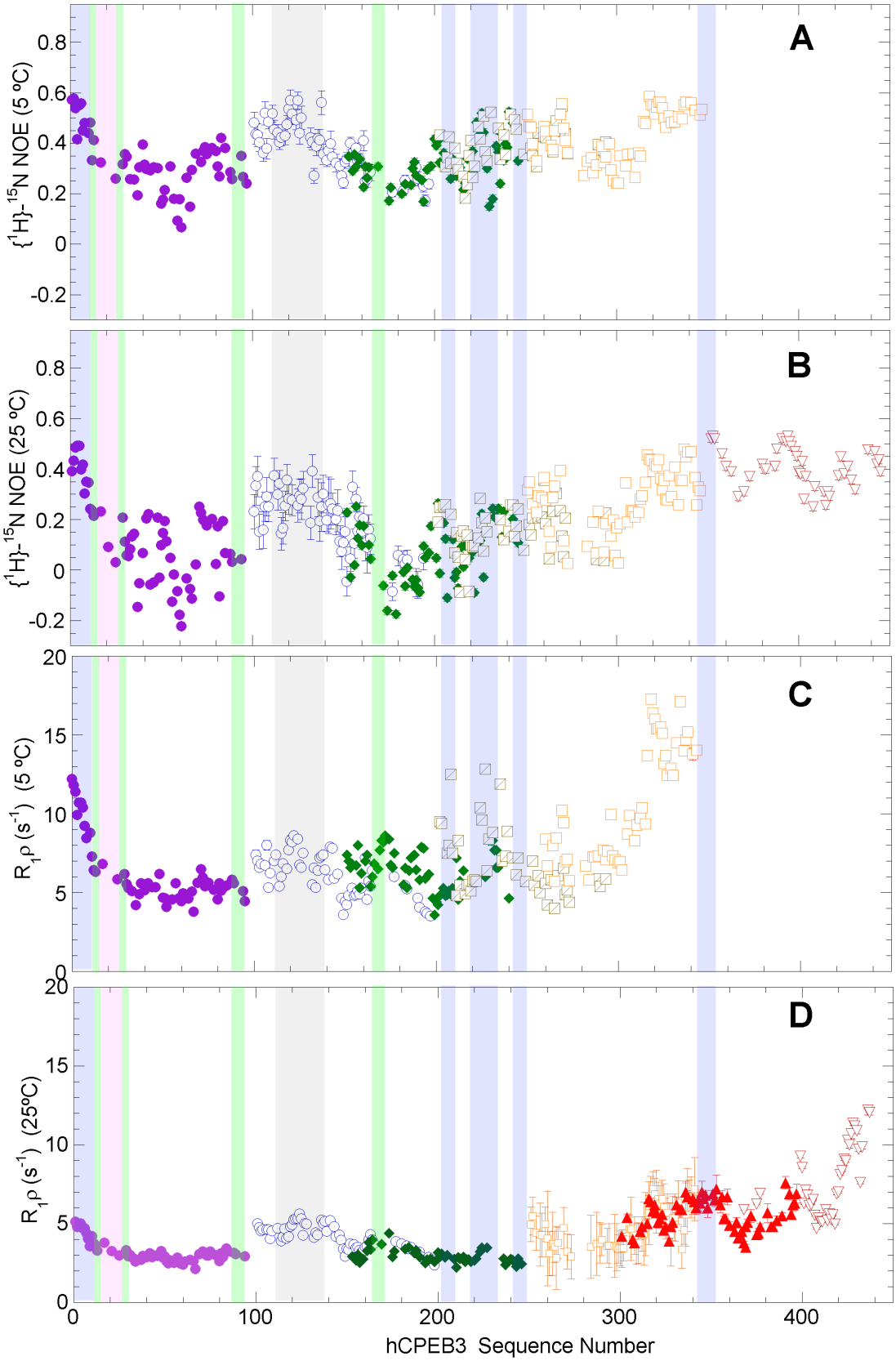




A. and **B**. Conformational chemical shifts (Δδ) for ^{13}Cα (**blue** bars), ^{13}Cβ (**red**) and ^{13}CO (**green**) nuclei for residues in two Pro-rich stretches, H84-Q94 and Q165-175, each of which contains four consecutitive Proline residues which will be locked in the polyproline II helical conformation. 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.

D. Fluorescence emission spectra of 8.3 μM human profilin 1 in the absence (blue spectrum) or presence (black) of 4.5 mM acGPPPPAPAPQPam peptide at 20 ºC in 10 mM K_2HPO_4, 75 mM KCl, (pH 8.0) at 20 ºC. The enhanced, blue-shifted emission is indicative of binding [34].

Sup. Fig. 12. Residue Level Dynamics of hCPEB3's Instrinsically Disordered Region

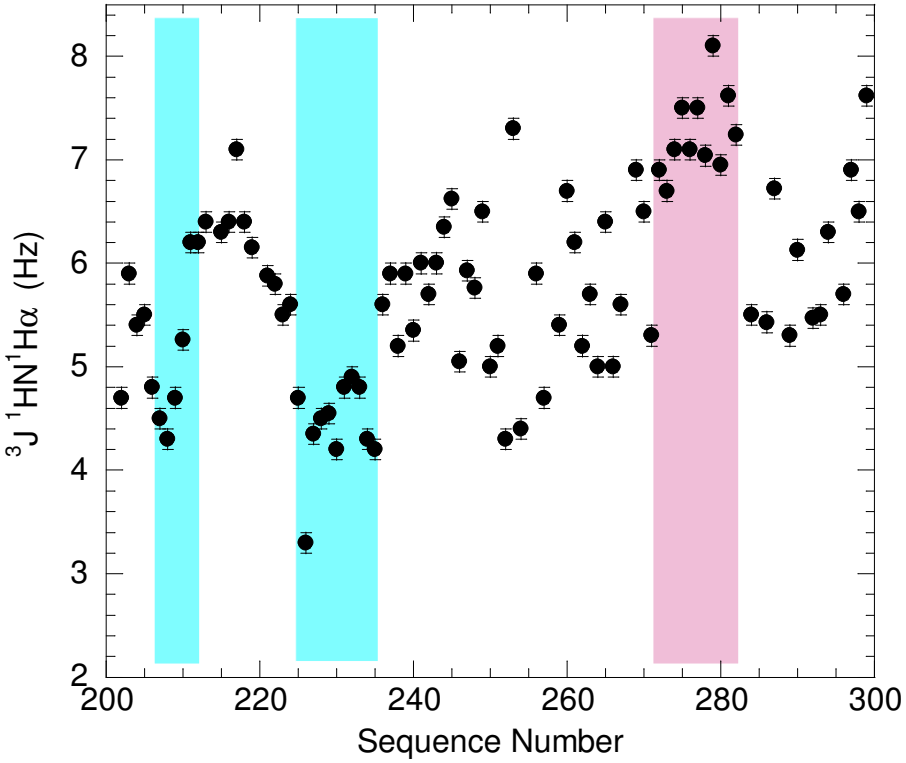


Residue-level dynamics on ns/ps time scales at 5ºC (**A**) and 25ºC (**B**) and on µs/ms at 5ºC (**C**) and 25 ºC (**D**). In panels **A** and **B**, a value of 0.86 is expected for completely rigid H-N groups; negative values

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.

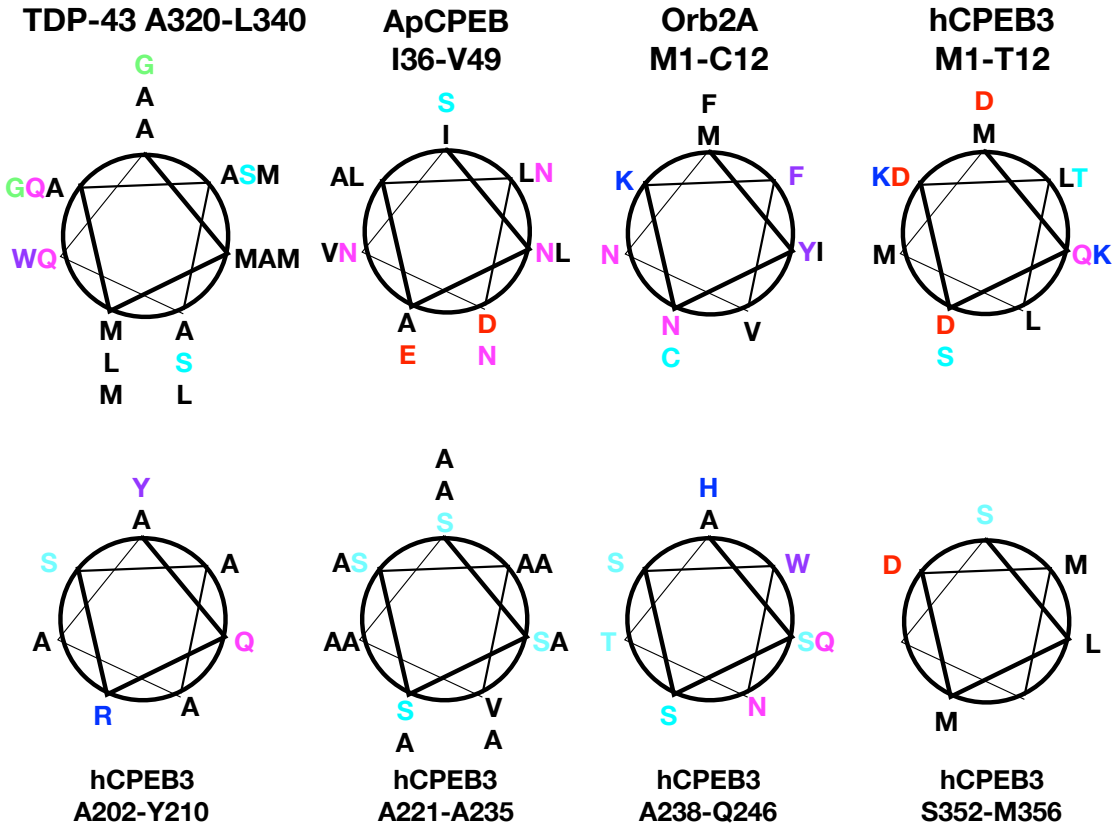
Elements of partial structure/interest are shaded **blue** for α-helices and **gray** for hydrophobic segments; these elements generally show decreased mobility. By contrast, the polyQ tract (shaded **magenta**) is more flexible. Whereas the dynamics of the PPII tracts (**green**) can now be assessed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

Sup. Fig. 13. ^{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*). By contrast, the last, modestly populated α-helix identified by chemical shift deviations (A238-Q246), does not show significantly lower $^{3}J_{1HN1H\alpha}$ values. A zone with $^{3}J_{1HN1H\alpha} > 7$ Hz (shaded **rose**) indicates extended conformations. This zone spans the $(VG)_5$ motif.

Sup. Fig. 14. Helices from Pathological and Functional Amyloids Are Stabilized by Distinct Interactions.



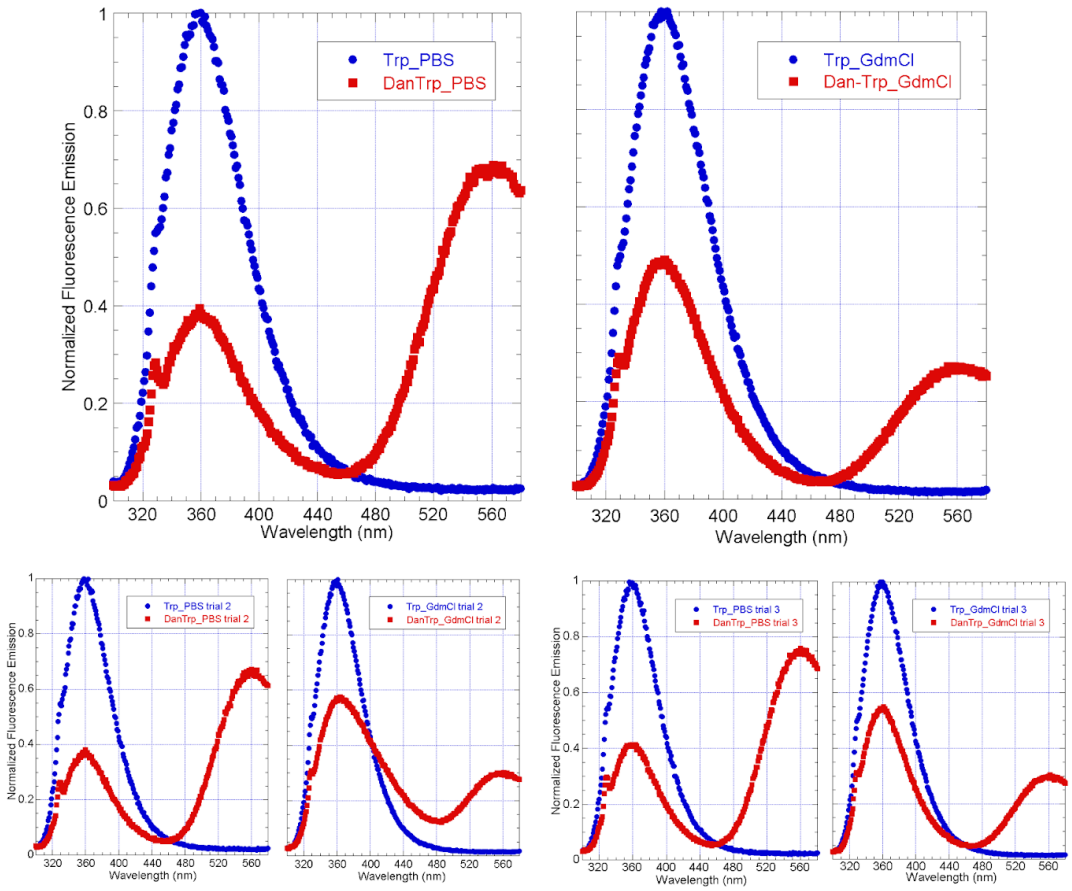
Helical wheel diagrams of TDP-43, ApCPEB, Orb2A and hCPEB3, nonpolar residues are colored **black**, aromatics=**purple**, Q/N=**magenta**, E/D=**red**, H/K/R=**blue**, G=**green**, C,S,T=cyan. The five hCPEB3 helices were identified here by NMR data. The TDP-43 helix was identified by NMR by Lim *et al.,* 2016. [51] The ApCPEB and Orb2A helices are putative as they have not yet been confirmed experimentally.

Sup. Fig. 15. Insight into Interhelix Interactions from FRET

Fluorescence spectra of two polypeptides whose concentration was 10.0 μM: MQDDLLMDKSKTGGGGASSSWNTHQ called "Trp" and Dansyl-MQDDLLMDKSKTGGGGASSSWNTHQ called "Dan-Trp", with residues corresponding to the first and fourth partly populated α-helices of hCPEB3 (shaded pink and cyan, respectively) were recorded.

It can be seen in the graph on the *left* (in buffer) that the emission of Trp is high in "Trp" but low in the "Dan-Trp" peptide. This is due to energy transfer of Trp to Dan and thanks to this the Dansyl shines and there is an emission centered at 560 nm which is not seen for the "Trp" peptide.

Applying the theory of Förster and making usually assumptions about the rapid movement and averaged orientation of the donor and acceptor groups, and using the known "Förster distance" (R_0) of 23.6 angstroms Lakowicz *et al.* (1990), the mean distance between the Dansyl and Trp can be calculated to be 22 angstroms, which is significantly lower than the mean value (41 angstroms) expected for a statistical coil of 21 residues (the number of residues separating the Dansyl and Trp groups) as calculated on the basis of the data complied by Fitzkee and Rose (2004) PNAS·USA **101**(34):12497-12502 as well as the results reported by Tanford *et al.* (1966) J. Biol. Chem. **241**(8) 1921-1923.



On the other hand, in the presence of 7.4 M GdmCl, in the graphs shown on the *right*, the Trp emission intensity decrease due to Dansyl is smaller, and also the emission intensity of Dansyl is less intense. Two repetitions of the experiment are shown in the lower panels. Based on these results, one can conclude that the Trp and Dansyl groups are farther away in the presence of the denaturant, due to the unfolding of a more compact structure which is present in the absence of GdmCl. Thus, these new results support the possibility of interhelical contacts in hCPEB3.