

## MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation

Cristina Batlle<sup>1</sup>, Isabel Calvo <sup>2</sup>, Valentin Iglesias <sup>1</sup>, Cian J. Lynch<sup>2</sup>, Marcos Gil-Garcia<sup>1</sup>, Manuel Serrano <sup>2,3</sup> & Salvador Ventura<sup>1</sup>✉

A disordered to  $\beta$ -sheet transition was thought to drive the functional switch of Q/N-rich prions, similar to pathogenic amyloids. However, recent evidence indicates a critical role for coiled-coil (CC) regions within yeast prion domains in amyloid formation. We show that many human prion-like domains (PrLDs) contain CC regions that overlap with polyQ tracts. Most of the proteins bearing these domains are transcriptional coactivators, including the Mediator complex subunit 15 (MED15) involved in bridging enhancers and promoters. We demonstrate that the human MED15-PrLD forms homodimers in solution sustained by CC interactions and that it is this CC fold that mediates the transition towards a  $\beta$ -sheet amyloid state, its chemical or genetic disruption abolishing aggregation. As in functional yeast prions, a GFP globular domain adjacent to MED15-PrLD retains its structural integrity in the amyloid state. Expression of MED15-PrLD in human cells promotes the formation of cytoplasmic and perinuclear inclusions, kidnapping endogenous full-length MED15 to these aggregates in a prion-like manner. The prion-like properties of MED15 are conserved, suggesting novel mechanisms for the function and malfunction of this transcription coactivator.

<sup>1</sup>Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain. <sup>2</sup>Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain. <sup>3</sup>Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain. ✉email: [salvador.ventura@uab.es](mailto:salvador.ventura@uab.es)

Prions were first described as protein-only infectious agents in the context of mammalian neurological disorders<sup>1</sup>. Nevertheless, increasing evidence indicates that prion-like conformational conversion is not always deleterious; instead, it can be exploited for beneficial purposes<sup>2</sup>. The best characterized nonpathogenic prions are those of yeast, where they provide increased fitness in fluctuating environments<sup>3,4</sup>, mainly by regulating transcription and translation<sup>5</sup>. Yeast prions are modular proteins, and their conformational promiscuity is encoded at their prion domains (PrDs), which are both sufficient and necessary for prion conversion<sup>6</sup>. PrDs are long and disordered sequences of low complexity, often enriched in glutamine and/or asparagine (Q/N) residues. Polypeptides containing similar Q/N-rich prion-like domains (PrLDs) have been identified in other organisms, including humans, and they are generically named prion-like proteins<sup>2</sup>.

The aggregated state of yeast prions is macroscopically indistinguishable from that of pathogenic prions and amyloid proteins involved in neurodegenerative diseases<sup>7</sup>. Thus, it has been assumed that they all form amyloids through a common mechanism, which involves misfolding of the native structure, either folded or disordered, into  $\beta$ -sheet aggregation-prone conformations<sup>8</sup>. However, this generic and uncontrolled misfolding mechanism was difficult to reconcile with the evidence that, in contrast to pathogenic amyloids, the structural transitions of functional prions should be regulated. Fiumara and co-workers demonstrated that yeast Q/N-rich PrDs have an intrinsic propensity to form coiled-coils (CC), and they proposed that CC-based structural changes might explain better the physiological conformational switches of functional Q/N-rich prions than stochastic misfolding<sup>9</sup>.

As yeast PrDs, polyQ tracts can adopt CC structures<sup>10</sup>; however, in this case, CC-mediated conformational changes might be pathogenic. Several rare hereditary neurodegenerative diseases are originated when Q stretches exceed a critical length, and their severity increases with polyQ expansion<sup>11,12</sup>. These extensions would result in longer CC-prone helices, with a stronger polymerization propensity, which might ultimately cause their aggregation and the formation of the recurrent protein deposits observed in these diseases<sup>9</sup>.

Although CC-mediated aggregation provides a plausible mechanism for functional transitions in yeast PrDs and for the pathogenesis of polyQ-expansion disorders, it remains to be demonstrated that the aggregation of human prion-like proteins also accommodates to the CC model. Therefore, we undertook a bioinformatic analysis of the human proteome, which revealed that a significant proportion of human PrLDs indeed contained sequences of high CC propensity, which, in most cases, overlapped with Q-rich regions. This subgroup of prion-like proteins works in the regulation of transcription, and it comprises key transcription coactivators, including Mediator complex (Mediator) subunits 12 and 15 (MED12 and MED15).

Mediator is a large multi-protein complex that regulates most, if not all, enhancer-driven gene transcription<sup>13–15</sup>. Mediator is highly conserved among eukaryotes and consists of four modules: head, middle, tail, and the CDK8 kinase module<sup>16</sup>. The head and middle modules execute recruitment to the promoter regions, while the tail module mediates protein interactions with the transcription regulators at enhancers, and the CDK8 module is involved in RNA Polymerase II (Pol II) recruitment and release<sup>15</sup>. The tail module consists of 7 subunits, one of them being MED15, which, as described above, we identified as a CC forming prion-like protein.

MED15 is located in the cell nucleus, and its knockdown causes slow growth and reduced transcriptional activation<sup>17</sup>. MED15 is overexpressed in a wide range of human cancers: castration-resistant prostate cancer, head, and neck squamous cell carcinoma,

hepatocellular carcinoma, breast cancer, renal cell carcinoma, and testicular germ cell tumors<sup>18–25</sup>. Patients with MED15 overexpression in tumor tissues exhibit a more aggressive phenotype, associated with significantly shorter survival time<sup>18</sup>. Consequently, MED15 may serve as a prognostic marker as well as a potential therapeutic target in cancer.

The MED15 subunit was first discovered in yeast as Gal11<sup>26</sup> and later on renamed as yMED15<sup>27</sup>. yMED15 is involved in a variety of biological processes, such as the expression of galactose-inducible genes, and it is an interaction hub of many transcription factors (TFs)<sup>28</sup>. Interestingly, yMED15 contains a central Q-rich region reported to form amyloid-like intracellular inclusions spontaneously in yeast, whereas the aggregation of the full-length yMED15 occurs under stress conditions<sup>29</sup>. The aggregation of yMED15 provokes the loss of the entire tail module, reducing transcription levels, which has been suggested to be an epigenetic mechanism for transcription regulation, similar to those of classical yeast prions<sup>29</sup>.

Human MED15 (MED15) has a Q-rich region, consisting of discontinuous polyQ tracts at its N-terminus, which maps to its PrLD and is responsible for the predicted CC propensity of this domain (Fig. 1A). yMED15 and MED15 have low sequence identity and different distribution of their Q-rich regions; still, this kind of low complexity sequences have been reported to exert similar functions in orthologs without a strict positional requirement<sup>10</sup>. Therefore, the MED15 prion-like domain (MED15-PrLD) might possess cryptic amyloid and prion-like features yet to be discovered.

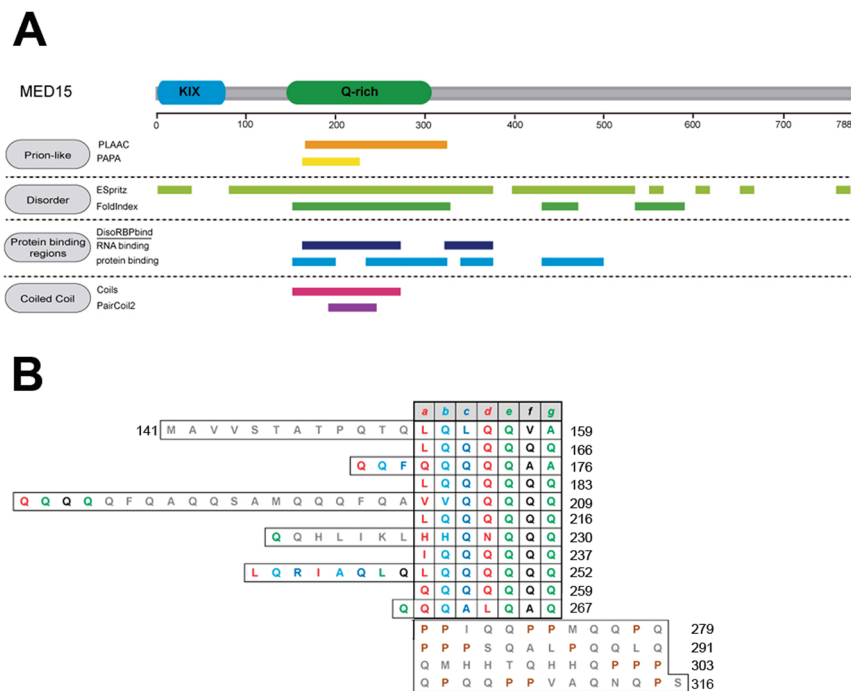
Here, we provide experimental evidence that the MED15-PrLD forms a dimer with CC structure. This domain bears the ability to self-assemble into Congo Red (CR) positive,  $\beta$ -sheet enriched amyloid fibrils, displaying concentration-dependent aggregation kinetics, and self-seeding activity. Moreover, as in functional yeast prions, a GFP globular domain adjacent to the PrLD maintains its fold and activity in the amyloid state. We demonstrate that it is the CC structure that regulates MED15-PrLD amyloid formation, its chemical or genetic disruption abolishing aggregation. Expression of MED15-PrLD in human cells promotes the formation of cytoplasmic and perinuclear inclusions, kidnapping the endogenous full-length protein to these aggregates in a prion-like manner.

Overall, this study reveals that amyloid formation by human prion-like proteins can follow the CC-mediated model reported in yeast and that the amyloid and prion-like properties of MED15 are conserved across species, which suggests novel mechanisms for MED15 function and malfunction.

## Results

**Human PrLDs with predicted coiled-coil domains overlap with polyQ tracts.** Coiled-coils (CC) are formed between proteins that contain repeats of seven amino acids (a-b-c-d-e-f-g) in which hydrophobic residues often occupy positions a/d to form a hydrophobic layer between the coiling helices; these heptad repeats are often discontinuous<sup>30,31</sup>. CC regions are over-represented in yeast prions, overlapping with their Q/N-rich prion domains, suggesting that the CC structure might represent a defining feature of these domains in yeast<sup>9</sup>. This observation prompted us to assess whether human prion-like proteins might share the same architecture.

First, we performed a prion-like amino acid composition (PLAAC) analysis to identify human prion-like proteins and their respective PrLDs at the proteome level. PLAAC is an algorithm that allows the identification of polypeptide regions with a composition similar to that of a set of well-characterized yeast PrDs<sup>32</sup>. We obtained a list of 193 reviewed human genes encoding for proteins with PrLDs (Supplementary Data 1).



**Fig. 1 MED15 contains an N-terminal Q-rich coiled-coil PrLD.** **A** MED15 diagram showing the location of the KIX (blue) and Q-rich (green) regions. Prediction of prion-like domains using PLAAC<sup>32</sup> and PAPA<sup>72</sup> predictors, disorder regions using ESpritz<sup>38</sup> and FoldIndex<sup>37</sup> predictors, protein binding regions using DisoRBPbind<sup>39</sup> predictor, and coiled-coil regions using COILS<sup>33</sup> and PairCoil2<sup>70</sup> predictors are shown below the diagram. **B** Schematic diagram of the heptad repeats (a-b-c-d-e-f-g) predicted by COILS<sup>33</sup> (14-residue window) of MED15-PrLD (residues 141–316). Gray residues indicate COILS score <0.2.

We analyzed which of the identified PrLDs might fold into a CC using COILS<sup>33</sup>, an algorithm that detects CC heptad repeats in primary sequences (Supplementary Data 1). We found that 12% of human PrLDs ( $n = 22$ ) were predicted to have CC regions. In all, 86% of these proteins ( $n = 19$ ) is involved in transcriptional regulation (Supplementary Data 1). Accordingly, a Gene Ontology (GO) analysis of the 22 candidates<sup>34</sup> indicated a significant enrichment in the following ontologies: (i) transcription initiation from RNA polymerase II promoter in biological process, (ii) transcription coactivator activity in molecular function, and (iii) nucleus and nucleoplasm in cellular component (Supplementary Data 1). Importantly, 90% of the predicted CC regions ( $n = 20$ ) correspond to or overlap with polyQ tracts in the PrLDs (Supplementary Data 1 and Fig. S1), which is consistent with the observation that polyQ proteins are functionally biased towards transcriptional regulation<sup>10,35</sup>. These 20 prion-like candidates include proteins whose polyQ expansion is connected to autosomal dominant cerebellar ataxia, such as the TATA-box-binding protein (TBP) and ATXN1<sup>11</sup>, as well as the CREB-binding protein (CBP), a transcriptional coactivator necessary for the survival of many neurons which is recruited into aggregates in polyQ diseases<sup>36</sup>. Overall, our analysis indicates that, as it occurs in yeast prions, a significant fraction of human prion-like proteins might contain CC that overlap with polyQ tracts at their PrLDs. Proteins displaying these features include important transcriptional coactivators like CBP, TBP, KMT2D, FOXP2, NCOA2, NCOA3, MAML2, MED12, and MED15.

Apart from those involved in disease, the amyloid and prion-like properties of the PrLDs identified in these transcriptional regulators remain to be experimentally demonstrated. However, the yMED15 polyQ can form amyloid-like aggregates on its own<sup>29</sup>. Although human and yeast MED15 proteins display very low sequence identity (12%), we decided that it was worthwhile to explore whether the identified human MED15-PrLD can access

an amyloid state and propagate this conformation in a prion-like manner.

**MED15 contains an N-terminal Q-rich coiled-coil PrLD.** MED15 is highly enriched in Gln residues (Q, 20% of residues), especially at its N-terminus, with multiple short polyQ tracts (of 2–16 Qs; Fig. 1). It is precisely this N-terminal Q-rich region that is identified as a PrLD by PLAAC (Fig. 1A and Supplementary Data 1).

PrLDs are thought to be intrinsically disordered in their monomeric states, in such a way that their self-assembly can occur spontaneously, without an energy-dependent conformational change<sup>2</sup>. The MED15-PrLD is consistently predicted to be disordered by FoldIndex and ESpritz predictors (Fig. 1A)<sup>37,38</sup>. Disordered regions can be cryptic protein binding sites that, upon interaction, undergo a disorder to order transition; DisoRDPbind is an algorithm aimed to predict such transitions, scoring putative associations between the disordered region of interest and RNA, DNA, or protein<sup>39</sup>. MED15-PrLD was predicted to interact with both RNA and protein (Fig. 1A), suggesting that it is susceptible to experiment conformational switches. Indeed, polyQ tracts are known to mediate protein–protein interactions (PPI) and multimerization, which can lead to the formation of homomeric or heteromeric CC structures<sup>10</sup>. Accordingly, the MED15 N-terminal Q-rich region is predicted to adopt a CC conformation (Fig. 1A and Supplementary Data 1) and has been shown to exhibit  $\alpha$ -helical structure when fused to a globular domain<sup>40</sup>. This CC domain is discontinuous and consists of three consecutive CC regions (Fig. S2). As expected, in the MED15-PrLD CC model, almost all “a” and “d” positions in the helix are occupied by hydrophobic amino acids or Gln, which acts as an ambivalent hydrophobe (Fig. 1B)<sup>41</sup>. The N-terminal Q-rich segment is the less conserved region in MED15 vertebrate orthologs, with significant differences in the number, length, and

location of polyQ tracts among species; however, in all cases, it is predicted to form CC (Fig. S2), suggesting that this region is structurally conserved, likely, because of its functional significance.

Overall, the bioinformatics analysis of MED15 suggests that we are in front of a prion-like protein with a disordered N-terminal Q-rich PrLD in its monomeric state, but with the ability to undergo a transition to a dimeric or higher-order CC structure.

### MED15 interactors are predicted to have coiled-coil domains.

Recently, a list of Mediator complex interactors was identified by performing FLAG-MED15 immunoprecipitation in mice<sup>42</sup>. Therefore, at least a fraction of the identified binders may contact MED15 directly. All mouse MED15 potential interactors display >75% sequence identity with their human homologs (Supplementary Data 2). In the light of the predicted propensity of MED15 to form CC, we examined the CC propensity of their putative interactors using COILS<sup>33</sup>. Interestingly enough, 50% of these candidates resulted in positive predictions, indicating that CC-containing proteins are overrepresented among MED15 binders (Supplementary Data 2). Not surprisingly, a GO term enrichment analysis indicated that this subset of proteins is preferentially located at the nucleoplasm, acting as transcription coactivators and with major involvement in Pol II transcription (Supplementary Data 2). Importantly, among these polypeptides, we found four of the CC-containing prion-like proteins identified in the previous section: NCOA2, MAML2, CBP, and MED12, which suggests that their PrLDs compositional bias and propensity to populate CC conformations might result in their eventual interaction with MED15. Indeed, an analysis with the STRING protein interactions database<sup>43</sup>, indicates that all these five proteins are closely associated (Fig. S3; PPI enrichment  $p$ -value:  $7.5 \times 10^{-08}$ ).

**Soluble MED15-PrLD characterization.** Long polyQ stretches are commonly followed by polyproline (polyP) tracts at their C-terminus. These repeats stabilize and stop the CC region<sup>10,44</sup>, which suggests that the two tracts form a functional/structural unit and that the expression of the polyQ alone may not recapitulate the physiological context of the sequence. In MED15, a Pro-rich segment is also adjacent to the Q-rich region (Fig. 1B), and both stretches are predicted to be part of the PrLD. The presence of Pro residues at the C-terminus of polyQ tracts is also frequent in MED15 orthologs (Fig. S4).

Most proteins with polyQ tracts are modular, and these stretches are in close vicinity to globular domains. Indeed, these low complexity regions are difficult to express recombinantly and to purify in the absence of a globular partner that modulates their inherent aggregation propensity<sup>45</sup>. The MED15-PrLD is flanked at its N-terminus by a globular KIX domain (Fig. 1A), a feature conserved in MED15 orthologs (Fig. S4).

With the above considerations in mind, we expressed the MED15-PrLD domain, containing both Q-rich and Pro-rich regions, fused at the C-terminus of GFP, which would act as a functionally traceable globular domain, hereinafter named as GFP-MED15CC. This construct should allow us to characterize the soluble, and eventually, the aggregated states of MED15-PrLD in a mimic of its sequential framework.

As intended, GFP-MED15CC was purified from the soluble cellular fraction. The GFP globular domain integrity was evaluated by monitoring the GFP absorption and fluorescence emission spectra (Fig. 2A, B). GFP-MED15CC spectral properties were identical to those of GFP alone, with an absorption maximum at 490 nm and an emission fluorescence maximum at 510 nm. In order to verify that GFP-MED15CC was not

aggregated in the soluble state, we used synchronous light scattering (Fig. 2C). Neither GFP nor GFP-MED15CC showed a significant increase in light scattering signal, compared to the buffer alone.

The GFP-MED15CC secondary structure was assessed by far-UV circular dichroism (CD; Fig. 2D). GFP alone showed a  $\beta$ -sheet spectrum with a characteristic minimum at 218 nm. In contrast, the GFP-MED15CC spectrum exhibited two minima at 208 and 222 nm, characteristic of  $\alpha$ -helices. The deconvolution of the spectra using the K2D program<sup>46</sup> indicated a predominant  $\alpha$ -helix conformation for GFP-MED15CC (100%) and  $\beta$ -sheet for GFP (78%). Thus, the strong  $\alpha$ -helix signal in GFP-MED15CC completely masks the signal of GFP, something that is characteristic of globular domains-CC fusions<sup>47</sup>.

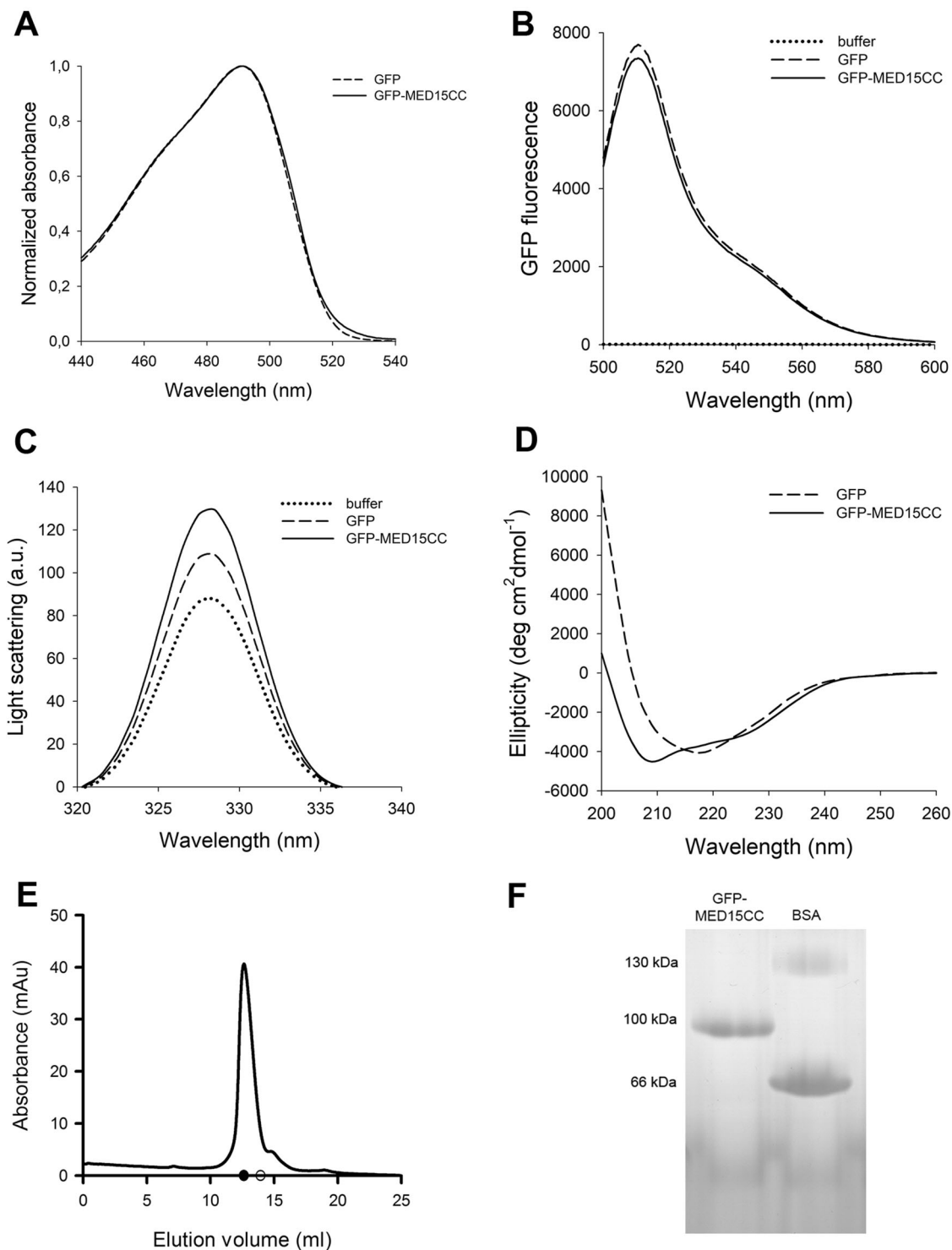
As described above, MED15-PrLD is predicted to be intrinsically disordered in its monomeric state. The fact that we observed a predominant  $\alpha$ -helical secondary structure by far-UV CD is consistent with this region adopting a CC. This necessarily implies that, in solution, GFP-MED15CC acquires quaternary structure. GFP-MED15CC is 458 residues long, with an expected MW of 52 kDa. Size exclusion chromatography (SEC) and native PAGE were performed to evaluate the oligomerization state of GFP-MED15CC (Fig. 2E, F). Its SEC elution pattern corresponded to a protein size of  $\sim$ 100 kDa (elution at 12.65 ml), indicating that the protein was a dimer (Fig. 2E). In contrast, GFP (246 residues, MW 28 kDa) was eluted as a monomer (elution at 15 ml). In agreement with the SEC data, GFP-MED15CC migrates as a unique band of  $\sim$ 100 kDa in a native PAGE, located between the monomeric and dimeric forms of bovine serum albumin (BSA) used as control (Fig. 2F). Besides, dynamic light scattering (DLS) analysis revealed major populations with average diameters of 13 and 7 nm for GFP-MED15CC and GFP, respectively, further supporting that the MED15 CC domain drives the formation of a dimeric structure.

In conclusion, in solution, GFP-MED15CC is a dimer sustained by a CC interaction, where the adjacent globular domain retains its integrity and function.

**MED15-PrLD aggregates into  $\beta$ -sheet enriched amyloid-like fibrillar structures.** Both PrLDs and polyQ tracts are well known for their ability to aggregate into amyloid  $\beta$ -sheet conformations<sup>45,48</sup>. Therefore, we assessed if GFP-MED15CC aggregates spontaneously using synchronous light scattering, GFP fluorescence, far-UV CD, and infrared spectroscopy (ATR-FTIR; Fig. 3A–D). GFP was used as a control in all these assays.

The proteins were incubated at 5  $\mu$ M (0.25 mg/ml) for 24 h at 37 °C under quiescent conditions, and the light scattering of the correspondent solutions measured (Fig. 3A, B). In contrast to GFP, GFP-MED15CC exhibited a substantial increase in the light scattering signal upon incubation (Fig. 3A). GFP fluorescence was examined in the samples' supernatant after their centrifugation (see Methods section). In agreement with the light scattering data, the fluorescence of the GFP-MED15CC in the soluble fraction decreased significantly upon incubation, whereas that of GFP remained mostly the same (Fig. 3B). Together, these results demonstrate that GFP-MED15CC aggregates within 24 h.

We monitored the secondary structure of incubated GFP-MED15CC solutions by far-UV CD and fourier transform infrared (FTIR; Fig. 3C, D). GFP-MED15CC suffered a conformational change from the initial  $\alpha$ -helical to a  $\beta$ -sheet structure after 24 h of incubation, displaying a spectrum minimum at 218 nm, similar to that of GFP (Fig. 3C). Prolonged incubation for 4 days resulted in a weak  $\beta$ -sheet signal, indicating that a significant proportion of the protein was aggregated out of the solution, becoming undetectable by CD. The conformation of

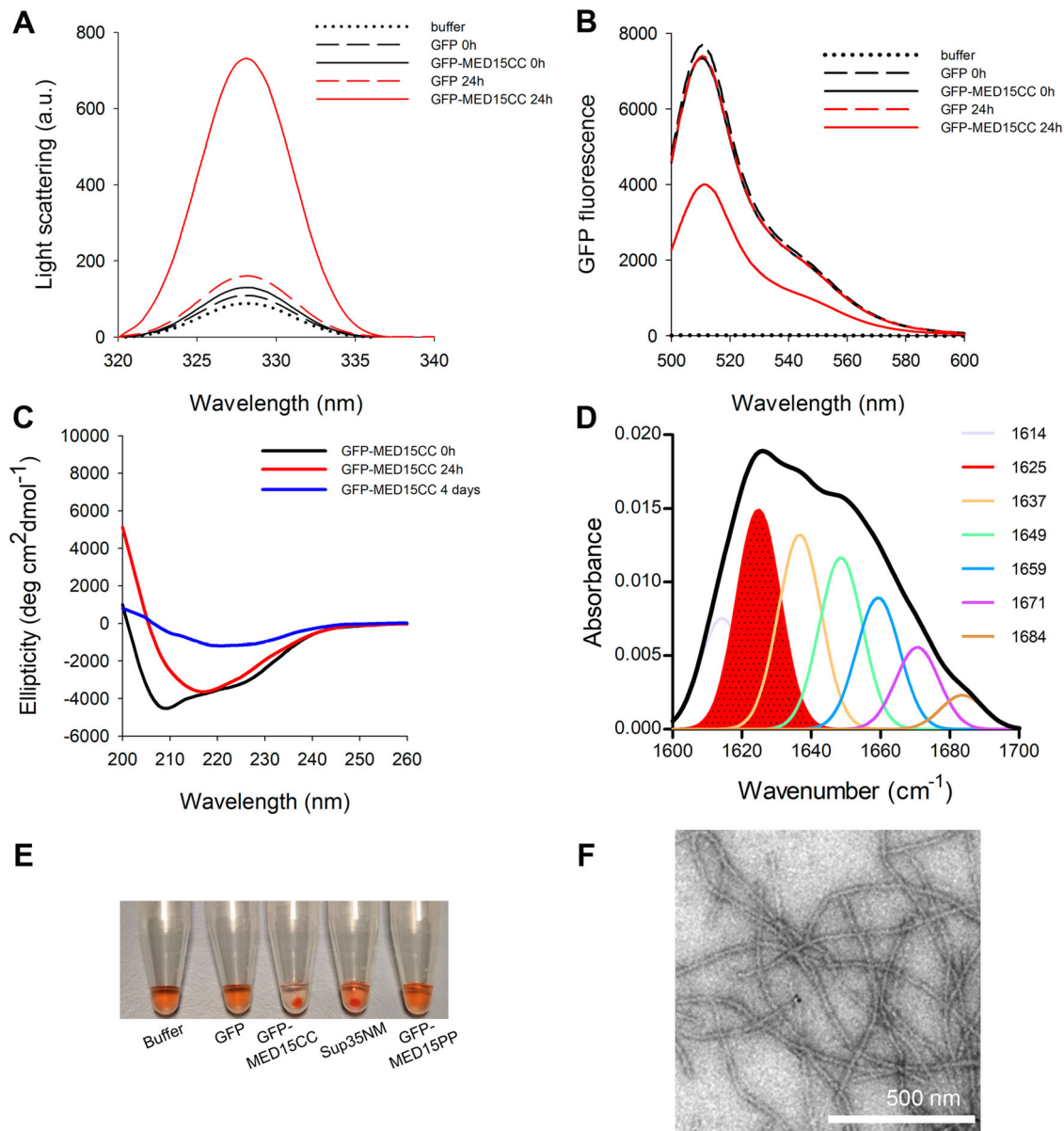


**Fig. 2** MED15-PrLD soluble characterization. **A** Absorbance, **B** fluorescence, **C** light scattering, and **D** circular dichroism of 5  $\mu$ M GFP and GFP-MED15CC. **E** GFP-MED15CC size exclusion chromatography elution profile. Filled circle and empty circle in the x axis indicates 100 and 50 kDa position, respectively. **F** 12.5% blue native PAGE of GFP-MED15CC and bovine serum albumin (BSA) at 0.5 mg/ml. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

control GFP remained unaffected after 4 days of incubation (Fig. S5). To further confirm the putative  $\beta$ -sheet secondary structure of incubated GFP-MED15CC, we analyzed its FTIR spectrum in the amide I region ( $1700\text{--}1600\text{ cm}^{-1}$ ), which corresponds to the absorption of the main chain carbonyl group and is sensitive to the protein conformation. Deconvolution of the FTIR spectrum of incubated GFP-MED15CC resulted in a

major peak at  $1625\text{ cm}^{-1}$ , which is typically attributed to the presence of intermolecular  $\beta$ -sheet structure (Fig. 3D), in good agreement with the CD data.

Next, we evaluated whether GFP-MED15CC aggregates were arranged into an amyloid-like architecture. GFP-MED15CC emits green fluorescence; hence, we could not use Thioflavin-T<sup>49</sup>. Instead, we used a CR precipitation assay since this dye



**Fig. 3** MED15-PrLD aggregates into  $\beta$ -sheet enriched amyloid-like fibrillar structures. **A** Light scattering, **B** GFP fluorescence (supernatant fraction) and **C** circular dichroism of soluble ( $t = 0$  h, black) and aggregated ( $t = 24$  h, red, or  $t = 4$  days, blue)  $5 \mu\text{M}$  GFP-MED15CC or GFP. **D** FTIR of aggregated ( $t = 24$  h)  $5 \mu\text{M}$  GFP-MED15CC. **E** Congo Red precipitation of incubated ( $t = 9$  h)  $5 \mu\text{M}$  GFP, GFP-MED15CC, GFP-MED15PP, and  $10 \mu\text{M}$  Sup35NM. **F** Transmission electron microscopy image of aggregated ( $t = 9$  h)  $5 \mu\text{M}$  GFP-MED15CC. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

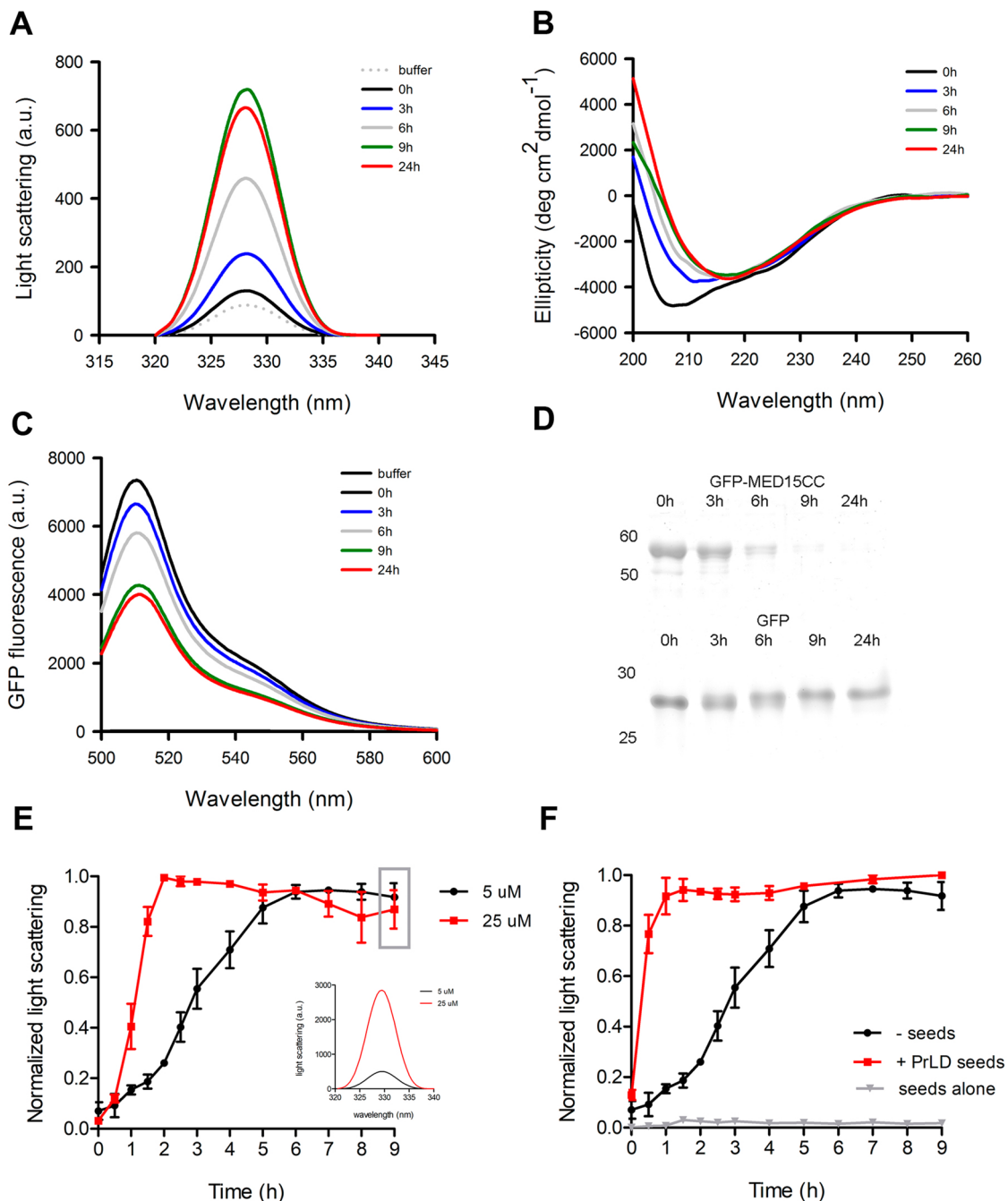
deposits on top of amyloid fibrils<sup>50</sup>. We incubated CR with buffer only, or with GFP, as negative controls, and with Sup35NM yeast prion amyloid fibrils as a positive control<sup>5</sup>. After 1 h of incubation, only GFP-MED15CC and Sup35NM formed a red pellet, suggesting that MED15CC aggregates are amyloid-like (Fig. 3E). Next, we addressed the morphological features of GFP-MED15CC aggregates by transmission electron microscopy (TEM) to further confirm its amyloidogenic nature (Fig. 3F). Negative staining showed long, thin, and unbranched amyloid-like fibrillar structures, without any significant accumulation of amorphous material. The fibrils exhibited a diameter of  $19 \pm 4$  nm and a length of  $2 \pm 0.5 \mu\text{m}$ .

In contrast to pathogenic proteins, a remarkable property of yeast PrDs is that their assembly into amyloids does not necessarily imply the misfolding of the attached globular domains, which might keep their native conformation in the aggregated state<sup>51</sup>. In the same manner, GFP-MED15CC fibrils

were green fluorescent, indicating the maintenance of the native GFP fold within the fibril phase (Fig. S6).

Overall, these results indicate that the PrLD of the initially dimeric GFP-MED15CC protein promotes the spontaneous assembly into supramolecular  $\beta$ -sheet enriched fibrillar amyloid structures in which the adjacent globular domain keeps its functional conformation, all these features being consistent with a prion-like nature.

**MED15-PrLD aggregation kinetics.** In the previous section, we showed that GFP-MED15CC aggregates into amyloid fibrils. To further characterize this reaction, we followed its kinetics, at  $5 \mu\text{M}$  protein concentration and  $37^\circ\text{C}$  without agitation. We monitored the reaction by synchronous light scattering, far-UV CD, and GFP fluorescence (Fig. 4A–D). Upon incubation, GFP-MED15CC showed a progressive increase in light scattering, indicative of protein aggregation (Fig. 4A). Far-UV CD analysis indicated that



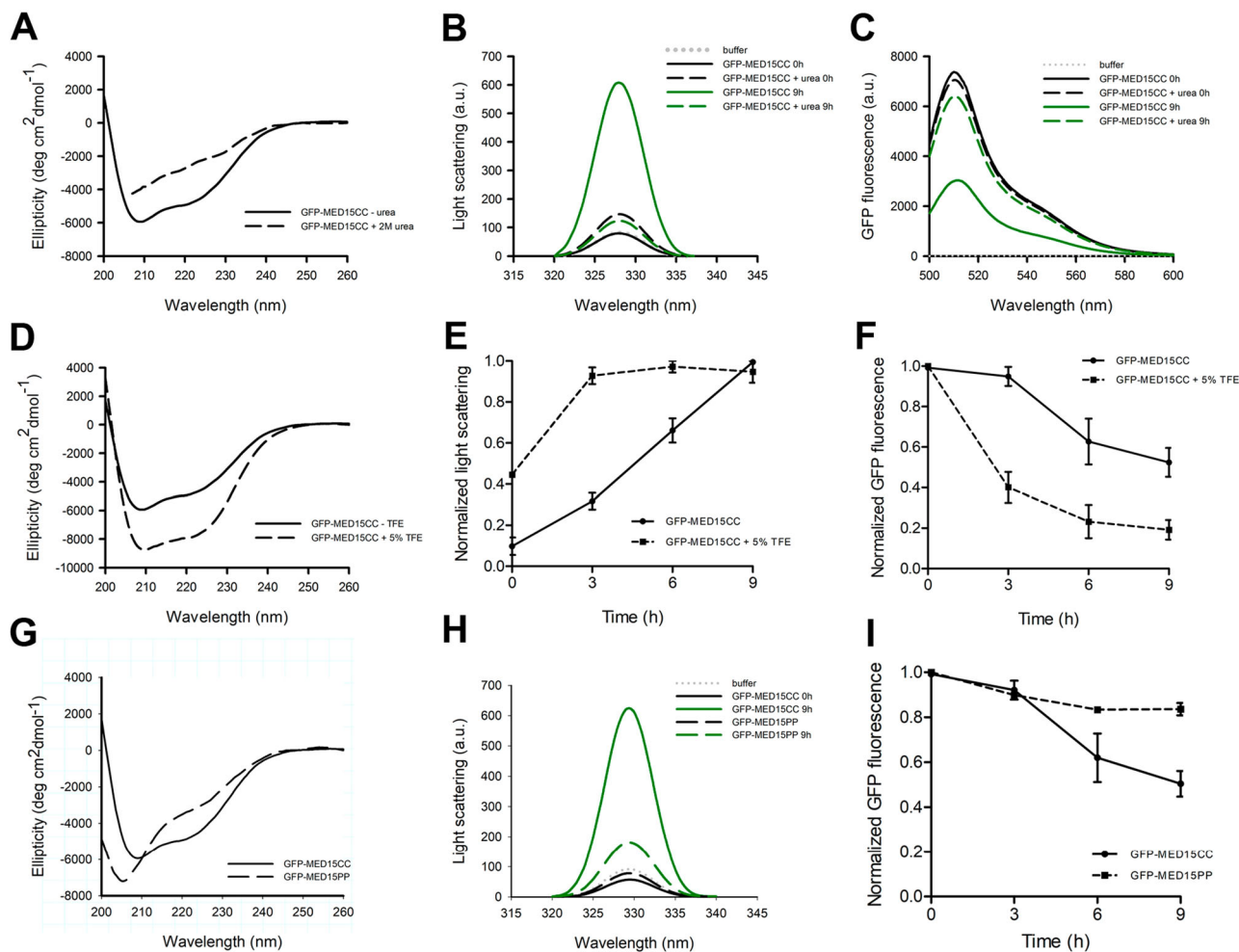
**Fig. 4 MED15-PrLD aggregation kinetics characterization.** **A** Light scattering, **B** circular dichroism, and **C** GFP fluorescence (supernatant fraction) of 5  $\mu\text{M}$  GFP-MED15CC incubated at 37  $^{\circ}\text{C}$  at the indicated time points. **D** 12% SDS-PAGE of 5  $\mu\text{M}$  GFP and GFP-MED15CC supernatant fraction after centrifugation at the indicated time points of incubation. **E** Light scattering kinetics of 5 and 25  $\mu\text{M}$  GFP-MED15CC at 37  $^{\circ}\text{C}$ . The inset shows the real scattering intensities at the final time point (9 h, gray rectangle). **F** Light scattering kinetics of 5  $\mu\text{M}$  GFP-MED15CC incubated at 37  $^{\circ}\text{C}$  in the absence or presence of 2% GFP-MED15CC seeds. Kinetics of seeds alone is shown as control. All assays are performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

this is accompanied by a conformational change from an  $\alpha$ -helical to a  $\beta$ -sheet secondary structure (Fig. 4B). GFP-MED15CC sample centrifugation at the same time points confirms a gradual decrease in the amount of soluble protein, with a concomitant decrease of GFP fluorescence in the supernatant (Fig. 4C, D). All the techniques converge to indicate that 9 hours of incubation are enough to reach the plateau phase of the aggregation reaction. In contrast, the parallel GFP control did not show signs of aggregation by any of the used techniques (Figs. 3A–C and 4D).

Aggregation into amyloids responds to a second or higher reaction and thus is, typically, very sensitive to the protein

concentration. We evaluated if this is the case for GFP-MED15CC. We monitored GFP-MED15CC aggregation kinetics at 5 and 25  $\mu\text{M}$  by synchronous light scattering and observed a clear acceleration and signal increase at the higher concentration, as expected for a canonical amyloid protein (Fig. 4E).

Many amyloids, and specially prion-like proteins, have the ability to self-propagate by seeding the aggregation of their soluble counterparts<sup>52</sup>. In this way, the presence of small amounts of preformed amyloid fibrils usually accelerates the protein aggregation kinetics, mostly by reducing or even abrogating the lag phase of the reaction, which corresponds to



**Fig. 5 MED15-PrLD aggregation is governed by its coiled-coil conformation.** **A** Circular dichroism, **B** light scattering, and **C** GFP fluorescence (supernatant fraction) of 5  $\mu\text{M}$  GFP-MED15CC in the absence or presence of 2 M urea. **D** Circular dichroism, **E** light scattering kinetics, and **F** GFP fluorescence kinetics (supernatant fraction) of 5  $\mu\text{M}$  GFP-MED15CC in the absence or presence of 5% TFE. **G** Circular dichroism, **H** light scattering, and **I** GFP fluorescence kinetics (supernatant fraction) of 5  $\mu\text{M}$  GFP-MED15CC or GFP-MED15PP. In all cases, aggregation was performed for 9 hours at 37  $^{\circ}\text{C}$  without agitation in 20 mM Tris pH 7.5 and 150 mM NaCl.

the formation of the initial aggregation nuclei. In Fig. 4F we compare the aggregation kinetics of GFP-MED15CC at 5  $\mu\text{M}$  in the absence or presence of 2% (w/w) of its preformed fibrils. It can be observed how the presence of fibrils dramatically accelerates GFP-MED15CC aggregation. This result indicates that GFP-MED15CC amyloid formation follows a nucleation-polymerization mechanism and that, at least in vitro, the MED15-PrLD can propagate its amyloid conformation.

**MED15-PrLD aggregation is governed by its coiled-coil conformation.** Pathogenic polyQ tracts are generally thought to misfold spontaneously into aggregation-prone  $\beta$ -sheet conformations. However, recent evidence suggests that their aggregation depends instead on the population of metastable CC structures<sup>9</sup>. To test whether this mechanism drives GFP-MED15CC amyloid formation, we evaluated its aggregation in the presence of two reagents aimed at weakening or strengthening the CC domain selectively.

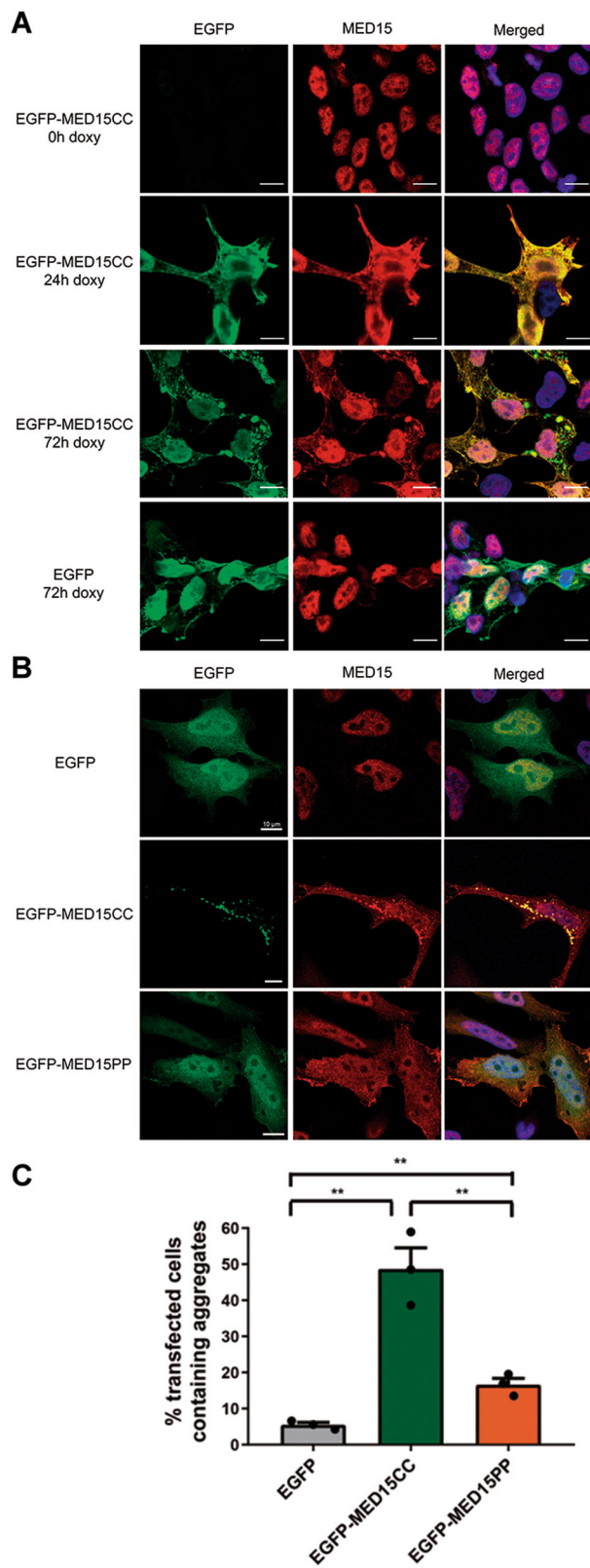
In the presence of 2 M urea, the GFP-MED15CC  $\alpha$ -helical conformation was significantly disrupted, as assessed by far-UV CD, and the spectrum shifted towards the disordered region (Fig. 5A). Importantly, the control GFP conformation was not affected by this denaturant (Fig. S7A). We analyzed the GFP-MED15CC aggregation propensity in the presence or absence of

2 M urea by synchronous light scattering and GFP fluorescence (Fig. 5B, C). GFP-MED15CC did not show any signs of aggregation in the denaturing agent's presence, indicating that CC disruption prevented protein aggregation. It is important to point out that, as expected, the integrity of the GFP moiety of GFP-MED15CC was not affected by urea because the protein remained fluorescent. The control GFP did not show any aggregation with, or without, urea (Fig. S7B, C).

TFE is a reagent used to enhance or stabilize  $\alpha$ -helical conformations. However, the presence of >20% TFE might induce non-native  $\alpha$ -helices, blurring any physiologically relevant interpretation of the data; to avoid this effect, we used an unusually low concentration of this cosolvent in our experiments. In the presence of 5% TFE, the CC of GFP-MED15CC was significantly enhanced as observed by far-UV CD (Fig. 5D), whereas the structure of the control GFP was unaffected (Fig. S7D). Evaluation of GFP-MED15CC aggregation kinetics by synchronous light scattering and GFP fluorescence showed a significant acceleration of the process in the presence of TFE (Fig. 5E, F). Control GFP did not show signs of aggregation with, or without, TFE (Fig. S7E, F).

MED15-PrLD contains discontinuous CC segments, accounting for a total of 11 heptad repeats, according to the COILS predictor (Fig. 1B). To confirm the predominant role of MED15 CC on the PrLD aggregation, we used structure-guided





**Fig. 6 MED15-PrLD forms insoluble cytoplasmic inclusions in mammalian cells.** **A** Cellular localization by immunofluorescence of EGFP or EGFP-MED15CC in HEK293T cells after addition of doxycycline (doxy) for 0, 24, or 72 h. **B** Cellular localization by immunofluorescence of EGFP, EGFP-MED15CC, and EGFP-MED15PP in HeLa cells after expression for 24 h. In both **A** and **B**, cells were stained with MED15 antibody (red) to evaluate the effect of transfection in the endogenous mediator subunit, and with DAPI (blue) as nuclear marker. **C** Graph of percentage transfected cells containing cytoplasmic or perinuclear aggregates after EGFP, EGFP-MED15CC, and EGFP-MED15PP expression in HeLa cells. Dots represent the mean of independent triplicates.

We first analyzed GFP-MED15PP secondary structure in solution by far-UV CD. The variant exhibited a minimum at 205 nm correspondent to a random coil conformation, although a small shoulder at 222 nm was still evident. In any case, the spectrum confirmed that the CC fold was significantly destabilized/disrupted by the introduced amino acid changes (Fig. 5G). Consistent with this observation, the GFP-MED15PP aggregation propensity was clearly reduced, relative to the GFP-MED15CC protein, as evidenced by the synchronous light scattering and GFP fluorescence signals after 9 h of incubation at 37 °C (Fig. 5H, I). Indeed, when incubated under the same conditions that GFP-MED15CC, the mutant protein does not bind to CR (Fig. 3E).

All in all, these results indicate that its CC conformation mediates MED15-PrLD amyloid formation and that disruption of this structure significantly reduces its aggregation potential. This aggregation mechanism is consistent with the one described for the functional switch of Q/N-rich yeast prions<sup>9</sup>.

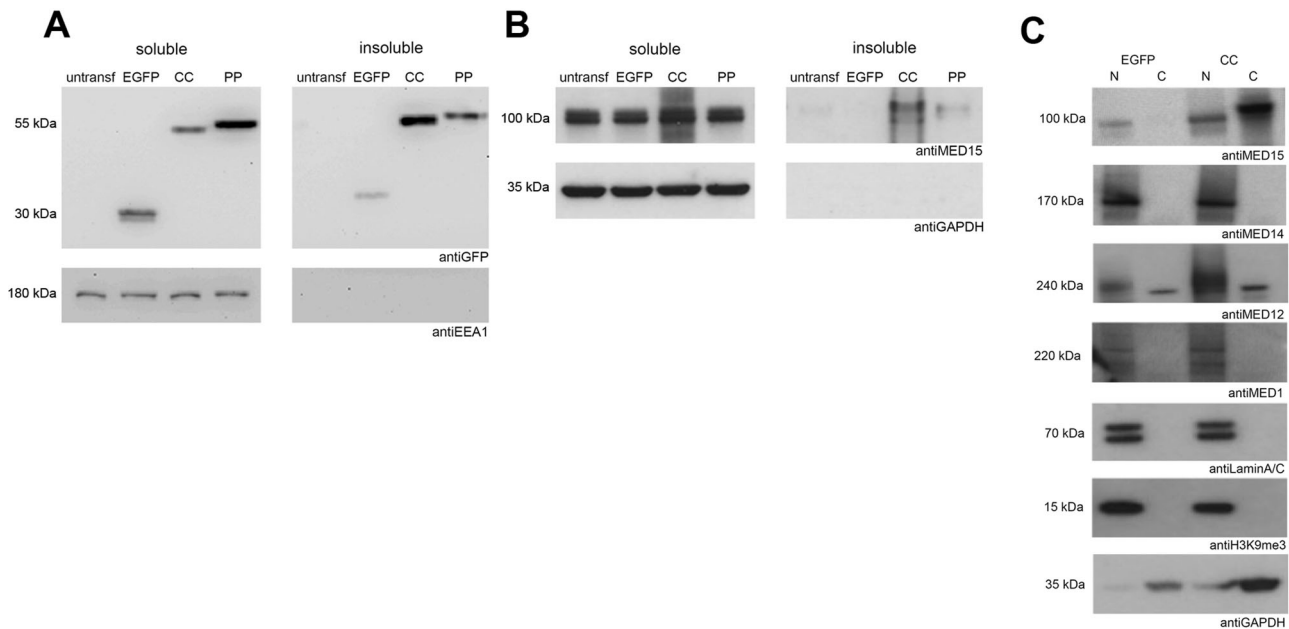
**MED15-PrLD forms insoluble cytoplasmic inclusions in mammalian cells and kidnaps the endogenous protein.** In the previous sections, we have shown that MED15-PrLD forms amyloids and that the CC is a central player in this reaction. To determine whether MED15-PrLD may aggregate in a cellular context, we stably transfected HEK293T cells with lentiviral vectors containing EGFP-tagged MED15-PrLD (EGFP-MED15CC) or EGFP, as a control, and expressed the two proteins using a doxycycline-inducible system (TetO).

MED15 is a component of the Mediator tail module, and, as expected, it displays a diffuse nuclear localization in non-induced cells (Fig. 6A). In all, 24 h after induction of protein expression EGFP-MED15CC formed small aggregates distributed homogeneously in the cytoplasm, whereas at 72 hours multiple larger cytoplasmic and perinuclear inclusions were observed, whereas EGFP remained diffusely distributed in the cell (Fig. 6A).

As in the case of HEK293T cells, transfection of HeLa cells with lentiviral vectors containing EGFP-MED15CC or EGFP and expression of the two proteins using the TetO-system resulted in the aggregation of EGFP-MED15CC in the cytoplasm, with EGFP remaining soluble (Fig. S9A).

To demonstrate that the EGFP-MED15CC ability to form intracellular inclusions is expression system independent and that this feature depends on the CC region of the PrLD, we transiently transfected EGFP, EGFP-MED15CC, and the EGFP-MED15PP variant, containing the five disrupting Pro substitutions, in HeLa cells and expressed the three proteins for 24 hours. EGFP-MED15CC expression resulted again in the formation of cytoplasmic and perinuclear inclusions, which occur in ~50% of the transfected cells (Figs. 6B, C and S9B). In contrast, EGFP-MED15PP formed inclusions in only ~15% of the transfected cells ( $p$ -value = 0.0063), and they were small (Fig. 6B, C). Inclusion formation is a hallmark of protein aggregation, and aggregated amyloids exhibit protein detergent insolubility. In agreement with their relative in vitro amyloid-forming and

mutagenesis to decrease the CC propensity by introducing Pro substitutions in the wild-type protein, a strategy commonly used for  $\alpha$ -helix disruption<sup>53</sup>. We substituted five residues in positions “a” of the heptad repeats with a single Pro (Fig. S8A). These substitutions are predicted by COILS to disrupt the CC domain significantly (Fig. S8B). Hereinafter we named this mutant as GFP-MED15PP.



**Fig. 7 MED15-PrLD solubility and endogenous MED15 localization in HeLa cells.** Cell extracts of HeLa were processed for soluble examination of **A** EGFP, EGFP-MED15CC (CC), and EGFP-MED15PP (PP) or **B** endogenous MED15 by western blot in untransfected (untransf) cells or after expression of EGFP, EGFP-MED15CC (CC), or EGFP-MED15PP (PP) proteins. EEA1 and GAPDH were blotted as loading controls. **C** Cell extracts of HeLa were processed for nuclear/cytoplasmic (N/C) examination of endogenous MED15, MED14, MED12, and MED1 mediator subunits by western blot after expression of EGFP or EGFP-MED15CC (CC) proteins. LaminA/C, H3K9me3, and GAPDH were blotted as loading controls.

cellular inclusion-forming propensities, EGFP-MED15CC was localized mainly in the detergent-insoluble fraction upon HeLa cells fractionation, whereas EGFP-MED15PP remained mostly soluble (Fig. 7A).

Overall, the cellular data converge to indicate that, as it happens *in vitro*, the CC conformation mediates MED15-PrLD aggregation and the formation of protein inclusions in human cells.

We have shown that, *in vitro*, the MED15-PrLD seeds the amyloid formation of its soluble counterpart efficiently. This activity immediately suggested that, when expressed in cells, it may transmit its amyloid-like state and potentially seed the aggregation of the endogenous, soluble, and functional MED15 protein. Western blot and immunostaining of the endogenous full-length MED15 indicate that EGFP-MED15CC transient expression in HeLa cells results in a fraction of the otherwise soluble endogenous MED15 being recruited into the detergent-insoluble fraction (Fig. 7B). This pro-aggregation effect was reduced when we expressed the CC-disrupted EGFP-MED15PP variant instead (Fig. 7B).

Despite MED15 is assumed to be a nuclear protein, its presence in the cytosol has been described<sup>24</sup> and it has been suggested that this subunit might experiment indirect nuclear export/import through a piggyback mechanism<sup>54</sup>. We speculated that sequestration of the endogenous protein by aggregated MED15-PrLD should increase its levels in the cytosol. HeLa cells fractionation into cytoplasmic and nuclear fractions confirmed the expected redistribution of the endogenous protein upon EGFP-MED15CC expression (Fig. 7C). Mislocalization of endogenous MED15 in the cytosol might result from the retention of newly synthesized protein and/or by an alteration of the nucleo-cytoplasmic shuttling. In both cases, the interaction with aggregated cytosolic EGFP-MED15CC will difficult the transit to the nucleus.

Next, we assessed if the mislocalization of endogenous MED15 might also alter the distribution of other subunits of the Mediator complex, particularly MED14, MED12, and MED1, placed at the tail, CDK8, and middle modules, respectively. In contrast to MED15, none of these endogenous subunits mislocalizes upon

expression of EGFP-MED15CC (Fig. 7C), which indicates that as expected for a prion-like protein mechanism, cytoplasmic recruitment is sequence-specific.

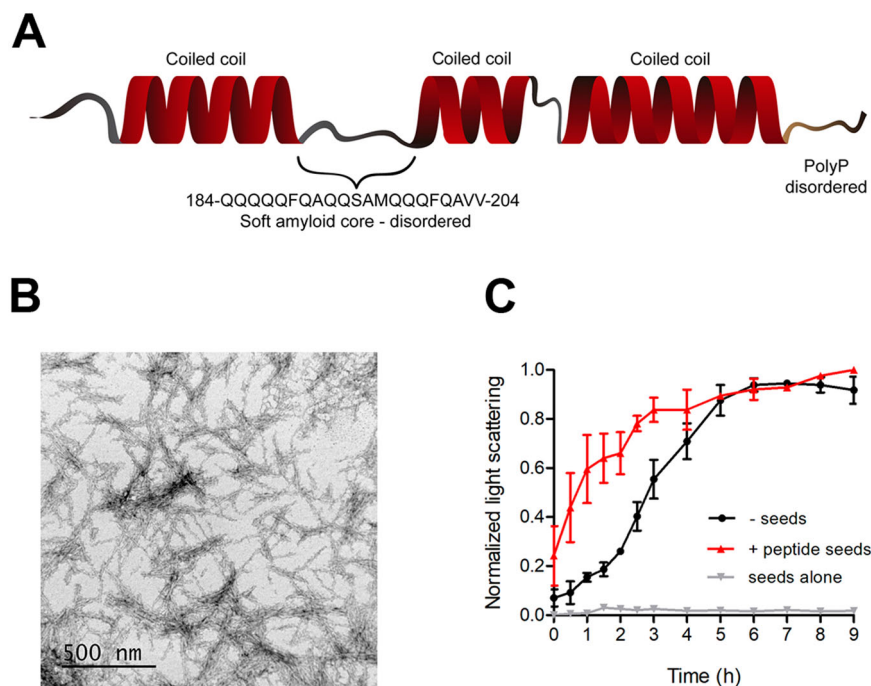
Overall, the data in this section converge to indicate that MED15-PrLD possesses intracellular prion-like activity, related to the CC propensity of its Q-rich segment.

## Discussion

We have identified the presence of CC domains in a significant number of human PrLDs. This result extends the previous observation that these  $\alpha$ -helical domains were recurrent in yeast prions<sup>9</sup>, indicating that CC constitute a generic feature of many functional prion and prion-like proteins.

PrLDs are considered and consistently predicted to be disordered in their soluble states. Accordingly, a disordered to  $\beta$ -sheet transition is seen as the generic mechanism behind their aggregation, similar to pathogenic amyloids<sup>3,4</sup>. Instead, our data are consistent with CC domains playing a critical role in the structural dynamics of at least a fraction of human PrLDs. Accordingly, disrupting or destabilizing the MED15-PrLD CC, either genetically, by Pro mutations, or chemically, with urea, abolishes amyloid formation, whereas its stabilization with TFE accelerates aggregation.

To the best of our knowledge, this is the first time that the transition of a complete PrLD from an initial soluble CC conformation to an amyloid state has been characterized biophysically since previous reports involved relatively short peptides (17 residues) of the yeast Ure2p PrD<sup>9</sup>. These studies did not allow to discriminate if  $\alpha$ -helical CC were self-sufficient mediators of the aggregation of functional prions or just intermediates or facilitators in the process of  $\beta$ -sheet formation. Our results support this second mechanism because the spontaneous aggregation of MED15-PrLD into amyloid fibrils involves a progressive conformational shift from an initial  $\alpha$ -helical state to a  $\beta$ -sheet-rich structure, which occurs concomitantly with the aggregation of the domain.



**Fig. 8 MED15 soft amyloid core seeds MED15-PrLD aggregation.** **A** Schematic diagram of MED15-PrLD structure. It consists of three coiled-coil regions predicted by COILS<sup>33</sup> (14-residue window) separated by disordered regions. The disorder region between the first and second coiled-coil corresponds to the soft amyloid core (SAC) predicted by pWaltz<sup>73</sup>. The final disordered region corresponds to the Pro residues flanking the polyQ tracts (polyP disordered). **B** Transmission electron microscopy image of aggregated 100  $\mu$ M MED15CC SAC in 20 mM Tris pH 7.5 and 150 mM NaCl. **C** Light scattering kinetics of 5  $\mu$ M GFP-MED15CC incubated at 37  $^{\circ}$ C in the absence or presence of 2% MED15 SAC seeds (peptide seeds). Kinetics of seeds alone is shown as control. Aggregation was performed in 20 mM Tris pH 7.5 and 150 mM NaCl.

The molecular mechanism by which a given CC converts into a  $\beta$ -sheet and forms amyloids is still unclear. In an elegant study, Kammerer et al.<sup>55</sup> designed a short peptide (17 residues), referred to as cc $\beta$ , that folds into a CC under ambient conditions but transforms into amyloid fibrils at elevated temperatures. This allowed the authors to trigger the CC to  $\beta$ -sheet transition in a controlled manner and to dissect the relative importance of the forces driving this conformational change. The study revealed that the nature of residues at position “f” of the heptad repeat, the most exposed region in a CC, was the most critical feature for the peptide transition. Hydrophobic residues at this position dramatically accelerated aggregation, whereas polar amino acids abolished it. These hydrophobic residues would facilitate multimerization, a reaction that depends on local contacts and would be favored at increased protein concentrations. Besides, a hydrophobic amino acid at position “f” would allow forming a continuous hydrophobic patch in a  $\beta$ -sheet, together with hydrophobic residues in positions “a” and “d”, all lying in one side of the sequence in a fully extended conformation<sup>56</sup>. The same mechanism would likely apply for polyQ tracts, since Gln behaves as an ambivalent hydrophobe<sup>41</sup>, with contacts becoming more favorable as the polyQ, and therefore the CC, expands.

Although our model system is far more complex, it shares significant features with the behavior of cc $\beta$ . In both cases, CD data indicated a transition from an initial  $\alpha$ -helical to a  $\beta$ -sheet structure, the aggregated states corresponded to ordered amyloid fibrils, and conformational conversions were significantly accelerated by increased protein concentrations. Importantly, according to COILS, all residues at position “f” of the MED15-PrLD CC domain are either hydrophobic or Gln (Fig. 1B). The need for specific spatial positioning of exposed hydrophobic residues to allow multimerization might explain the need to maintain the integrity of CC for MED15-PrLD aggregation. An

essential difference between cc $\beta$  and MED15-PrLD is that the structural transition of the former requires native state destabilization by temperature, whereas in the second case, it occurs spontaneously at room temperature. This indicates that the energy barrier between the CC and amyloid states of MED15-PrLD is significantly lower than that of cc $\beta$ , something expected for a domain that should experiment conformational shifts under physiological conditions.

A CC-mediated model does not exclude the interplay of CCs with other PrLD elements in the aggregation process. The predicted CC regions in MED15-PrLD are discontinuous, with two short regions devoid of significant  $\alpha$ -helical propensity linking them (Fig. 1B). We have shown that a majority of yeast PrDs contain sequence stretches able to assemble into highly ordered amyloid fibrils<sup>57</sup>. These regions differ from the classical amyloid cores of pathogenic proteins in that they are slightly longer and more polar, in such a way that the amyloid potential is less concentrated, allowing the containing PrD to remain soluble under most physiological conditions, while still keeping a certain amyloid propensity that might facilitate assembly in certain circumstances<sup>57,58</sup>. These soft amyloid cores (SAC) are necessary to sustain yeast prions conversion in human cells<sup>59</sup>. We recently identified and characterized similar regions within several human PrLDs, including that of MED15<sup>60</sup>. The SAC of MED15-PrLD comprises residues 184-QQQQFQAQQSAMQQQFQAVV-204 and a peptide correspondent to this stretch spontaneously forms highly ordered,  $\beta$ -sheet-rich, amyloid fibrils<sup>60</sup>. This SAC corresponds almost precisely to the connecting region between the first and second CC in MED15-PrLD (Figs. 1B and 8A). This sequence is predicted to be devoid of any CC propensity by COILS and completely disordered by both FoldIndex and Espritz, and accordingly to be significantly exposed to the solvent. Thus, contacts between these accessible regions in different MED15-

PrLD molecules are possible. Indeed, when we formed fibrils of the 21-residues peptide (Fig. 8B) and added them (2%) to a solution of soluble GFP-MED15CC, they efficiently seeded its aggregation, indicating that this region was recognized in the soluble and dimeric protein (Fig. 8C). However, it is important to note that this amyloidogenic sequence alone is not sufficient to promote aggregation since this reaction does not occur in the absence of a previous CC conformation. Thus, either  $\alpha$ -helical regions drive the initial MED15-PrLD multimerization, and intermolecular SAC contacts occur afterward, or both regions act in parallel to mediate aggregation. In both instances, the presence of the SAC would contribute to decrease the energy barrier for amyloid formation, respect that of a CC-only driven reaction. This leads us to propose a mechanism in which both the misfolding of short disordered regions to  $\beta$ -sheet conformations and CC multimerization cooperate to promote MED15-PrLD amyloid formation, unifying thus two apparently contradictory models for prion conversion. Such a cooperative model would explain why MED15-PrLD fibrillates spontaneously, even if its polyQ tracts are much shorter than those in triplet expansion genetic diseases<sup>61</sup>.

We show that a dimeric CC structure is the thermodynamically dominant MED15-PrLD form in solution, immediately suggesting that MED15 might form homodimers spontaneously. No studies on the oligomeric state of full-length MED15 are available, but it has been demonstrated that  $\gamma$ MED15 assembles into a dimer over a broad range of concentrations, suggesting that this state is also accessible to MED15<sup>62</sup>. The existence of functional MED15 as a dimeric protein stabilized by CC contacts is consistent with the observation that proteins containing CC are overrepresented among its interactors, since apart from homotypic interactions, CC also facilitate heterotypic PPI<sup>63</sup>. Human prion-like proteins containing CC at their PrLDs function in transcriptional regulation and comprise important coactivators, including NCOA2, MAML2, CBP, and MED12, all interconnected with MED15. As in humans, transcription regulators are prevailing in the yeast prion family<sup>6</sup>, and they include four *bona fide* prions: Ure2p, Swi1, Cyc8, Mot3. Not surprisingly, Fiumara et al.<sup>9</sup> demonstrated that all these proteins contain CC at their PrDs. Thus, transcription appears to be an important cellular event that is influenced by prions and prion-like proteins across species, likely thanks to their ubiquitous CC domains. As it occurs in yeast, human prion-like transcription coactivators might permit the generation of phenotypic heterogeneity, allowing cells to adapt in front of intrinsic and extrinsic changes.

For most yeast prions, the new phenotype arising from their conversion to the amyloid state arises from a loss of function. In this way, Ure2p is a dimeric protein whose globular functional domain binds to the transcription factor Gln3p, preventing its migration to the nucleus<sup>64</sup>. This interaction is lost in the Ure2p prion amyloid state, allowing Gln3p to activate the transcription of a series of genes that were previously silent. However, the Ure2p globular domain is not misfolded in the process and maintains its original fold but becomes occluded from interaction and loses its original function<sup>51</sup>. Similarly, when MED15-PrLD accesses a highly ordered amyloid state, the contiguous GFP moiety remains folded and fluorescent, suggesting that a mechanism equivalent to that of Ure2p might regulate MED15 activity.

In vitro, preformed fibrils of MED15-PrLD can seed the aggregation of their soluble counterpart, implying that they facilitate a conformational conversion from an initial  $\alpha$ -helical to a  $\beta$ -sheet structure. When MED15-PrLD is expressed in human cells, a significant fraction of the endogenous MED15 protein localizes to the detergent-insoluble cytoplasmic fraction suggesting that an analogous mechanism applies inside cells and thus

that MED15 can act in a prion-like manner in a physiological context. The ability to aggregate inside cells and to potentially transmit this conformation is, again, strongly dependent on the integrity of the N-terminal CC domain, as demonstrated by the reduced ability to form cellular inclusions and to interact with the endogenous protein of the MED15PP variant.

MED15 is part of the mediator tail module and is known to serve as a hub for signaling pathways. MED15 is overexpressed in different cancers and correlates with the clinical outcome and the recurrence of the disease<sup>18–20,23,65</sup>. The molecular mechanism by which increased levels of MED15 contributes to these malign phenotypes remains unknown, but it is assumed that they result from an increased and sustained transcriptional activation<sup>19,20,23</sup>. MED15-PrLD forms intracellular inclusions in different human cell lines, including laryngeal and oral squamous cell carcinoma (Fig. S10), two tumors with MED15 implication<sup>19</sup>. Under the assumption that, as it occurs for  $\gamma$ MED15, the aggregation propensity of the PrLD is maintained in the context of the full-length protein, our results suggest alternative mechanisms to explain the association of this prion-like protein with cancer. Overexpression would increase the number of MED15 pre-associating N-terminal CC domains and thus their effective local concentration, facilitating both the establishment of CC homotypic and heterotypic interactions and PrLD conformational conversion.

MED15-PrLD mediated homotypic interactions would favor multimerization, and this might prevent the establishment of other relevant MED15 PPI, occluding the access to TFs interacting motifs, like those in the KIX domain, as it happens in yeast Ure2p<sup>51</sup>. Alternatively, the self-association of MED15 might result in enhanced formation of transcriptional hubs by phase separation, as described for the ENL chromatin reader<sup>66</sup>. Both situations are expected to result in widespread transcriptional changes. Because a significant fraction of MED15 interactors contains CC domains, increased heterotypic interactions and subsequent multimerization by side-to-side CC interactions, or from the swapping between protomers, are expected to exert a similar effect to homotypic contacts, either decreasing or exacerbating the interactors function. Another possibility is that the MED15-PrLD would establish promiscuous interactions with other Q-rich CC-containing proteins that are not usual MED15 partners, sequestering them. Since many of these proteins are expected to be TFs, these degenerated CC interactions would significantly impact transcription and signaling pathways, as demonstrated for polyQ-expanded proteins in neurodegenerative diseases<sup>11,67</sup>. In extreme cases, the MED15-PrLD might template and enhance the transition of other Q-rich proteins towards the formation of toxic amyloid aggregates, as previously shown for human ATXN1<sup>40</sup>. Deciphering if any of these mechanisms or a combination of them is behind MED15 association with different cancers might offer novel therapeutic avenues for their treatment.

## Methods

**Bioinformatics analysis.** Human reviewed proteins from the standard proteome (Proteome ID UP000005640, release 2019\_06) were downloaded from UniProtKB/Swiss-Prot<sup>68,69</sup>. MED15 mouse interactors were obtained from<sup>42</sup>. Their human homologous were obtained by blasting each sequence against their human UniProtKB/Swiss-Prot counterparts. For each protein, the highest sequential candidate (>75% homology) was obtained.

Human reviewed proteins were screened with PLAAC algorithm<sup>32</sup> using as background probability the precompiled frequencies of human proteome to obtain the list of putative proteins with PrLDs.

Coiled-coil prediction was performed using the C version of COILS<sup>33</sup> or PairCoil<sup>20</sup> with window length of 21 residues, unless is otherwise indicated.

Prion-like proteins and MED15 interactors datasets were analyzed for enrichment with the Functional Annotation Tool of DAVID 6.8 (Database for Annotation, Visualization and Integrated Discovery)<sup>34</sup> and ordered by p-value. The default "Direct GO" category was selected for the three ontologies: molecular function, biological process and cellular component. Direct GO uses mappings

directly provided by the source annotation, thus minimizing redundant parent terms or very high-level classes.

**MED15 sequences.** MED15-PrLD sequence (residues 141–316) was obtained from Eric E. Wanker research group and fused to GFP sequence in pET21b plasmid. GFP-MED15CC sequence is shown below with MED15-PrLD sequence underlined.

MA SMTGGQOMGRDPNSSKGEELFTGVVPIILVELDGDVNGHKF  
SVSGEGEGDATYGLTLKFICTTGKLPVWPVPTLVTTLTYGVCQCF  
SRYPDHMKRHFDFK SAMPEGYVQERTISFKDDGNYKTRAEVKFE  
GDTLVNRIELKGI DF KEDGNILGHKLEYNYSHNVIYITADKQKN  
GIKANFKIRHNI EDGSVQLADHYQQNTP IGDGPVLLPDNHYLST  
QSALS KDPNE KRDHMLLEFVTAAGITHGMDLEYKKLLEVLFOG  
PMAVVSTATPOTQLQLQOVALQQQQQQQFQQQQQAAALQQQQQ  
QQQQQFOAQQSAMQQQFAVVQQQQQLQQQQQQOHLIKLHHON  
QQQQQQQQQLORLAQLQQQQQQQQQQQQQQQQQQQALQAAPPIC  
QPPMQPQPFPSPALPQOOLQOMHHTQHHPQPQPQPVAQNP  
SAAALEHHHHHHH

MED15 PP sequence was purchased from Genscript as GFP-MED15PP fusion in a pET21b plasmid. The sequence is the same as GFP-MED15CC except 5 Pro substitutions in MED15 CC region (Fig. S6A).

**Protein expression and purification.** In all, 100 ml of Luria Broth (LB) with 100 µg/ml ampicillin (amp) and 34 µg/ml chloramphenicol (cm) were inoculated by small colony of BL21 Rosetta cells with pET21b-GFP-MED15CC/PP and incubated overnight at 37 °C and 250 rpm. In total, 25 ml of saturated overnight culture was transferred into 1L LB-amp-cm and incubated at 37 °C and 250 rpm. Protein expression was induced at OD<sub>600</sub> = 0.5 by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The induced culture was incubated O/N at 20 °C and 250 rpm. Cells were then harvested by centrifugation for 15 min at 4000×g and 4 °C (Beckman CoulterTM Avanti Centrifuge J-26XPI). The cell pellet was frosted at –80 °C.

Pellets from 2L cell culture were resuspended in 30 ml binding buffer (20 mM TrisHCl pH 8, 500 mM NaCl, 20 mM imidazole, 10% glycerol) supplemented with 0.2 mg/ml lysozyme, 20 µg/ml DNase, and 1 tablet of protease inhibitor cocktail EDTA free (Roche). The suspension was incubated for 20 min at 4 °C with slow agitation and then lysed by 5 min sonication on ice (Branson Digital Sonifier). Lysate cells were centrifuged for 30 min at 30,000×g and 4 °C. The supernatant was filtered through a 0.45-µm PVDF membrane and loaded onto a HisTrap FF Ni-column (GE Healthcare) at a flow rate of 4 ml/min. Protein was eluted by one-step procedure using elution buffer (20 mM TrisHCl pH 8, 500 mM NaCl, 250 mM imidazole, and 10% glycerol). His purified protein was treated with 0.2 mg/ml RNase (Thermo Scientific) for 15 min at 37 °C and 1 mM PMSF was added to avoid protein degradation. Native buffer (20 mM TrisHCl pH 8, 500 mM NaCl, and 10% glycerol) was added to the protein sample to reduce imidazole to 20 mM and a second His trap step was performed as before. Then, protein was filtered using 0.22 µm PVDF membrane, concentrated using 10 K Amicon (Millipore) to 2.5 ml, filtered again and subjected to PD10 column equilibrated in 20 mM TrisHCl pH 7.5, 500 mM NaCl, 10% glycerol. Purified protein was concentrated to 400 µM, filtered and stored in small volume aliquots at –80 °C. Protein concentration and absence of RNA was confirmed by 260 nm and 280 nm absorbance values using Specord® 200 Plus spectrophotometer (Analytischjena).

GFP was purified as described in Gil-Garcia et al.<sup>71</sup> and Sup35NM was purified as described in Sant'anna et al.<sup>57</sup>.

**Size exclusion chromatography.** In all, 2 mg/ml of purified protein was filtered with 0.22 µm PVDF membrane and loaded on a Superdex 200 HR 10/30 (GE Healthcare) equilibrated with 20 mM TrisHCl pH 7.4, 150 mM NaCl, and 10% glycerol. The fractions were analyzed by 12% SDS-PAGE.

**Native page.** In all, 0.5 mg/ml of purified protein was loaded in a 12.5% blue native polyacrylamide gel electrophoresis for the analysis of protein oligomers.

**Dynamic light scattering.** GFP and GFP-MED15CC protein size was determined using a Malvern Zetasizer Nano Series (Malvern instruments, UK) in 20 mM TrisHCl pH 7.5 and 150 mM NaCl at different protein concentrations per duplicate. Protein diameter was extracted from volume size.

**GFP absorption.** GFP absorbance was monitored from 400 to 600 nm using a Specord® 200 Plus spectrophotometer (Analytischjena).

**Protein aggregation.** Right before each experiment, the stock solutions were diluted to 5 µM in 20 mM TrisHCl pH 7.4 and 150 mM NaCl. For aggregation assays the samples were incubated at 37 °C without agitation. In some experiments, protein aggregation was performed in the presence of 2 M urea or 5% 2,2,2-trifluoroethanol (TFE) (Fluka). All aggregation experiments were performed per triplicate.

**GFP fluorescence.** In all, 50 µl sample was centrifuged for 5 min at max speed and RT. Then, supernatant fraction was used for GFP fluorescence measurements unless indicated in the text. GFP fluorescence was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 485 nm, emission range from 500 to 600 nm, slit widths of 5 nm, 0.5 nm interval, 1 s response, and 1000 nm/min scan rate. The supernatant fraction was also analyzed by 12% SDS-PAGE.

**Synchronous light scattering.** Synchronous light scattering was monitored using a JASCO Spectrofluorometer FP-8200. The conditions of the spectra acquisition were: excitation wavelength of 330 nm, emission range from 320 to 340 nm, slit widths of 5 nm, 0.5 nm interval, 1 s response, and 1000 nm/min scan rate.

**Congo Red precipitation.** CR 200 µM in H<sub>2</sub>O was centrifuged before the experiment to remove any possible precipitate. Buffer, GFP, GFP-MED15CC, or Sup35NM aggregates were mixed 1:1 (v/v) with CR 200 µM. Samples were incubated for 1 h at RT covered from light. Then, we centrifuged the samples for 5 min at max speed and RT. Appearance of a red pellet was indicative of the presence of amyloid fibrils.

**Circular Dichroism.** CD experiments were performed using a JASCO J-715 spectropolarimeter. Measurements of the far-UV CD spectra (260–200 nm) were made by the addition of 200 µl of 5 µM 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 5 scans. The contribution of the buffer was subtracted.

**Fourier transform infrared spectroscopy.** In all, 100 µl aggregated protein was centrifuged and washed one time with H<sub>2</sub>O to remove the presence of salts. The final pellet was resuspended in 10 µl H<sub>2</sub>O. 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 32 independent scans, measured at a spectral resolution of 4 cm<sup>-1</sup> within 1800–1500 cm<sup>-1</sup> 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.

**Transmission electron microscopy.** The morphology of the aggregated proteins was evaluated by negative staining and using a JEOL TEM-1400plus Transmission Electron Microscope, 80 KV. 5 µM aggregated protein solution was diluted to 1 µM final concentration in H<sub>2</sub>O. In total, 5 µl of the diluted solution was placed on carbon-coated copper grids and incubated for 5 min. The grids were then washed and stained with 5 µl of 2% w/v uranyl acetate for 1 min. Then, grids were washed again before analysis.

**Mammalian molecular cloning.** MED15 CC and PP sequence were amplified by PCR from pET21b-GFP-MED15CC/PP and assembled to pEGFP-C3 (Clontech, plasmid #6082-1) HindIII and BamHI digested plasmid using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). EGFP and EGFP-MED15CC were amplified by PCR from pEGFP-C3 and pEGFP-C3-MED15CC, respectively, and assembled to pTetO-FUW-OSKM (Addgene, plasmid #20321) EcoRI digested plasmid using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). The final vectors were transformed into XL1Blue chemically competent *E. coli* cells.

**Mammalian cell culture.** HeLa cells were grown and maintained in MEMα Glutamax medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) at 37 °C and 5% CO<sub>2</sub>. HEK-293T were grown and maintained in DMEM medium supplemented with 10% FBS (Gibco) with antibiotics (penicillin/streptomycin 100 U/ml) on gelatin-coated plates at 37 °C and 5% CO<sub>2</sub>. Laryngeal HNSCC (UT-SCC-42B) and oral HNSCC (UT-SCC-2) were grown and maintained in DMEM medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) at 37 °C and 5% CO<sub>2</sub>. Cells were passaged and plated using 1x TrypLE Express (Thermo Fisher Scientific) or Trypsin-EDTA (0.05% Trypsin 0.53 mM EDTA•4Na) liquid (Invitrogen).

**Production of lentivirus and infection of recipient cells.** Lentiviral supernatants were produced in HEK-293T cells (5 × 10<sup>6</sup> cells per 100 mm diameter dish). Vector transfections were performed using Eugene-6 transfection reagent (Roche) according to the manufacturer's protocol.

For lentiviral production, per dish, 293T cells were transfected with 3 plasmids: (i) the ecotropic lentiviral envelope packaging plasmid pMD2.G (0.3 µg; Addgene, plasmid #12259; containing the VsVg gene); (ii) the lentiviral packaging plasmid pCMV-dR8.91 (3.0 µg); (from: Harvard Medical School, plasmid #516); (iii) plus one of the following lentiviral expression constructs (3.0 µg) expressing either the FUW-M2-rtTA vector (Addgene #20342), or, pTetO-FUW-EGFP-MED15CC or pTetO-FUW-EGFP.

Two days later, viral supernatants (10 ml) were collected serially during the subsequent 48 h, at 12-h intervals, each time adding fresh medium to the cells (10

ml). The recipient 293T cells were seeded the previous day ( $1.5 \times 10^5$  cells per well in a six-well plate) and each well received 1.0 ml of the corresponding lentiviral supernatants. This procedure was repeated every 12 h for 2 days (a total of 3 additions). Recipient cells received both the FUW-M2-rtTA plus either pTetO-FUW-EGFP-MED15CC or pTetO-FUW-EGFP.

In all, 24 h after lentiviral infection was completed; lentiviral expressing cells were used directly. Cell samples were tested for expression of EGFP or EGFP-MED15CC by immunofluorescence 72 h after addition of 1  $\mu$ g/ml doxycycline.

**Mammalian cells transfection.** HeLa, laryngeal HNSCC (UT-SCC-42B) and oral HNSCC (UT-SCC-2) cells were transfected 24 h after seeding, on eight-well glass slides for immunofluorescence or on 10 cm dish for western blot, with 400 ng DNA using Lipofectamine2000 (Invitrogen). After 4 h of transfection, we changed cellular media to fresh one.

**Immunofluorescence.** Cells were fixed with 4% paraformaldehyde (141451, Pan-ReacAppliChem), permeabilized with 0.2% TritonX100 and blocked with 1% bovine serum albumin (BSA) or 5% filtered FCS (Foetal Calf Serum). Primary antibody used was against MED15 (11566-1-AP, Proteintech). For visualization, the appropriate host-specific Alexa Fluor 555 (Thermo Fisher Scientific) secondary antibody was used. Slides were mounted using ProLong Gold antifade reagent with DAPI (Invitrogen). Images were captured using a Leica TCS SP5 confocal microscope (Leica Biosystems) with a  $\times 63$  oil objective.

**Soluble/insoluble fractionation.** Transfected HeLa cells were trypsinized, washed with PBS1x and resuspended in 0.5 ml RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% TritonX100, 2 mM EDTA, 0.5% sodium deoxycholate, and 0.1% SDS) supplemented with protease inhibitors. After 10 min incubation on ice, cells were centrifuged for 10 min at max speed and 4 °C. The supernatant corresponding to the soluble fraction was placed in a new eppendorf tube. The pellet corresponding to the insoluble fraction was washed with RIPA supplemented with protease inhibitors, resuspended in 0.5 ml urea buffer (30 mM Tris pH 8.5, 8 M urea, 2% CHAPS) and incubated for 20 min at RT and agitation. Protein concentration of the soluble fraction was measured by Bradford and 20  $\mu$ g were loaded for western blot analysis. We loaded the same volume for the insoluble fraction. Primary antibody used was against MED15 (11566-1-AP, Proteintech), GFP (G6795, Sigma), GAPDH (AM4300, Invitrogen), and EEA1 (610457, BD Biosciences). The appropriate host-specific HRP conjugated secondary antibody (BioRad) was used. Western blot was revealed using Immobilon® Forte Western HRP substrate (Millipore) in a VersaDoc (BioRad).

**Nuclear/cytoplasmic fractionation.** Nuclear and cytoplasmic fractionation was performed with the NE-PER kit (ThermoFisher #78833). In all, 10  $\mu$ g protein was loaded for western blot analyses. Primary antibody used was against MED15 (11566-1-AP, Proteintech), GFP (ab13970, Abcam), MED1 (51613, Cell signaling), MED14 (ab72141, Abcam), MED12 (A300-774A, Bethyl Laboratoris), H3K9me3 (2118 s, Cell Signaling), and LaminA/C (sc-6215 (N-18) or sc-376248 (E-1), Santa Cruz Biotechnology).

**Statistics and reproducibility.** Statistical analyses were performed using GraphPad Prism v8 or SigmaPlot v10. Unless otherwise indicated data are shown as the mean  $\pm$  standard deviation (SD). In Fig. 6C the data correspond to the mean  $\pm$  SEM and significance in difference between two groups were tested by Student *t* test. Triplicate samples were analyzed per experiment.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The unprocessed immunoblotting images are shown in Supplementary Fig. 11. The source data behind the graphs are provided in Supplementary Data 3. All datasets in the current study are available from the corresponding author on reasonable requests.

Received: 22 September 2020; Accepted: 2 March 2021;

Published online: 26 March 2021

## References

- Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–44 (1982).
- Battle, C., Iglesias, V., Navarro, S. & Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Rev. Proteomics* **14**, 335–350 (2017).
- Wickner, R. B., Edskes, H. K., Shewmaker, F. & Nakayashiki, T. Prions of fungi: Inherited structures and biological roles. *Nat. Rev. Microbiol.* **5**, 611–618 (2007).
- Shorter, J. & Lindquist, S. Prions as adaptive conduits of memory and inheritance. *Nat. Rev. Genet.* **6**, 435–450 (2005).
- Liebman, S. W. & Chernoff, Y. O. Prions in yeast. *Genetics* **191**, 1041–1072 (2012).
- 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).
- Jeon, J. et al. Superb resolution and contrast of transmission electron microscopy images of unstained biological samples on graphene-coated grids. *Biochim. Biophys. Acta* **1830**, 3807–15 (2013).
- Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
- Fiumara, F., Fioriti, L., Kandel, E. R. & Hendrickson, W. A. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143**, 1121–1135 (2010).
- Schaefer, M. H., Wanker, E. E. & Andrade-Navarro, M. A. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res.* **40**, 4273–4287 (2012).
- Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30**, 575–621 (2007).
- Pearce, M. M. P. & Kopito, R. R. Prion-like characteristics of polyglutamine-containing proteins. *Cold Spring Harb. Perspect. Med.* **8**, a024257 (2018).
- Kelleher, R. J., Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61**, 1209–15 (1990).
- Malik, S. & Roeder, R. G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761–772 (2010).
- Verger, A., Monté, D. & Villeret, V. Twenty years of Mediator complex structural studies. *Biochem. Soc. Trans.* **47**, 399–410 (2019).
- Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).
- Nakatsubo, T. et al. Human mediator subunit MED15 promotes transcriptional activation. *Drug Discov. Ther.* **8**, 212–217 (2014).
- Wang, K. et al. Increased mediator complex subunit 15 expression is associated with poor prognosis in hepatocellular carcinoma. *Oncol. Lett.* **15**, 4303–4313 (2018).
- Shaikhbrahim, Z. et al. Clinical and molecular implications of MED15 in head and neck squamous cell carcinoma. *Am. J. Pathol.* **185**, 1114–1122 (2015).
- Shaikhbrahim, Z. et al. MED15, encoding a subunit of the mediator complex, is overexpressed at high frequency in castration-resistant prostate cancer. *Int. J. Cancer* **135**, 19–26 (2014).
- Ovchinnikov, D. A. et al. DNA methylation at the novel CpG sites in the promoter of MED15/PCQAP gene as a biomarker for head and neck cancers. *Biomark. Insights* **9**, 53–60 (2014).
- Brown, J. J. et al. Potential biomarkers for head and neck squamous cell carcinoma. *Laryngoscope* **113**, 393–400 (2003).
- Zhao, M. et al. Mediator MED15 modulates transforming growth factor beta (TGF $\beta$ )/Smad signaling and breast cancer cell metastasis. *J. Mol. Cell Biol.* **5**, 57–60 (2013).
- Klümper, N. et al. Differential expression of Mediator complex subunit MED15 in testicular germ cell tumors. *Diagn. Pathol.* **10**, 1–6 (2015).
- Weiten, R. et al. The Mediator complex subunit MED15, a promoter of tumour progression and metastatic spread in renal cell carcinoma. *Cancer Biomark.* **21**, 839–847 (2018).
- Fassler, J. S. & Winston, F. The *Saccharomyces cerevisiae* SPT13/GAL11 gene has both positive and negative regulatory roles in transcription. *Mol. Cell. Biol.* **9**, 5602–5609 (1989).
- Bourbon, H.-M. et al. A Unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell* **14**, 553–557 (2004).
- Cooper, D. G. & Fassler, J. S. Med15: glutamine-rich mediator subunit with potential for plasticity. *Trends Biochem. Sci.* **44**, 737–751 (2019).
- Zhu, X. et al. Mediator tail subunits can form amyloid-like aggregates in vivo and affect stress response in yeast. *Nucleic Acids Res.* **43**, 7306–7314 (2015).
- Parry, D. A. D., Bruce Fraser, R. D. & Squire, J. M. Fifty years of coiled-coils and  $\alpha$ -helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258–269 (2008).
- Mason, J. M. & Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **5**, 170–176 (2004).
- 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).
- Lupas, Andrei, Van Dyke, Marc & Stock, Jeff Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

35. Gemayel, R. et al. Variable glutamine-rich repeats modulate transcription factor activity. *Mol. Cell* **59**, 615–627 (2015).
36. Jiang, H., Nucifora, F. C., Ross, C. A. & DeFranco, D. B. Cell death triggered by polyglutamine-expanded huntingtin in a neuronal cell line is associated with degradation of CREB-binding protein. *Hum. Mol. Genet.* **12**, 1–12 (2003).
37. Prilusky, J. et al. FoldIndex@: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438 (2005).
38. Walsh, I., Martin, A. J. M., Di Domenico, T. & Tosatto, S. C. E. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **28**, 503–509 (2012).
39. Peng, Z., & Kurgan, L. & Kurgan, L. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. *Nucleic Acids Res.* **43**, e121–e121 (2015).
40. Petrakis, S. et al. Identification of human proteins that modify misfolding and proteotoxicity of pathogenic ataxin-1. *PLoS Genet.* **8**, e1002897 (2012).
41. Sodek, J., Hodges, R. S., Smillie, L. B. & Jurasek, L. Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure. *Proc. Natl Acad. Sci. USA* **69**, 3800–4 (1972).
42. Quevedo, M. et al. Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nat. Commun.* **10**, 2669 (2019).
43. Szklarczyk, D. et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
44. Bhattacharyya, A. et al. Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* **355**, 524–535 (2006).
45. Adegbuyiro, A., Sedighi, F., Pilkington, A. W., Groover, S. & Legleiter, J. Proteins containing expanded polyglutamine tracts and neurodegenerative disease. *Biochemistry* **56**, 1199–1217 (2017).
46. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
47. Walavalkar, N. M., Gordon, N. & Williams, D. C. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine Binding domain 2 (MBD2) homologues and gata zinc finger domain containing 2A (GATAD2A/p66a). *J. Biol. Chem.* **288**, 3419–3427 (2013).
48. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474**, 1417–1438 (2017).
49. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta* **1804**, 1405–1412 (2010).
50. Li, H., Rahimi, F., Sinha, S., Maiti, P. & Bitan, G. *Encyclopedia of Analytical Chemistry*. 1–32 (John Wiley & Sons, Ltd, 2009).
51. Baxa, U., Speransky, V., Steven, A. C. & Wickner, R. B. Mechanism of inactivation on prion conversion of the *Saccharomyces cerevisiae* Ure2 protein. *Proc. Natl Acad. Sci. USA* **99**, 5253–5260 (2002).
52. Wickner, R. B. et al. Yeast prions act as genes composed of self-propagating protein amyloids. *Adv. Protein Chem.* **57**, 313–334 (2001).
53. Chang, D. K., Cheng, S. F., Trivedi, V. D. & Lin, K. L. Proline affects oligomerization of a coiled coil by inducing a kink in a long helix. *J. Struct. Biol.* **128**, 270–279 (1999).
54. Tessier, T. M., MacNeil, K. M. & Mymryk, J. S. Piggybacking on classical import and other non-classical mechanisms of nuclear import appear highly prevalent within the human proteome. *Biology* **9**, 188 (2020).
55. Kammerer, R. A. et al. Exploring amyloid formation by a de novo design. *Proc. Natl Acad. Sci. USA* **101**, 4435–40 (2004).
56. Dong, H. & Hartgerink, J. D. Role of hydrophobic clusters in the stability of alpha-helical coiled coils and their conversion to amyloid-like beta-sheets. *Biomacromolecules* **8**, 617–23 (2007).
57. Sant’anna, R. et al. Characterization of amyloid cores in prion domains. *Sci. Rep.* **6**, 34274 (2016).
58. Fernández, M. R., Batlle, C., Gil-García, M. & Ventura, S. Amyloid cores in prion domains: key regulators for prion conformational conversion. *Prion* **11**, 31–39 (2017).
59. Duernberger, Y. et al. Prion replication in the mammalian cytosol: functional regions within a prion domain driving induction, propagation, and inheritance. *Mol. Cell. Biol.* **38**, e00111–18 (2018).
60. Batlle, C., De Groot, N. S., Iglesias, V., Navarro, S. & Ventura, S. Characterization of soft amyloid cores in human prion-like proteins. *Sci. Rep.* **7**, 1–16 (2017).
61. Bañez-Coronel, M. et al. RAN translation in huntingtin disease. *Neuron* **88**, 667–677 (2015).
62. Wands, A. M. et al. Transient-state kinetic analysis of transcriptional activator-DNA complexes interacting with a key coactivator. *J. Biol. Chem.* **286**, 16238–45 (2011).
63. Parry, D. A. D., Fraser, R. D. B. & Squire, J. M. Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258–69 (2008).
64. Thual, C. et al. Stability, folding, dimerization, and assembly properties of the yeast prion Ure2p. *Biochemistry* **40**, 1764–1773 (2001).
65. Syring, I. et al. The knockdown of the Mediator complex subunit MED15 restrains urothelial bladder cancer cells’ malignancy. *Oncol. Lett.* **16**, 3013–3021 (2018).
66. Wan, L. et al. Impaired cell fate through gain-of-function mutations in a chromatin reader. *Nature* **577**, 121–126 (2020).
67. Kwon, M. J. et al. Coiled-coil structure-dependent interactions between polyQ proteins and Foxo lead to dendrite pathology and behavioral defects. *Proc. Natl Acad. Sci. USA* **115**, E10748–E10757 (2018).
68. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
69. Bateman, A. et al. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
70. McDonnell, A. V., Jiang, T., Keating, A. E. & Berger, B. Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* **22**, 356–8 (2006).
71. Gil-Garcia, M. et al. Combining structural aggregation propensity and stability predictions to redesign protein solubility. *Mol. Pharm.* **15**, 3846–3859 (2018).
72. Toombs, J. A. et al. De novo design of synthetic prion domains. *Proc. Natl Acad. Sci. USA* **109**, 6519–6524 (2012).
73. Sabate, R., Rousseau, F., Schymkowitz, J. & Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Comput. Biol.* **11**, e1004013 (2015).

## Acknowledgements

We are grateful to UAB microscopy service of Barcelona for their technical advice. We thank Erich E. Wanker research group for providing us GST-MED15SCC plasmid. We thank Marcos Gil for providing us GFP purified protein. I.C. was funded by Secretaria d’Universitats i Recerca de la Generalitat de Catalunya and European Social Fund. Work in the laboratory of M.S. was funded by the IRB, and by the Spanish Ministry of Science co-funded by European Regional Development Fund (ERDF; SAF2017-82613-R), the European Research Council (ERC-2014-AdG/669622), “la Caixa” Foundation, and Secretaria d’Universitats i Recerca del Departament d’Empresa i Coneixement of Catalonia (Grup de Recerca consolidat 2017 SGR 282). S.V. acknowledges funding from MINECO (BIO2016-78310-R) and ICREA (ICREA-Academia 2016). C.B. acknowledges funding from “Ministerio de Educación y Formación Profesional”.

## Author contributions

C.B. designed and performed the experiments, analyzed the data, and wrote the manuscript. I.C., C.L., and M.G.-G. contributed to specific experiments, V.I. performed the bioinformatics analysis. M.S. designed the stable cell line experiments. V.I., I.C., C.L., and M.S. edited the manuscript. S.V. designed the project, analyzed the data, and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-01930-8>.

**Correspondence** and requests for materials should be addressed to S.V.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021