SCIENTIFIC REP🔴RTS

OPEN

Characterization of Soft Amyloid Cores in Human Prion-Like Proteins

Cristina Batlle[1], Natalia Sanchez de Groot [ID][2,3], Valentin Iglesias[1], Susanna Navarro[1] & Salvador Ventura [ID][1]

Prion-like behaviour is attracting much attention due to the growing evidences that amyloid-like self-assembly may reach beyond neurodegeneration and be a conserved functional mechanism. The best characterized functional prions correspond to a subset of yeast proteins involved in translation or transcription. Their conformational promiscuity is encoded in Prion Forming Domains (PFDs), usually long and intrinsically disordered protein segments of low complexity. The compositional bias of these regions seems to be important for the transition between soluble and amyloid-like states. We have proposed that the presence of cryptic soft amyloid cores embedded in yeast PFDs can also be important for their assembly and demonstrated their existence and self-propagating abilities. Here, we used an orthogonal approach in the search of human domains that share yeast PFDs compositional bias and exhibit a predicted nucleating core, identifying 535 prion-like candidates. We selected seven proteins involved in transcriptional or translational regulation and associated to disease to characterize the properties of their amyloid cores. All of them self-assemble spontaneously into amyloid-like structures able to propagate their polymeric state. This provides support for the presence of short sequences able to trigger conformational conversion in prion-like human proteins, potentially regulating their functionality.

A broad range of human pathologies, ranging from neurodegenerative conditions such as Alzheimer's and Parkinson's diseases, to non-neuronal disorders such as type II diabetes and cataracts, are associated with protein misfolding and aggregation into amyloid-like structures[1]. The self-assembly of proteins into β-sheet-enriched amyloid conformations appears to obey, in many cases, the so-called amyloid-short-stretch hypothesis, according to which, aggregation is first nucleated by the intermolecular contacts formed by a reduced number of specific short regions in the protein[2]. In pathogenic proteins, these stretches are generally around 5–10 residues in length, with a high aggregation propensity and a predominant hydrophobic character[3].

Prions are proteins able to adopt multiple structural conformations, from which at least one has self-propagating properties, usually an amyloid state[4]. Prions have been traditionally associated with the onset of mammalian neurophatologies[5]. Nevertheless, there are evidences that prion-like mechanisms are not always deleterious and instead they can be used for beneficial purposes[6]. The best characterized set of functional prions has been found in yeast, where they can behave as epigenetic elements, facilitating adaptation to fluctuating environments[7,8]. The conformational promiscuity of yeast prions is encoded in Prion Forming Domains (PFDs)[9]. PFDs are both sufficient and necessary for prion conversion and usually correspond to long and intrinsically disordered segments of low complexity[9].

The information on the common features shared by yeast PFDs has fueled the development of algorithms aimed to identify similar prion-like domains (PrLDs) and the proteins that contain them at the proteome level[9–17]. Despite PFDs bear the capacity to shift to an amyloid state, classical amyloid prediction algorithms fail to identify them[12]. This observation led to suggest that PFDs assembly is governed by the low complexity and compositional bias common to these domains[9,12]. Examples of algorithms exploiting these features are DIANA[10], LPS[11], PrionScan[13], PLAAC[16] and PAPA[12]. Remarkably, these computational approaches predict the existence of PrLD-containing proteins in a wide variety of organisms, from prokaryotes to higher eukaryotes[18–23]. In humans, this sub-proteome is enriched in nucleic acid-binding proteins[18,19]. A fraction of these proteins seem to be involved in the formation of membraneless intracellular compartments, like RNA granules, through PrLDs

[1]Institut de Biotecnologia i de Biomedicina i Departament de Bioquímica i Biologia Molecular, Universitat Autónoma de Barcelona, Bellaterra, 08193, Spain. [2]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003, Barcelona, Spain. [3]Universitat Pompeu Fabra (UPF), Barcelona, Spain. Correspondence and requests for materials should be addressed to S.V. (email: salvador.ventura@uab.es)

mediated liquid-liquid phase separation[24,25]. Mutations in these domains have been shown to promote an aberrant transition to an aggregated amyloid-like state, the formation of which might led to the onset of neurodegenerative diseases[26,27].

We have recently proposed that, in addition to composition, the presence of soft amyloid cores inside PFDs and PrLDs could be important for their assembly[28,29]. We rationalized that these assembly-nucleating regions should be longer than classical amyloid stretches, in such a way that the amyloid potential would be more diffusely distributed; each residue having an average lower potency, but with more residues contributing to the assembling force. This will make their aggregation sensitive to protein concentration and seeding. Thus, in our view, the aggregation of prion-like domains shares mechanistic features with those of classical amyloidogenic proteins. This notion was implemented in pWALTZ, an algorithm that predicts the 21-residues long sequence stretch with the highest average amyloid potential in PrLDs[14,15]. This concept was further validated experimentally by demonstrating the existence of such soft amyloid stretches in the prion domains of four of the best characterized yeast prions[30], as well as in the predicted PrLD of the Rho termination factor of *Clostridium botulinum*[31], which later led to the discovery of the first bacterial prion-like protein[32].

Here, we study the presence of soft amyloid cores in novel putative human prion-like proteins. For this purpose, we performed a stringent computational analysis of the human proteome in the search of domains that, while fulfilling the compositional bias characteristic of PrLDs, would also exhibit a sequence stretch that can potentially nucleate their self-assembly. From this set, we selected seven nucleic acid-binding proteins associated to disease (DDX5, EYA1, ILF3, MED15, NCOA2, PHC1 and TIA1) to structurally characterize the nature of their putative nucleating cores. The results herein indicate that the PrLDs of all these proteins include 21-residues long stretches able to self-assemble spontaneously into non-toxic, β-sheet enriched, Thioflavin-T positive amyloid-like structures displaying self-seeding activity. Therefore, the present work provides compelling experimental evidence for the existence of specific sequences with the potential to trigger the conformational conversion of PrLDs in human proteins.

Results

Identification of PrLD soft amyloid cores in human prion-like proteins.
The analysis of the 70940 protein sequences in the human proteome was initially performed with PAPA[12] and further refined with pWALTZ[14]. Both PAPA and pWALTZ algorithms were trained on top of yeast prions; however, they are based on radically different concepts, a suitable composition of the PrLD and the presence of an embedded soft amyloid core, respectively. Sequences identified by these two orthogonal approaches are expected to recapitulate the conformational properties of yeast prions. A total of 663 human proteins were identified by PAPA and later shortlisted using pWALTZ to render a total of 535 polypeptides, encoded by 336 different genes (Table S1).

We wanted to test whether the soft amyloid cores predicted inside the PrLDs of these putative human prion-like proteins could spontaneously self-assemble into amyloid-like conformations and propagate their aggregated state, as we observed before for yeast prions[30] and the prion-like *C. botulinum* Rho factor[31]. We focused on nucleic acid-binding proteins, both because this molecular function is enriched in our dataset and because most of the experimentally validated yeast prions act in translational or transcriptional regulation. We selected six proteins associated to disease whose prionogenic properties have not been reported before: DDX5, EYA1, ILF3, MED15, NCOA2 and PHC1. We also included in the analysis TIA1, an RNA-binding protein identified by the orthogonal approach for which a prion-like behavior has been already suggested[33] (Tables 1 and 2).

DDX5 (p68) (Fig. 1A) is a member of the DEAD box family of RNA helicases involved in transcriptional regulation[34] and it is overexpressed in various types of cancers such as those of prostate, breast and colon; promoting cell proliferation and metastasis[35,36]. The C-terminal region containing the predicted PrLD seems to play a role in the interaction of the protein with other components of the transcriptional machinery[37].

EYA1 (Fig. 2A) is a transcriptional coactivator and a protein phosphatase with regulatory roles in nephrogenesis[38]. The protein is overexpressed in breast cancer[39] and mutations in EYA1 gene are associated with branchio-oto-renal syndrome[40]. The N-terminal region containing the PrLD has been reported to function as a transactivation domain[41].

ILF3 (Fig. 3A) works in RNA metabolism, from transcription to degradation and it appears to be essential for cellular development[42]. Interestingly, ILF3 participates in ribonucleoprotein (RNP) granules assembly[43] and it interacts with FUS, a well-characterized prion-like protein[44].

MED15 (Fig. 4A) is one part of the Mediator complex involved in the transcription of RNA-polymerase II dependent genes[45]. The identified PrLD corresponds to a Q-rich region similar to those accounting for conformational conversion in yeast prions. Indeed, its yeast homolog has been already classified as a prion[46] with the ability to form amyloid-like structures *in vivo* under stress conditions[47].

NCOA2 (Fig. 5A) is a transcriptional coactivator for steroid receptors and nuclear receptors acting in the upregulation of DNA expression[48]. It has a tissue-specific role in tumorigenesis, acting as an oncogene in prostate cancer[49] and as a tumor suppressor in liver cancer[50]. Moreover, NCOA2 is a key player in glucose homeostasis, being involved in Mediator recruitment for glucokinase expression[51].

PHC1 (Fig. 6A) is one component of the Polycomb complex responsible for cellular differentiation during development. This complex is constituted by gene silencing proteins that repress important developmental regulator genes, including homeotic (HOX) genes[52]. PHC1 is associated with primary microcephaly[53].

TIA1 (Fig. 7A) is an RNA-binding protein and a component of stress granules required to regulate alternative splicing and mRNA translation and turnover. The predicted PrLD lies in a Q-rich region. TIA1 has already been suggested to be a functional prion-like protein whose dysregulation is involved in Amyotrophic Lateral Sclerosis (ALS)[33]. However, the presence of an amyloid core inside its PrLD has not been addressed before.

To further confirm the presence of PrLDs in these proteins and to define more precisely their boundaries we used PLAAC, yet another composition-based predictor, in which, in contrast to PAPA, the length of the predicted

PROTEIN	FUNCTION	DISEASE
DDX5	RNA helicase protein involved in transcriptional regulation	Prostate, breast and colon cancer, leukemia, hepatitis and others.
EYA1	Transcriptional coactivator and protein phosphatase	Branchio-oto-renal syndrome, breast cancer, cataract and others.
ILF3	Facilitates gene expression regulation from transcription to degradation	Cancer, myocardial infarction, chronic kidney disease and others.
MED15	Component of the mediator complex involved in the transcription of RNA-pol II dependent genes	Epicondylitis, prostate cancer, prostatitis and others.
NCOA2	Nuclear receptors and steroid receptors coactivator	Prostate cancer, mesenchymal chondrosarcoma and others.
PHC1	Component of a Polycomb group involved in the maintenance of the transcriptionally repressive state of HOX genes	Microcephaly.
TIA1	Regulates alternative splicing, associated with apoptosis and 3'UTR mRNA binding protein	Amyotrophic Lateral Sclerosis, Hodgkin's Lymphoma, myopathy, Spinal Muscular Atrophy and others.

Table 1. Function and implication in disease of the selected human PrLD-containing proteins. Uniprot[77] and Malacards Human Disease database[78] were used to determine the function and disease-association for each human PrLD-containing candidate, respectively. All of them are involved in transcriptional o translational regulatory functions and have been associated to common diseases such as cancer or degenerative disorders.

PROTEIN	UNIPROT ID	PrLD AMYLOID CORE	pWALTZ SCORE	Q/N/S/G (%)
DDX5	P17844	530-TQNGVYSAANYTNGSFGSNFV-550	71.62	52.38
EYA1	Q99502	184-MQGSSFTTSSGIYTGNNSLTN-204	74.69	57.14
ILF3	Q12906	672-YGSYGYGGNSATAGYSQFYSN-692	71.01	57.14
MED15	Q96RN5	184-QQQQQFQAQQSAMQQQFQAVV-204	70.58	61.90
NCOA2	Q15596	1374-HFGQQANTSMYSNNMNINVSM-1394	70.39	52.38
PHC1	P78364	382-QQQQIHLQQKQVVIQQQIAIH-402	74.58	47.62
TIA1	P31483	331-AYGMYGQAWNQQGFNQTQSSA-351	67.81	57.14

Table 2. Selected human PrLD amyloid cores. For each PrLD-containing protein it is shown its Uniprot ID, its 21 residues-long amyloid core, the pWALTZ score for this protein region and its Q/N/S/G content.

PrLD also depends on the protein composition. PLAAC detected PrLDs, overlapping with the regions previously identified by PAPA, in all the above described proteins (Figs 1A to 7A).

The predicted soft amyloid cores for these proteins are shown in Table 2. Remarkably, well-validated aggregation predictors like Aggrescan[54], Tango[55] and Zyggregator[56] failed to classify these stretches as aggregation-prone, the exception being the PHC1 core, which was identified by Tango (Table S2). The underlying reason explaining why these amyloid predictors fail to score properly the putative cores is likely their much lower hydrophaticity, when compared with the amyloid stretches present in pathogenic proteins like Aβ42 or α-synuclein (Table S3). Indeed, the predicted cores are enriched in polar residues like Gln, Asn and Ser as well as in Gly (Table 2), all amino acids considered conferring low aggregation propensities to protein sequences.

An analysis of the structural context in which these soft amyloid cores are embedded in their respective sequences using the prediction algorithms FoldIndex[57], IUPRED[58], PondR-FIT[59] and RONN[60] indicates that they are preferentially located in disordered protein segments (Table S4).

Predicted human PrLD soft amyloid cores assemble into β-sheet rich structures.

We synthetized 21-residues long peptides correspondent to the detected amyloid cores and analyzed their secondary structure content by far-UV circular dichroism (CD) immediately after their dilution at $100 \mu M$ in sodium phosphate buffer at pH 7.4 (Figs 1B to 7B). Five out of the seven peptides (DDX5, EYA1, ILF3, MED15 and TIA1) exhibited spectra consistent with a mostly disordered conformation. In ILF3, the high content in Tyr residues (23%) renders a characteristic aromatic signal at 230 nm. The other two peptides (NCOA2 and PHC1) already exhibited a β-sheet CD spectrum immediately after dilution, thus suggesting that they experience a very fast assembly in aqueous buffer.

Next, we incubated the peptides at $100 \mu M$ for 2 days at 37 °C and monitored their ability to form macromolecular structures using synchronous light scattering, bis-ANS binding, infrared spectroscopy (ATR-FTIR) and far-UV CD.

The formation of high-order assemblies after incubation was confirmed for all peptides by measuring the light scattering of the correspondent solutions. All of them exhibited significant scattering signal after 2 days (Figure S1). Despite the amyloid cores in this study are less hydrophobic than those of pathogenic amyloids (Table S3), they still exhibit non-polar residues that might contribute to the initial assembly. We explored the presence of exposed hydrophobic clusters in the detected aggregated material by measuring their binding to bis-ANS (Figure S2A), a dye that increases its fluorescence emission upon interaction with these regions[61]. The bis-ANS fluorescence emission maximum increases and blueshifts from 530 nm in the absence of peptides to 510 nm in their presence, indicating the existence of hydrophobic-patches in the surface of all these assemblies. The binding of these assemblies to bis-ANS was, however, much lower than those of the amyloid fibrils formed by the Parkinson's associated α-synuclein protein at the same concentration (Figure S2B), consistent with the lower hydrophaticity of PrLDs amyloid stretches.
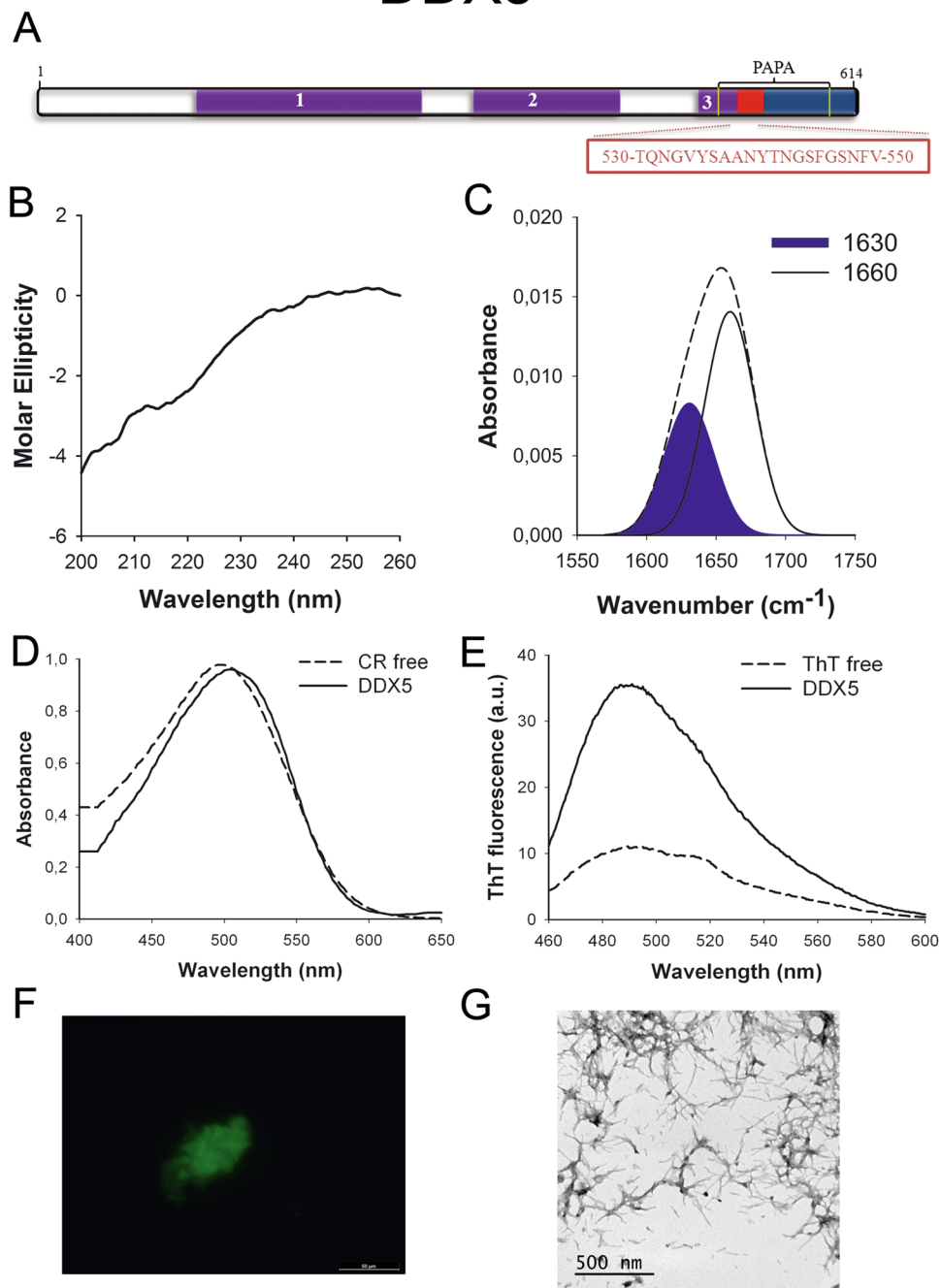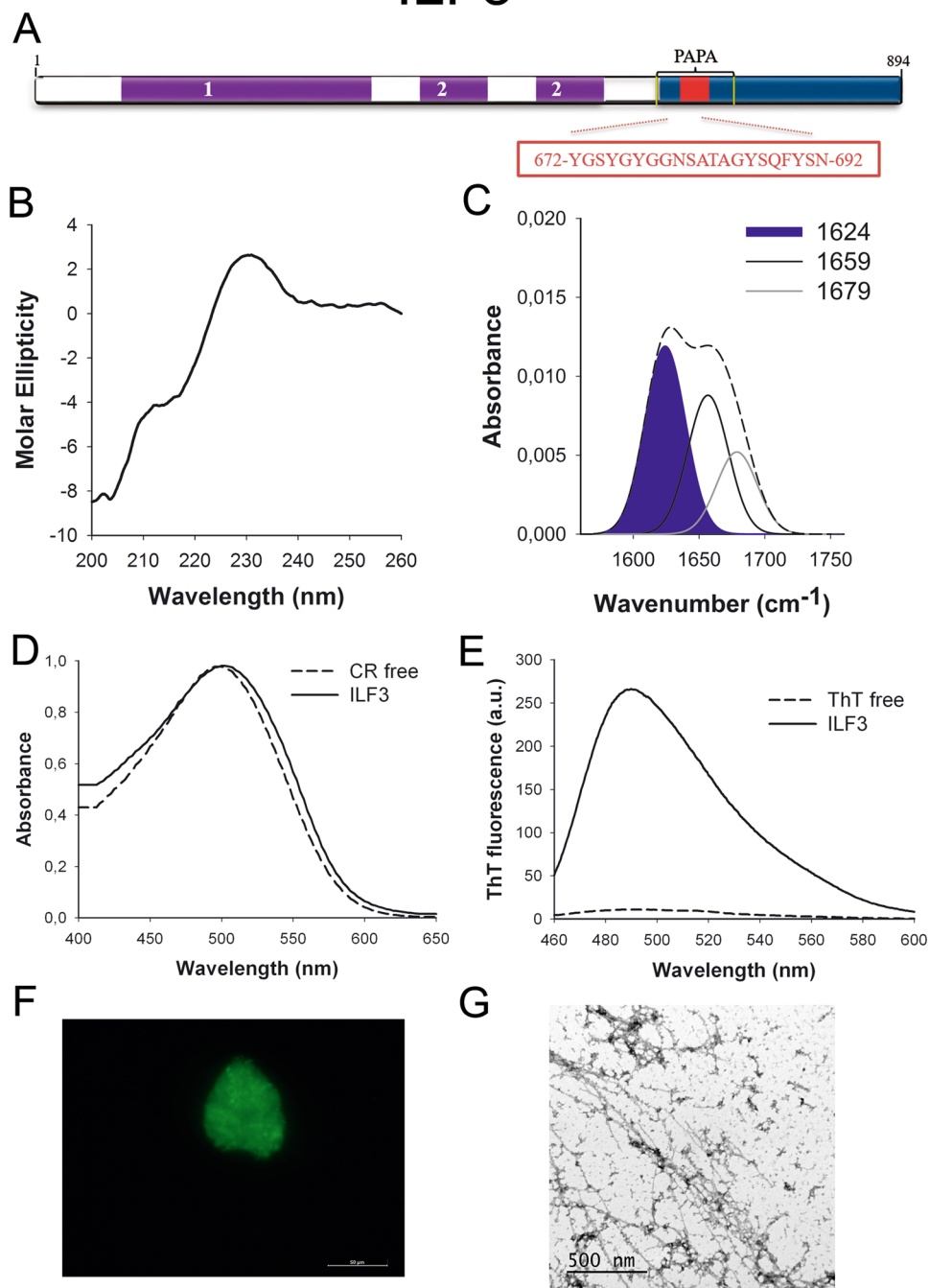
Figure 1. DDX5 PrLD amyloid core. (**A**) DDX5 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = DEAD domain. 2 = Helicase C domain. 3 = p68-like RNA helicase domain. The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of 100 μM DDX5 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) DDX5 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of DDX5 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of DDX5 peptide. (**F**) DDX5 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) DDX5 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of DDX5 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.
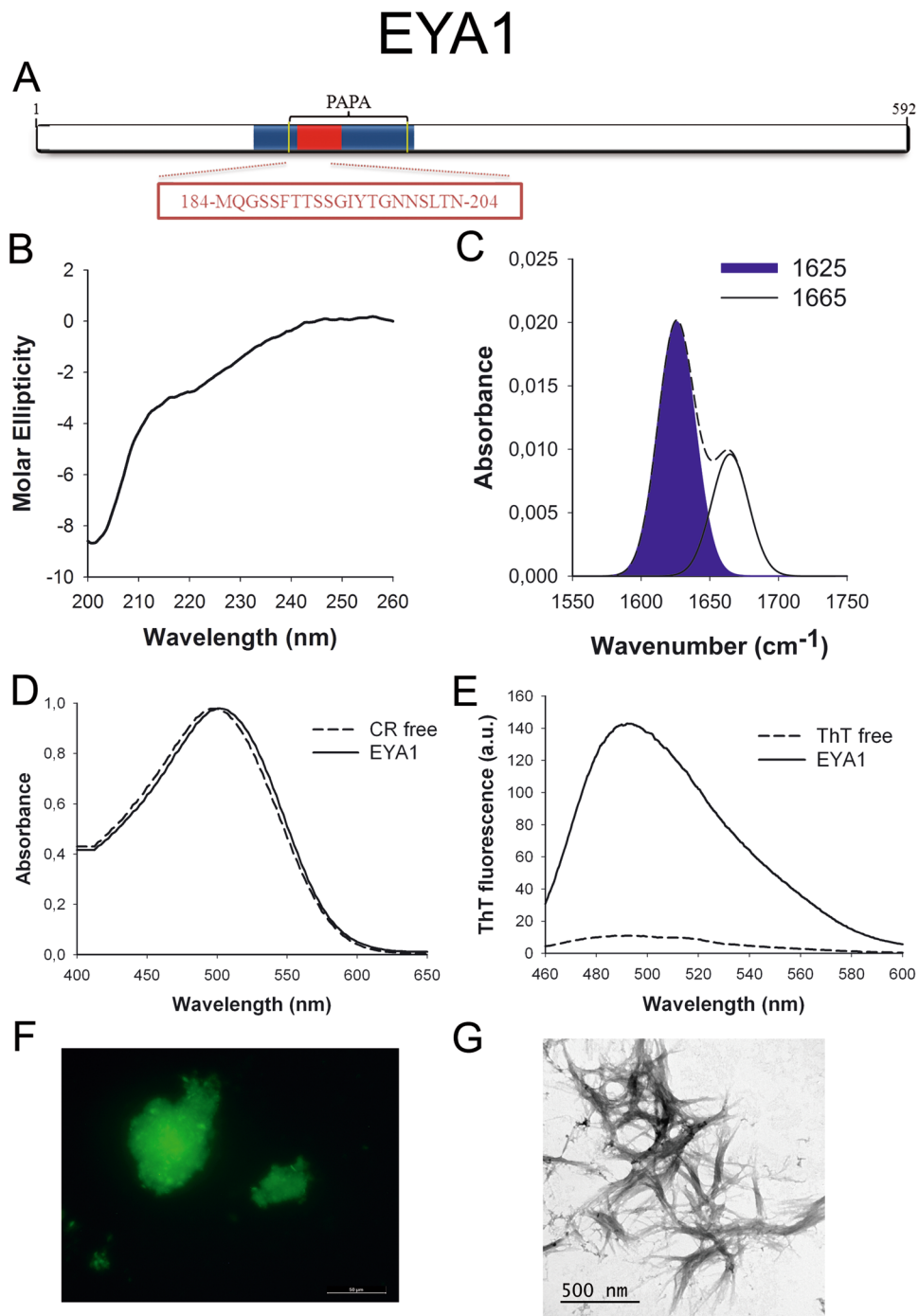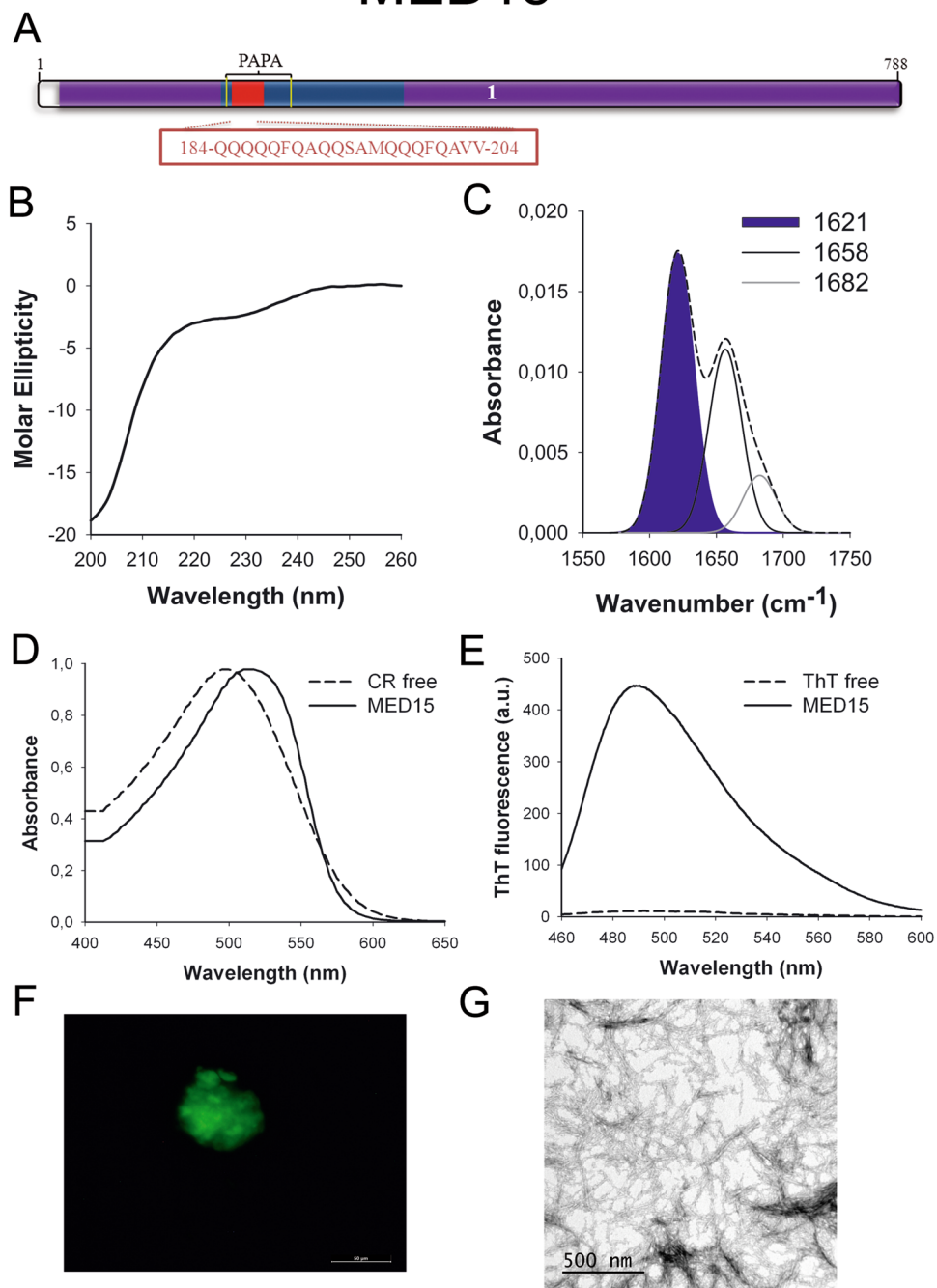
Figure 2. EYA1 PrLD amyloid core. (**A**) EYA1 diagram showing the location of the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of 100 μM EYA1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) EYA1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of EYA1 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of EYA1 peptide. (**F**) EYA1 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) EYA1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of EYA1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.
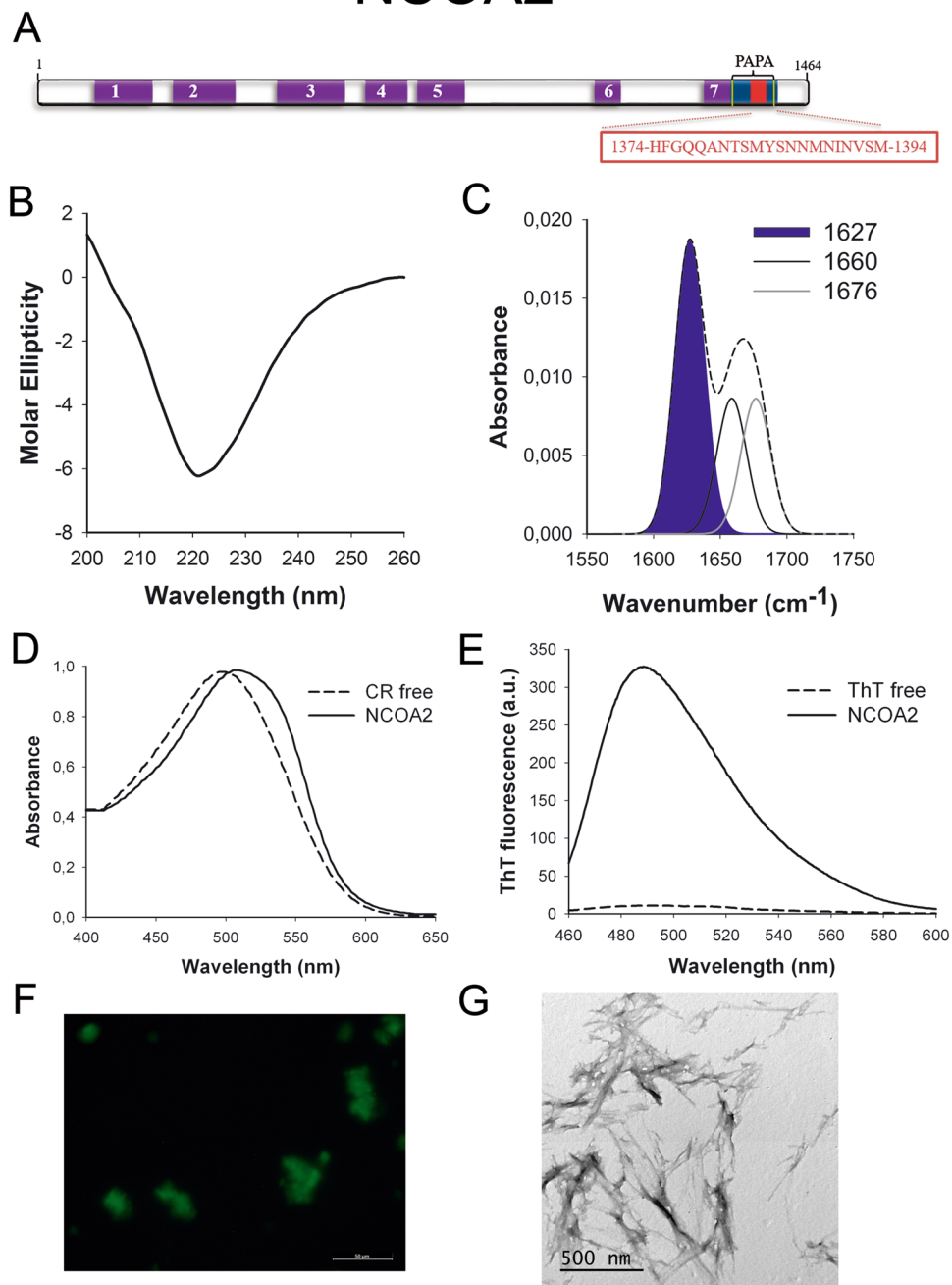
ILF3

Figure 3. ILF3 PrLD amyloid core. (**A**) ILF3 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = DZF domain. 2 = dsRNA binding motif. The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of 100 μM ILF3 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) ILF3 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of ILF3 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of ILF3 peptide. (**F**) ILF3 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) ILF3 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of ILF3 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.
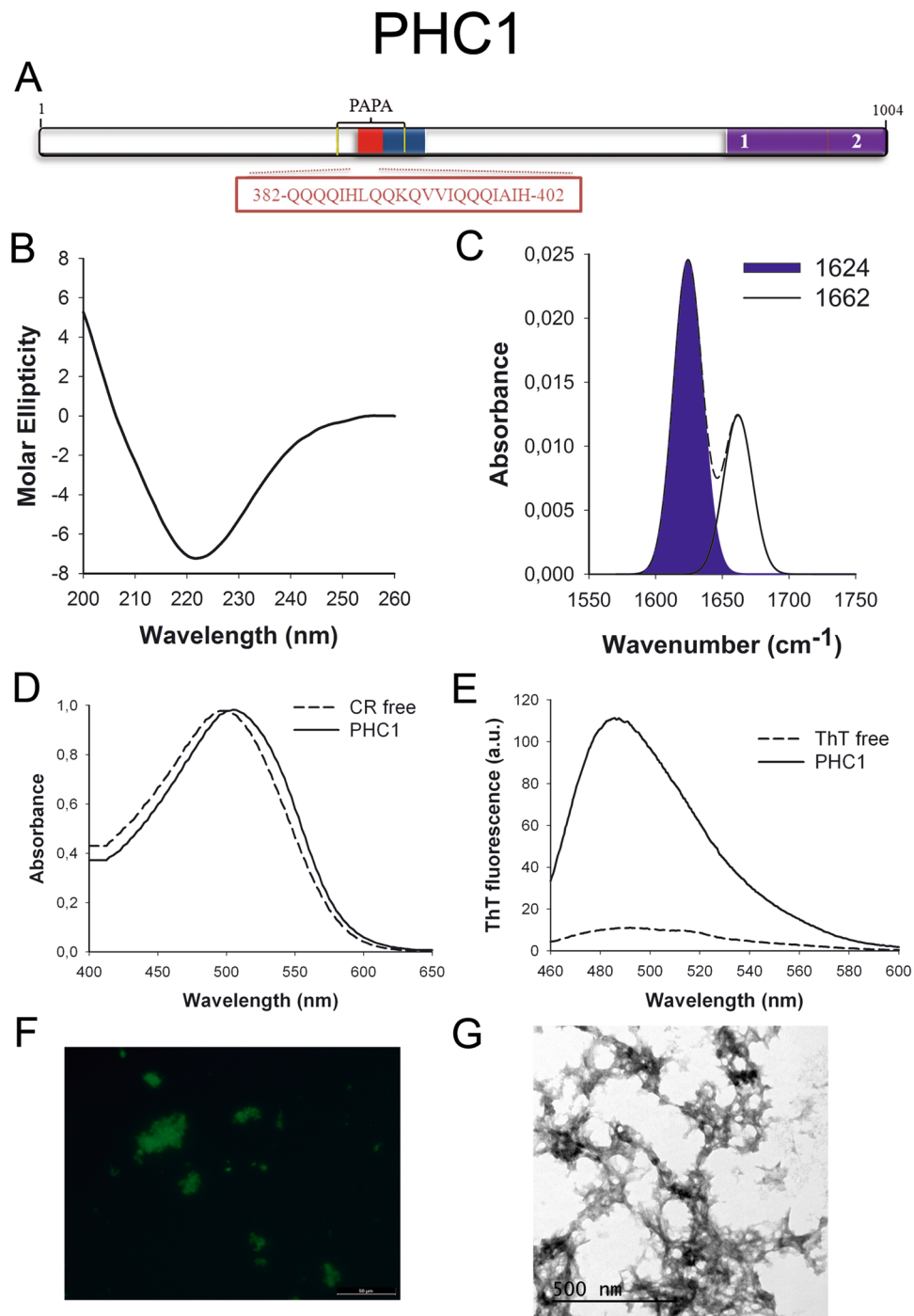
Figure 4. MED15 PrLD amyloid core. (**A**) MED15 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = MED15 domain. The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of 100 μM MED15 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) MED15 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of MED15 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of MED15 peptide. (**F**) MED15 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) MED15 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of MED15 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.
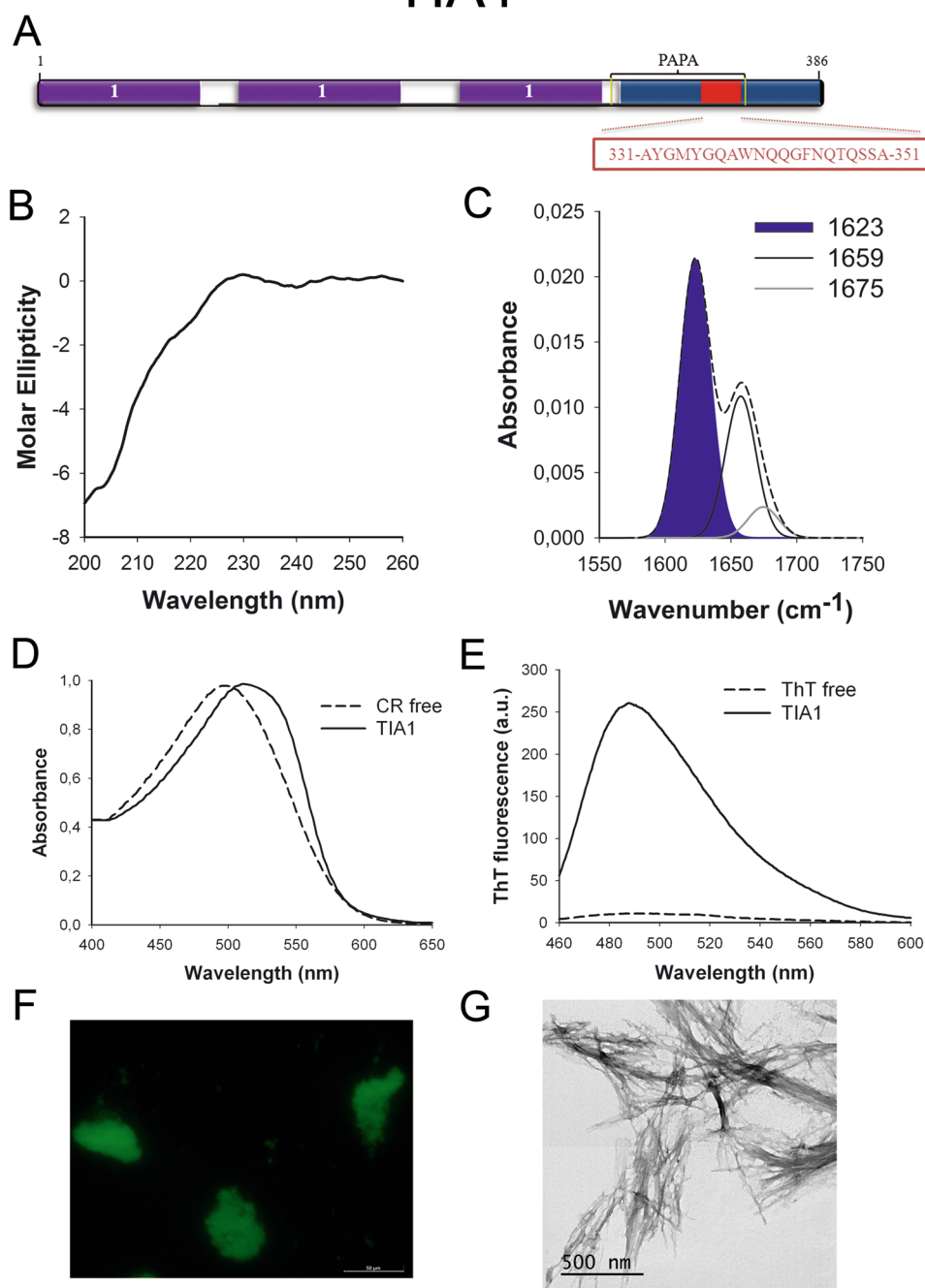
NCOA2

Figure 5. NCOA2 PrLD amyloid core. (**A**) NCOA2 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = PAS domain. 2 = PAS11 domain. 3 = NCOA_u2 domain. 4 = SRC1 domain. 5 = Duf4927 domain. 6 = Nuc rec co-act domain. 7 = DUF1518. The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of $100\,\mu M$ NCOA2 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) NCOA2 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of NCOA2 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of NCOA2 peptide. (**F**) NCOA2 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) NCOA2 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of NCOA2 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.

Figure 6. PHC1 PrLD amyloid core. (**A**) PHC1 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = PHC2 SAM assoc domain. 2 = SAM1 domain. The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of 100 μM PHC1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) PHC1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of PHC1 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of PHC1 peptide. (**F**) PHC1 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) PHC1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of PHC1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.

Figure 7. TIA1 PrLD amyloid core. (**A**) TIA1 diagram showing the location of the identified Pfam domains (purple), the amyloid core (red) and the PrLD as predicted by PLAAC (blue) and PAPA (yellow lines). 1 = RNA recognition motif (RRM). The sequence of the amyloid core is shown in the box. (**B**) CD spectrum in the far-UV region of $100\,\mu$M TIA1 peptide in 5 mM potassium phosphate buffer pH 7.4 before incubation. (**C**) TIA1 peptide FT-IR absorbance spectrum in the amide I region. The dashed line corresponds to the original spectrum, the blue area indicates the contribution of the inter-molecular β-sheet signal to the total area upon Gaussian deconvolution. (**D**) CR absorbance spectrum in the absence (dashed line) and in the presence (solid line) of TIA1 peptide. (**E**) Fluorescence emission spectrum of Th-T in the absence (dashed line) and in the presence (solid line) of TIA1 peptide. (**F**) TIA1 peptide stained with Th-S and observed at 40X magnification using fluorescence microscopy. (**G**) TIA1 peptide representative transmission electron micrograph. The data in panels C to G were collected upon incubation of TIA1 peptide for 2 days in 5 mM potassium phosphate buffer pH 7.4 at 37 °C.

Next, we recorded the amide I region of the FTIR spectrum (1700–1600 cm^{-1}) for these aggregates (Figs 1C to 7C). This region corresponds to the absorption of the carbonyl peptide bond group of the protein main chain and it is conformation sensitive. Deconvolution of the spectra allowed us to assign the individual secondary structure elements of incubated peptides and their relative contribution to the main absorbance (Table S5). In all the cases we could identify a strong band at 1620–1630 cm^{-1}, usually assigned to the presence of inter-molecular β-sheets. This signal is the largest contributor to the absorbance spectrum in all peptides, except for DDX5, where it contributes 37% of the area. Interestingly, no anti-parallel β–sheet band was detected (~1690 cm^{-1}) in any of the samples; thus suggesting that the detected β–strands in self-assembled peptides would adopt preferentially a parallel disposition. The other detected structural elements are associated with disordered structure and turns (Table S5). We also monitored the secondary structure of the incubated peptides using far-UV CD (Figure S3). In all cases we could detect a band at 215–220 nm consistent with the population of a β-sheet enriched conformation, despite in some cases the ellipticity was low, indicating that a significant proportion of the peptide was aggregated, therefore out of the solution and not detectable.

Overall, our data are consistent with the spontaneous assembly of the predicted human PrLD amyloid cores into supramolecular β-sheet enriched structures.

Predicted human PrLD amyloid cores form non-toxic amyloid-like fibrillar structures. We used the amyloid-specific dyes Congo red (CR), Thioflavin T (Th-T) and Thioflavin-S (Th-S) to confirm that the detected β–sheet enriched aggregates were organized into amyloid-like suprastructures.

The absorbance spectra of CR red shifts in the presence of amyloid aggregates[62]. Incubation of CR in the presence of aggregated peptides resulted in a red shift of its spectrum in all the cases (Figs 1D to 7D); despite for DDX5, EYA1 and ILF3 the spectral shift was small. To confirm the ability of these three peptides to bind CR they were incubated under the same conditions at 500 μM final concentration. The peptides in these three solutions promoted a clear shift of the absorption maximum of the die (Figure S4).

Th-T fluorescence emission is enhanced in the presence of amyloid fibrils[63]. All the peptides promoted an increase in the intensity of Th-T fluorescence spectral maximum at 488 nm (Figs 1E to 7E). Furthermore, binding of Th-S to the aggregates could be visualized by fluorescence microscopy for all incubated peptides. Areas rich in fibrous material were stained with Th-S to yield green-yellow fluorescence against a dark background in all cases (Figs 1F to 7F).

The dye binding results indicate that incubated peptide solutions might contain detectable amounts of amyloid-like structures. To confirm this extent, the morphological features of the peptide assemblies in these samples were analyzed using transmission electron microscopy (TEM). Negative staining indicated that all peptides effectively assemble into supramolecular structures (Figs 1G to 7G). The aggregates formed by DDX5, EYA1, MED15, NCOA2, and TIA1 correspond to amyloid-like fibrillar arrangements, without any significant accumulation of amorphous material. The fibrils exhibit a diameter that varies from 5 to 10 nm and a length that ranges from 2 to 10 μm. In the ILF3 sample long amyloid-like fibrils coexist with small aggregates that appear to attach to the fibrils, whereas, for PHC1, despite its amyloid-like tintorial properties, the material appears to be essentially protofibrillar. All the peptides were also able to form macromolecular aggregates when they were incubated at 10 μM (1/10 the concentration used in the previous assays) (Figure S5).

The above data indicate that the predicted amyloid cores exhibit a strong propensity to form amyloid-like assemblies. The amyloids formed by pathogenic protein fragments are usually highly cytotoxic. We tested if the aggregates formed by the PrLD amyloid cores display any toxicity when administered to neuroblastoma SH-SY5Y cells. All the aggregates were essentially innocuous when added to the cell cultures at up to 10 μM (Figures S6).

Predicted human PrLD amyloid cores form aggregates with self-seeding properties. Seeded protein aggregation is a well-established mechanism for *in vivo* amyloid fibril formation and underlies prion propagation[64]. The nucleation step of the amyloid assembly is shortened in the presence of preformed amyloid fibrils of the same protein, that can act as nuclei for the subsequent polymerization reaction[65]. Specific and short aggregation-prone regions have been shown to play a crucial role in this process[66,67]. To test whether preformed PrLDs core amyloid-like assemblies can seed the aggregation of the correspondent soluble peptides, we followed the aggregation kinetics of the peptides at 100 μM in the presence and absence of 2% (w/w) of preformed aggregates (Fig. 8). We could not monitor the aggregation kinetics of NCOA2 or PHC1 because they exhibited very high Th-T signal from the very beginning of the reaction, consistent with a very fast assembly into β-sheet structures, as suggested by the far-UV CD spectra they exhibit immediately upon dilution in aqueous solution (Figs 5B and 6B). The rest of peptides exhibited characteristic sigmoidal aggregation kinetics with lag-phases ranging from 20 to 120 min in the absence of seeds. The addition of preformed aggregates strongly accelerated the formation of Th-T positive assemblies in all cases (Fig. 8), supporting a nuclei-dependent aggregation mechanism and raising the possibility that such specific interactions could also occur in the context of the complete proteins in which these short regions are embedded.

Discussion

The prion phenomenon is best known by its association with spongiform encephalopathies in mammals, but in the last years a growing list of non-pathological prion-like proteins are being discovered. Functional prions were originally identified in yeast[68]. In these prion proteins, specific PFDs encode for their ability to switch between soluble and self-assembled states. Computational analysis are revealing the existence of polypeptides displaying similar domains in previously unexplored proteomes[17]. This suggests that this type of conformational conversion might be an evolutionary conserved mechanism exploited by different organisms, including humans, for beneficial purposes.

A characteristic feature of many pathogenic amyloids is the presence of short hydrophobic sequence stretches able to nucleate the assembly of these proteins into toxic aggregates both *in vitro* and *in vivo* under pathological conditions. These highly amyloidogenic regions seem to be absent in PFDs and PrLDs, likely because their
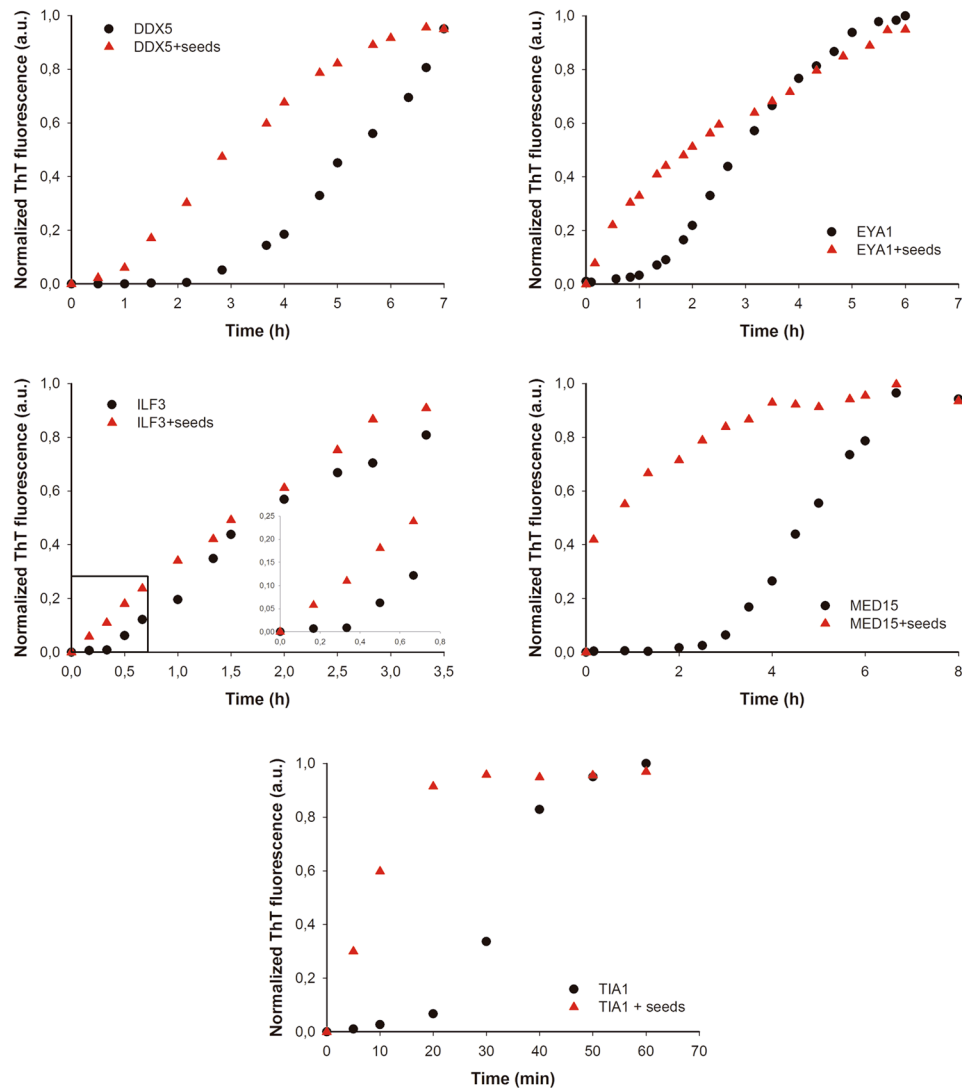
Figure 8. Human PrLD amyloid cores kinetics. Aggregation kinetics of DDX5, EYA1, ILF3, MED15 and TIA1 peptides at 100 μM in the absence (black circles) and presence (red triangles) of 2% (w/w) of preformed aggregates (seeds) were monitored by tracking the changes in Th-T fluorescence emission spectra. All the peptides exhibited accelerated kinetics in the presence of seeds.

potency will unbalance the equilibrium between the soluble and self-assembled states of the proteins towards an irreversibly aggregated and potentially toxic state, even in the absence of stress. However, we have recently proposed that in addition to a special compositional bias, PFDs and PrLDs contain indeed cryptic soft amyloid cores that can play an important role at the early stages of assembly by restricting inter-molecular interactions to specific regions of these long domains[28]. In these cores, the amyloid nucleating potency would be weaker and less concentrated that in pathogenic amyloids. This property would allow the protein to remain soluble in most physiological conditions, while being responsive to conditions that favor inter-molecular contacts or to the presence of preformed assemblies that target these specific segments. We have provided evidences for the presence of such regions in the PFDs of four of the best characterized yeast prions and for the ability of one of these soft amyloid cores to recruit the assembly of its correspondent full-length PFD *in vitro* and in the cell[30].

Using an approach analogous to the one described here, we previously screened the *C. Botulinum* proteome for the presence of regions displaying both compositional similitude to *bona fide* prions and containing a soft amyloid core, identifying a first bacterial PrLD in the Rho Terminator factor of this pathogen[31]. Later on, Yuan and Hochschild validated the prion-like nature of this protein, demonstrating that Rho can access alternative protein conformations in prokaryotes, including a self-assembled state with decreased Rho activity that results in genome-wide changes at the transcriptome level[32]. The seven human proteins selected in the present study for experimental characterization, exhibit predicted PrLDs with the same compositional properties than Rho or the yeast prions and are as well regulatory nucleic acid-binding proteins. We show here that they all contain a region able to autonomously self-assemble into amyloid-like structures displaying low cytotoxicity, a property that is likely linked to the weak hydrophobic patches in these assemblies, compared with those present in the

fibrils of pathogenic proteins like α-synuclein[1]. In addition, for all the cases we could test, the fibrillar assemblies had the capacity to seed and accelerate the aggregation reaction of their soluble counterparts. Despite our assays are *in vitro* and just with a piece of the protein, the results reveal the self-assembly potential encoded in human prion-like proteins.

Similar to Short Linear Interaction Motifs (SLIMs) in intrinsically disordered proteins, the assembling properties of soft amyloid cores might contribute to mediate PrLDs functional protein-protein interactions (PPI)[28]. Actually, it is known that the interaction between PrLDs of RNA granules proteins like TDP43, TAF15 and FUS is sufficient to induce liquid-liquid phase separation[7], but also reversible hydrogels[69], which have a high β-sheet content, but are labile to dilution. Interestingly enough, mutations that severely impede the formation of intermolecular β-sheets also inhibit the ability of PrLDs to activate transcription, indicating that, in this particular context, the formation of soft amyloid-like assemblies plays a functional role[70]. DDX5, ILF3 and TIA1 are also RNA granule proteins[43,71–73]; thus it is tempting to propose that the identified PrLDs and embedded soft amyloid cores might likewise play a role in the formation of functional β-sheet-containing assemblies. The other four proteins in our study are either structural components of macromolecular complexes (MED15 and PHC1) or they are involved in protein-protein interactions (NCOA2 and EYA1). The detected PrLDs and the sticky nature of their cores might facilitate these contacts. It has been suggested that the role of PrLDs would be to promote a primary level of organization through homotypic interactions and that, once oligomerized, these assemblies would facilitate the establishment of novel lateral interactions with other proteins[74]. Indeed, an analysis of the interaction networks of the seven proteins in this study using the STRING PPI network database[75] indicates that, as a trend, they tend to establish more PPI than the average human proteome (Table S6).

From an evolutionary point of view, the fact that the detected amyloid cores have been not purged out by natural selection support their functional role, since there is a strong selective pressure to reduce the amyloidogenic load of protein sequences, especially when they are located in disordered regions and thus exposed to solvent[76]. The presence of this kind of regions is inherently risky, since mutations that would increase their amyloid potency, making them more similar to classical pathogenic amyloid stretches, can promote irreversible protein aggregation, generating a deleterious phenotype resulting from deregulation of the homo or heterotypic PPIs in which they are involved. This is the case of hnRNPA1 and hnRNPA2, two human prion-like RNA-binding proteins in which point mutations mapping exactly in the pWALTZ predicted soft amyloid core increase the aggregation propensity of their PrLDs, causing the loss of their regulatory activity and leading to the onset of multisystem proteinopathy and ALS[14,27]. It is worth to explore whether the link to disease of the nucleic acid-binding human proteins in our subproteome and specifically of the seven polypeptides we studied here owes to a related mechanism. This would imply that a common process of mutation driven miss-assembly of cryptic amyloid cores in PrLDs might account for very different pathological phenotypes, from cancer to neurodegenerative disorders, depending on the affected protein and the pathways it regulates.

Methods

Computational identification of PrLDs in human proteins.
The human reference proteome dataset was downloaded from Uniprot[77] and scanned for PrLDs using PAPA[12] with the default parameters. From the initial 70940 proteins in the proteome, 663 prion-like candidates were identified. Their putative PrLDs were further evaluated with pWALTZ[14] with a threshold of 65 in order to identify those domains containing a soft amyloid core, which resulted in 535 final positive predictions. The protein sequences selected for experimental characterization were also analyzed with PLAAC[16] to define the boundaries of their PrLDs.

Peptide preparation.
We obtained the sequence of the 21 amino acid core region as predicted by pWALTZ[14] for the seven protein candidates selected for experimental characterization. The correspondent peptides were purchased from CASLO ApS (Scion Denmark Technical University). The lyophilized peptides were solubilized at a final concentration of 5 mM in dimethyl sulfoxide (DMSO) or in hexafluoro-2-propanol for CD analysis, in order to avoid the large increase in voltage caused by residual DMSO. Right before each experiment, the stock solutions were diluted to 100 μM in 5 mM sodium phosphate buffer pH 7,4. For aggregation assays the samples were incubated for 2 days at 37 °C with continuous agitation at 150 rpm in the presence of Teflon beads.

Synchronous light scattering.
Synchronous light scattering was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 360 nm, emission range from 350 to 370 nm, slit widths of 5 nm, 0.5 nm interval and 1000 nm/min scan rate. The peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before measurement. 100 μl of peptide solution was analyzed.

Bis-ANS (4,4-Dianilino-1,1-binaphthyl-5,5-disulfonate) binding.
The fluorescent spectrum of bis-ANS was analyzed using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 365 nm, emission range from 440 to 640 nm, slit widths of 5 nm, 0,5 nm interval and 1000 nm/min scan rate. The peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. 10 μl of peptide solution was added to 100 μl of 10 μM bis-ANS in H_2O. A 10 μM bis-ANS solution without peptide was used as a control.

Circular dichroism (CD) spectroscopy.
CD experiments were performed using a JASCO J-715 spectropolarimeter. Measurements of the far-UV CD spectra (260–190 nm) were made by the addition of 200 μl of the sample to a cuvette of 0.1 cm path-length. Spectra were recorded at room temperature, 1 nm band width and 100 nm/min scan rate. The resulting spectrum was the average of 10 scans. The contribution of the buffer was subtracted.

Fourier transform infrared (FT-IR) spectroscopy.   FTIR experiments were performed using a Bruker Tensor 27 FT-IR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Each spectrum consists of 16 independent scans, measured at a spectral resolution of $4\,cm^{-1}$ within the 1800–$1500\,cm^{-1}$ range. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data was afterwards deconvoluted using the Peak Fit 4.12 program. The buffer without peptide was used as a control and subtracted from the absorbance signal before deconvolution.

Transmission electron microscopy (TEM).   The morphology of the aggregated peptides was evaluated by negative staining and using a JEOL JEM-1400Plus Transmission Electron Microscope. $5\,\mu l$ of peptide solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with $5\,\mu l$ of 2% w/v uranyl acetate for 5 min. Then, grids were washed again before analysis.

Congo red (CR) binding.   CR binding to aggregated peptides was analyzed using a Specord® 200 Plus spectrophotometer (Analyticjena). The absorbance spectra were recorded from 400 to 650 nm. Spectra were acquired at 50 nm/sec scan rate. Peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. $10\,\mu l$ of the sonicated aggregated peptide was added to $100\,\mu l$ of $5\,\mu M$ CR in $5\,mM$ sodium phosphate buffer pH 7,4, and was incubated at room temperature for 5 min before the measurement. The same buffer with $5\,\mu M$ CR and without peptide was employed as a control.

Thioflavin-T (Th-T) binding.   The fluorescence spectra of Th-T were recorded using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm, slit widths of 5 nm, 0,5 nm interval and 1000 nm/min scan rate. Peptides were sonicated for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before dye addition. $5\,\mu l$ of the sonicated aggregated peptide was added to $100\,\mu l$ of $25\,\mu M$ ThT in $5\,mM$ sodium phosphate buffer pH 7,4. The same buffer with $25\,\mu M$ ThT and without peptide was employed as a control.

Thioflavin-S (Th-S) staining.   First, $150\,\mu l$ of aggregated peptides were incubated for 1 h in the presence of $125\,\mu M$ of ThS in $5\,mM$ sodium phosphate buffer pH 7,4. Then, the samples were washed two times with the same buffer. Finally, the precipitated fraction was resuspended in a final volume of $10\,\mu l$ and placed on a microscope slide and sealed. Images of the peptide aggregates bound to Th-S were obtained at 40-fold magnification in a Leica fluorescence microscope (Leica DMRB).

Aggregation kinetics and seeding assays.   Reactions were carried out at of $100\,\mu M$ final soluble peptide concentration in a solution containing $25\,\mu M$ of Th-T at 37 °C in the absence or presence of 2% seeds under quiescent conditions. The aggregation kinetics were followed monitoring the changes in Th-T fluorescence intensity at 488 nm over the time using a JASCO Spectrofluorometer FP-8200. Before each measure, the sample was mixed by pipetting up and down. The conditions of the spectra acquisition were: excitation wavelength of 440 nm, emission range from 460 to 600 nm, slit widths of 5 nm, 0,5 nm interval and 1000 nm/min scan rate. The seeds were prepared by sonicating the preformed aggregates of the corresponding peptide for 10 min in an ultrasonic bath (Fisher Scientific FB15052) before addition.

Cell viability assay.   Human SH-SY5Y cells were seeded into 96-well tissue culture plate with a density of 4,000 cells/well ($100\,\mu L$/well) in F-12 medium supplemented with 10% FBS, and maintained at 37 °C and 5% CO_2 atmosphere. Cell cultures were incubated in the presence of different concentrations of peptides for 72 hours. To control cells, the same volume of PBS1x was added. Following incubation, cells were stained by adding $20\,\mu L$ PrestoBlue® Cell Viability Reagent (Invitrogen) directly to the sample wells. After 30 min of incubation, cell viability was determined by measuring fluorescence exciting at 531 nm and collecting emission at 615 nm in a Victor fluorescent plate reader (Perkin Elmer).

References

1. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and hauman disease. *Annual Review of Biochemistry* **75**, 333–366 (2006).
2. Ventura, S. *et al*. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. USA* **101**, 7258–7263 (2004).
3. Rousseau, F., Serrano, L. & Schymkowitz, J. W. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* **355**, 1037–1047 (2006).
4. Aguzzi, A. & Calella, A. M. Prions: protein aggregation and infectious diseases. *Physiological reviews* **89**, 1105–1152 (2009).
5. Sikorska, B. & Liberski, P. P. Human prion diseases: from Kuru to variant Creutzfeldt-Jakob disease. *Sub-cellular biochemistry* **65**, 457–496 (2012).
6. Si, K. P. What Are They Good For? Annual Review of Cell and Developmental Biology 31, annurev-cellbio-100913-013409 (2015).
7. Molliex, A. *et al*. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
8. Chakrabortee, S. *et al*. Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits Article Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. *Cell* **167**, 1–13 (2016).
9. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **137**, 146–158 (2009).
10. Michelitsch, M. D. & Weissman, J. S. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11910–11915 (2000).
11. Harrison, P. M. & Gerstein, M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. *Genome biology* **4**, R40 (2003).
12. Toombs, J. a. *et al*. De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences* **109**, 6519–6524 (2012).

13. Espinosa Angarica, V. *et al*. PrionScan: an online database of predicted prion domains in complete proteomes. *BMC genomics* **15**, 102 (2014).

14. Sabate, R., Rousseau, F., Schymkowitz, J. & Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Computational Biology* **11**, e1004013 (2015).

15. Zambrano, R. *et al*. PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. Nucleic Acids Research, 1–7, (2015).

16. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).

17. Batlle, C., Iglesias, V., Navarro, S. & Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Rev Proteomics* **14**, 335–350 (2017).

18. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Research* **1462**, 61–80 (2012).

19. An, L. & Harrison, P. M. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biology direct* **11**, 32 (2016).

20. Malinovska, L., Palm, S., Gibson, K., Verbavatz, J.-M. & Alberti, S. Dictyostelium discoideum has a highly Q/N-rich proteome and shows an unusual resilience to protein aggregation. *Proceedings of the National Academy of Sciences* **112**, 201504459 (2015).

21. Singh, G. P. *et al*. Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum. *Molecular and Biochemical Parasitology* **137**, 307–319 (2004).

22. Espinosa Angarica, V., Ventura, S. & Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC genomics* **14**, 1–17 (2013).

23. Chakrabortee, S. *et al*. Luminidependens (LD) is an Arabidopsis protein with prion behavior. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 201604478 (2016).

24. Kato, M. *et al*. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).

25. Decker, C. J. & Parker, R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harbor perspectives in biology* **4**, a012286 (2012).

26. Patel, A. *et al*. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).

27. Kim, H. J. *et al*. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

28. Fernandez, M. R., Batlle, C., Gil-Garcia, M. & Ventura, S. Amyloid cores in prion domains: Key regulators for prion conformational conversion. *Prion* **11**, 31–39 (2017).

29. Sabate, R., Rousseau, F., Schymkowitz, J., Batlle, C. & Ventura, S. Amyloids or prions? That is the question. *Prion* **9**, 200–206 (2015).

30. Sant'Anna, R. *et al*. Characterization of Amyloid Cores in Prion Domains. *Scientific Reports* **6**, 34274 (2016).

31. Pallarès, I., Iglesias, V. & Ventura, S. The rho termination factor of Clostridium botulinum contains a prion-like domain with a highly amyloidogenic core. *Frontiers in Microbiology* **6**, 1–12 (2016).

32. Yuan, A. H. & Hochschild, A. A bacterial global regulator forms a prion. *Science* **355**, 198–201 (2017).

33. Li, X., Rayman, J. B., Kandel, E. R. & Derkatch, I. L. Functional Role of Tia1/Pub1 and Sup35 Prion Domains: Directing Protein Synthesis Machinery to the Tubulin Cytoskeleton. *Molecular Cell* **55**, 305–318 (2014).

34. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): Multi-tasking transcriptional regulators. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1829**, 756–763 (2013).

35. Janknecht, R. Multi-talented dead-box proteins and potential tumor promoters: P68 RNA helicase (DDx5) and its paralog, p72 RNA helicase (DDx17). *American Journal of Translational Research* **2**, 223–234 (2010).

36. Mazurek, A. *et al*. DDX5 regulates DNA replication and is required for cell proliferation in a subset of breast cancer cells. *Cancer discovery* **2**, 812–825 (2012).

37. Fuller-Pace, F. V. The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim Biophys Acta* **1829**, 756–763 (2013).

38. Xu, J. *et al*. Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. *Developmental Cell* **31**, 434–447 (2014).

39. Wu, K. *et al*. EYA1 phosphatase function is essential to drive breast cancer cell proliferation through cyclin D1. *Cancer Res* **73**, 4488–4499 (2013).

40. Buller, C., Xu, X., Marquis, V., Schwanke, R. & Xu, P. X. Molecular effects of Eya1 domain mutations causing organ defects in BOR syndrome. *Human molecular genetics* **10**, 2775–2781 (2001).

41. Xu, P. X., Cheng, J., Epstein, J. A. & Maas, R. L. Mouse Eya genes are expressed during limb tendon development and encode a transcriptional activation function. *Proc Natl Acad Sci USA* **94**, 11974–11979 (1997).

42. Castella, S., Bernard, R., Corno, M., Fradin, A. & Larcher, J. C. Ilf3 and NF90 functions in RNA biology. *Wiley Interdisciplinary Reviews: RNA* **6**, 243–256 (2015).

43. Shiina, N. & Nakayama, K. RNA granule assembly and disassembly modulated by nuclear factor associated with double-stranded RNA 2 and nuclear factor 45. *The Journal of biological chemistry* **289**, 21163–21180 (2014).

44. Saunders, L. R. *et al*. Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR. *The Journal of biological chemistry* **276**, 32300–32312 (2001).

45. Wang, X. *et al*. Redefining the modular organization of the core Mediator complex. *Cell Research* **24**, 796–808 (2014).

46. Halfmann, R. *et al*. Opposing Effects of Glutamine and Asparagine Govern Prion Formation by Intrinsically Disordered Proteins. *Molecular Cell* **43**, 72–84 (2011).

47. Zhu, X. *et al*. Mediator tail subunits can form amyloid-like aggregates *in vivo* and affect stress response in yeast. *Nucleic Acids Research* **43**, 7306–7314 (2015).

48. Xu, J. & Li, Q. Review of the *in vivo* functions of the p160 steroid receptor coactivator family. *Molecular Endocrinology* **17**, 1681–1692 (2003).

49. Silva, M. P. *et al*. NCOA2 is a candidate target gene of 8q gain associated with clinically aggressive prostate cancer. *Genes Chromosomes and Cancer* **55**, 365–374 (2016).

50. Suresh, S. *et al*. SRC-2-mediated coactivation of anti-tumorigenic target genes suppresses MYC-induced liver cancer. *PLoS genetics* **13**, e1006650 (2017).

51. Fleet, T. *et al*. SRC-2 orchestrates polygenic inputs for fine-tuning glucose homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6068–6077 (2015).

52. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nature reviews. Cancer* **6**, 846–856 (2006).

53. Awad, S. *et al*. Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. *Human Molecular Genetics* **22**, 2200–2213 (2013).

54. Conchillo-Sole, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC bioinformatics* **8**, 65 (2007).
55. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature biotechnology* **22**, 1302–1306 (2004).
56. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chemical Society reviews* **37**, 1395–1401 (2008).
57. Prilusky, J. *et al.* FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).
58. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
59. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et biophysica acta* **1804**, 996–1010 (2010).
60. Yang, Z. R., Thomson, R., McNeil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369–3376 (2005).
61. de Groot, N. S., Parella, T., Aviles, F. X., Vendrell, J. & Ventura, S. Ile-phe dipeptide self-assembly: clues to amyloid formation. *Biophys. J.* **92**, 1732–1741 (2007).
62. Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J. Histochem. Cytochem.* **37**, 1273–1281 (1989).
63. Sabate, R., Rodriguez-Santiago, L., Sodupe, M., Saupe, S. J. & Ventura, S. Thioflavin-T excimer formation upon interaction with amyloid fibers. *Chem. Commun. (Camb.)* **49**, 5745–5747 (2013).
64. Wickner, R. B. *et al.* Yeast prions act as genes composed of self-propagating protein amyloids. *Adv. Protein Chem.* **57**, 313–334 (2001).
65. Jarrett, J. T. & Lansbury, P. T. Jr. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055–1058 (1993).
66. Pastor, M. T., Esteras-Chopo, A. & Serrano, L. Hacking the code of amyloid formation: the amyloid stretch hypothesis. *Prion* **1**, 9–14 (2007).
67. Sabate, R., Espargaro, A., Grana-Montes, R., Reverter, D. & Ventura, S. Native structure protects SUMO proteins from aggregation into amyloid fibrils. *Biomacromolecules* **13**, 1916–1926 (2012).
68. Liebman, S. W. & Chernoff, Y. O. in Genetics Vol. 191, 1041–1072 (Genetics, 2012).
69. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
70. Kato, M. & Mcknight, S. L. Cross- b Polymerization of Low Complexity Sequence Domains. 1–11, (2016).
71. Gilks, N. *et al.* Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular biology of the cell* **15**, 5383–5398 (2004).
72. Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513–525 (2004).
73. Elvira, G. *et al.* Characterization of an RNA granule from developing brain. *Molecular & cellular proteomics: MCP* **5**, 635–651 (2006).
74. Kwon, I. *et al.* Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* **155**, 1049–1060 (2013).
75. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–452 (2015).
76. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737–742 (2007).
77. UniProt Consortium, T. U. UniProt: a hub for protein information. *Nucleic acids research* **43**, D204-212 (2015).
78. Rappaport, N. *et al.* MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res* **45**, D877–D887 (2017).

Acknowledgements

Author Contributions

Conceived and supervised: S.V. Performed the experiments: C.B., V.I. and S.N. Design, analysis and interpretation of experiments: C.B., N.S.G. and S.V. Wrote the paper: C.B. and S.V.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-09714-z

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.