



## OPEN ACCESS

## EDITED BY

Paul Martin Harrison,  
McGill University, Canada

## REVIEWED BY

Sean Cascarina,  
Colorado State University,  
United States  
Prashant Jethva,  
Washington University in St. Louis,  
United States

## \*CORRESPONDENCE

Salvador Ventura  
✉ salvador.ventura@uab.es

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Plant Proteomics and Protein  
Structural Biology,  
a section of the journal  
Frontiers in Plant Science

RECEIVED 03 October 2022

ACCEPTED 19 December 2022

PUBLISHED 16 January 2023

## CITATION

Pintado-Grima C, Santos J, Iglesias V,  
Mangano-Artuñedo Z, Pallarès I and  
Ventura S (2023) Exploring cryptic  
amyloidogenic regions in prion-like  
proteins from plants.  
*Front. Plant Sci.* 13:1060410.  
doi: 10.3389/fpls.2022.1060410

## COPYRIGHT

© 2023 Pintado-Grima, Santos, Iglesias,  
Mangano-Artuñedo, Pallarès and  
Ventura. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Exploring cryptic amyloidogenic regions in prion-like proteins from plants

Carlos Pintado-Grima<sup>1</sup>, Jaime Santos<sup>1†</sup>, Valentín Iglesias<sup>1,2,3†</sup>,  
Zoe Mangano-Artuñedo<sup>1</sup>, Irantzu Pallarès<sup>1</sup>  
and Salvador Ventura<sup>1\*</sup>

<sup>1</sup>Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain, <sup>2</sup>Barcelona Institute for Global Health, Barcelona Centre for International Health Research (ISGlobal, Hospital Clínic-Universitat de Barcelona), Barcelona, Spain, <sup>3</sup>Nanomalaria Group, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Barcelona, Spain

Prion-like domains (PrLDs) are intrinsically disordered regions (IDRs) of low sequence complexity with a similar composition to yeast prion domains. PrLDs-containing proteins have been involved in different organisms' regulatory processes. Regions of moderate amyloid propensity within IDRs have been shown to assemble autonomously into amyloid fibrils. These sequences tend to be rich in polar amino acids and often escape from the detection of classical bioinformatics screenings that look for highly aggregation-prone hydrophobic sequence stretches. We defined them as cryptic amyloidogenic regions (CARs) and recently developed an integrated database that collects thousands of predicted CARs in IDRs. CARs seem to be evolutionary conserved among disordered regions because of their potential to establish functional contacts with other biomolecules. Here we have focused on identifying and characterizing CARs in prion-like proteins (pCARs) from plants, a lineage that has been poorly studied in comparison with other prionomes. We confirmed the intrinsic amyloid potential for a selected pCAR from *Arabidopsis thaliana* and explored functional enrichments and compositional bias of pCARs in plant prion-like proteins.

## KEYWORDS

cryptic amyloidogenic regions, prion-like domains, plants, functional interactions, bioinformatics

**Abbreviations:** CAR, Cryptic amyloidogenic region; pCAR, Prionic cryptic amyloidogenic region; CARs-DB, Cryptic amyloidogenic regions database; PrLDs, Prion-like domains; LCD, Low complexity domain; IDPs, Intrinsically disordered proteins; IDRs, Intrinsically disordered regions; PPIs, Protein-protein interactions; LIPs, Linear interacting motifs; LLPS, Liquid-liquid phase separation; MLOs, Membraneless organelles; MED9, Mediator of RNA polymerase II transcription subunit 9.

# 1 Introduction

Since the description of the first self-perpetuating proteinaceous infectious pathogens in the late 20<sup>th</sup> century (Prusiner, 1982) and the subsequent discovery of yeast prions (Wickner, 1994), the prion concept has progressively evolved with our understanding of the phenomenon. Protein domains in a variety of polypeptides were found to share similar sequential composition to the domains required for yeast prion conversion, although they do not necessarily show conformational transitions or mechanisms of self-propagation and were termed prion-like domains (PrLDs). Despite initially thought to be linked to disease, further characterization of prion-like polypeptides revealed their role in regulating key biological functions in a wide variety of organisms (True and Lindquist, 2000; Chakrabortee et al., 2016; Yuan and Hochschild, 2017; Pallares et al., 2018; Tetz and Tetz, 2018). In this context, under physiological conditions, most PrLDs-containing proteins would not transition towards aggregated states, and instead, they tend to be involved in functional interactions with other proteins and nucleic acids (Alberti et al., 2009; King et al., 2012; Iglesias et al., 2015).

PrLDs are often enriched in glutamine, asparagine, serine, glycine, and tyrosine (Michelitsch and Weissman, 2000; Alberti et al., 2009; Toombs et al., 2010; Diaz-Caballero et al., 2018) and depleted of hydrophobic amino acids, thus having an overall polar character. The characteristic compositional bias of PrLDs is at the core of diverse predictive bioinformatics tools that aim at identifying such signatures in protein sequences (Toombs et al., 2012; Lancaster et al., 2014; Afsar Minhas et al., 2017). Most of the current protein aggregation predictors were designed to predict highly hydrophobic sequence stretches in disease-associated proteins, a feature that does not match the polar nature of PrLDs (Fernandez-Escamilla et al., 2004; Conchillo-Sole et al., 2007; Garbuzynskiy et al., 2010; Maurer-Stroh et al., 2010; Sormanni et al., 2015). Only a few of them, like pWaltz (Sabate et al., 2015b), and its implementation in PrionW (Zambrano et al., 2015), use adapted aggregation scales aimed to identify sequences of milder amyloid propensity, denominated soft amyloid cores (SACs). Albeit pWaltz has been successfully used to identify SACs within PrLDs whose correspondent peptides have been experimentally validated to form amyloids (Sant'anna et al., 2016; Battle et al., 2017a), this algorithm suffers from technical limitations, including the exclusion of proline residues on its predictions, the identification of a single hit per PrLD or the rigidity of its sequence scoring system that unavoidably employs a sliding window of 21 residues.

In recent work, we implemented a pWaltz-related approach to identify protein regions with amyloid potential in hydrophilic sequential contexts by decreasing the actual detection threshold of the Waltz algorithm (Maurer-Stroh et al., 2010), -a well-characterized amyloidogenic predictor- without any further

arbitrary residue exclusion or window-size length assumption. This prediction scheme allowed us to identify and characterize cryptic amyloidogenic regions (CARs) in experimentally validated intrinsically disordered regions (IDRs), which were previously considered to be depleted of aggregation-prone regions (Santos et al., 2021). CARs are sequences of mild amyloidogenic potential, yet able to assemble into fibrils when in isolation. These stretches have likely been conserved within IDPs because their potential to act as functional interacting motifs with other biomolecules overpasses a low risk for pathogenic aggregation. In fact, CARs consistently overlap with interacting motifs in IDPs. We collected these predictions in a comprehensive database named CARs-DB (Pintado-Grima et al., 2022a).

Given the generic disordered nature of PrLDs, in this work, we conducted a computational screening to identify CARs in plant PrLDs (pCARs) from different plant species to understand their role in this kingdom. We discovered that CARs are widespread in model plants' PrLDs, with gene ontology enrichment analysis suggesting they are connected to critical biological processes. In addition, we validated amyloid formation for a selected pCAR from *Arabidopsis thaliana* (*A. thaliana*), demonstrating that this class of plant protein sequences bears the potential to spontaneously self-assemble, at least *in vitro*.

## 2 Materials and methods

### 2.1 Data collection and dataset generation

The reference proteomes of five different plant model organisms, including *Arabidopsis thaliana* (mouse-ear cress; UP000006548), *Zea mays* (maize; UP000007305), *Oriza sativa* (rice; UP000059680), *Solanum lycopersicum* (tomato; UP000004994) and *Nicotiana tabacum* (tobacco plant; UP000084051) were extracted from UniProt (UniProt, 2021) (Release 2022\_03). PrLDs were screened with PLAAC (Lancaster et al., 2014) using a core length of 60 amino acids and relative weighting of background probabilities obtained from input sequences. PLAAC searches protein sequences to identify probable prion subsequences exploiting a hidden-Markov model (HMM) algorithm using the selected background probabilities. pCARs longer than six residues were obtained from PrLDs by applying the Waltz algorithm at thresholds 73.5, 80, and 85, as optimized in our original study of CARs (Santos et al., 2021) and CARs-DB (Pintado-Grima et al., 2022a) to account for polar amyloidogenic regions. The Waltz algorithm (Maurer-Stroh et al., 2010) explores the sequence determinants of amyloid structure using position-specific scoring matrices derived from analyzing a database of

short amyloid-forming and non-forming sequences. Modified Waltz thresholds optimized in our original CAR work were selected based on previous experimental evidence (Sabate et al., 2015b).

## 2.2 Gene Ontology (GO) enrichment analysis

Functional enrichments were assessed with the database for annotation, visualization, and integrated discovery DAVID (Sherman et al., 2022) using each proteome as background. GO terms of biological process (BP\_DIRECT), molecular function (MF\_DIRECT), and cellular component (CC\_DIRECT) were selected for the study with a modified Fisher cutoff p-value of 0.1. The Benjamini false discovery rate method was used to control increased error rates for multiple tests (Benjamini and Hochberg, 1995).

## 2.3 Peptide preparation and aggregation

The peptide encoding the mediator of RNA polymerase II transcription subunit 9 (MED9) pCAR region Ac-QYQQFQQQHFIQQQFQ-NH<sub>2</sub> was purchased from SynPeptide (Shanghai, China) with a purity >95%. To mimic the protein environment, terminal charges were neutralized by N-terminal acetylation and C-terminal amidation. Peptide powder was dissolved in hexafluoroisopropanol to a concentration of 1 mg/mL, aliquoted, and vacuum dried with a SpeedVac (Thermo Fisher Scientific, Waltham, USA). For the aggregation reactions, aliquots were solubilized with 20  $\mu$ L hexafluoroisopropanol and diluted to a concentration of 50  $\mu$ M in 20 mM Tris and 100 mM NaCl pH 8. 150  $\mu$ L of peptide solutions were incubated in a 96 wells plate (non-treated) (Sarstedt, Germany) for two days at 37°C with continuous agitation at 100 RPM. A detailed description of the aggregation conditions is described in the supplementary MIRRAGGE spreadsheet (Martins et al., 2020).

## 2.4 Binding to amyloid dyes

End-point samples of the MED9 pCAR were incubated with 40  $\mu$ M thioflavin-T (Th-T) for 20 minutes, and their fluorescence emission spectra were recorded using a Spark plate reader (Tecan, Männedorf, Switzerland). Fluorescence intensity was measured by exciting at 440 nm, bandwidth of 5 nm, and collecting the emission from 460 to 600 nm with a 1 nm interval.

Congo-red (CR) binding to end-point samples recovered from the plate was tested using a Cary 100 UV/Vis Spectrophotometer (Varian, Palo Alto, United States). 100  $\mu$ L

of the aggregated sample was incubated for 20 minutes with 900  $\mu$ L of CR at a final concentration of 5  $\mu$ M before recording the spectra. CR absorbance spectra were recorded from 400 to 700 nm. A control sample with 100  $\mu$ L of the aggregation buffer was prepared (CR-Free).

## 2.5 Far circular dichroism spectroscopy

Far-UV circular dichroism (CD) spectra of the peptide before and after aggregation were recorded on a Jasco J-815CD spectrometer (Halifax, Canada) at 25°C. The spectra were acquired from 260 nm to 200 nm at 0.2 nm intervals, 2 nm bandwidth, 2 s of response time, and a 200 nm/min scan speed on a 0.1 cm quartz cell. To prevent unwanted self-association through the experiment and have a representative spectrum of the initial time-point, the initial aliquot was resuspended in 25 mM Sodium acetate 100 mM NaCl pH 5; 10 accumulations were recorded and averaged for each measurement.

## 2.6 Attenuated total reflectance Fourier transform infrared spectroscopy

Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy experiments were performed in a Bruker Tensor 27 FTIR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Samples were dried under an N<sub>2</sub> (g) stream and measured at a spectral resolution of 2 cm<sup>-1</sup> within the 1800–1500 cm<sup>-1</sup> range (16 accumulations). Data recording and normalization were performed using the OPUS MIR Tensor 27 software and deconvoluted with the Peak Fit 4.12 program (Systat Software Inc., San Jose, CA, USA).

## 2.7 Transmission electron microscopy

5  $\mu$ L of the aggregated sample, previously diluted to a concentration of 5  $\mu$ M, was placed onto glow-discharged carbon-coated copper grids for 1 min. Sample excess was blotted with ashless filter paper, and grids were washed in distillate water drops. Negative staining was performed with 2% (w/v) uranyl acetate for 1 min. A TEM JEM-1400 (JEOL, Peabody, USA) microscope was used, operating at an accelerating voltage of 120 kV.

# 3 Results

## 3.1 Prevalence of pCARs in plant PrLDs

Since the original characterization of prion domains in yeast (Chernoff et al., 1995; Serio et al., 1999), PrLDs with similar

features have been identified and studied in other organisms such as viruses (Tetz and Tetz, 2018), bacteria (Iglesias et al., 2015; Yuan and Hochschild, 2017; Choi et al., 2022), protozoan (Pallares et al., 2018) or humans (An and Harrison, 2016; Iglesias et al., 2019). Plants have received less attention, with only a few studies focusing on PrLDs' coverage and function (Chakrabortee et al., 2016; Dorone et al., 2021; Garai et al., 2021). To identify the potential presence of CARs in PrLDs (pCARs) from plants, the proteomes of five different model higher plants were first screened with the PLAAC algorithm. Around 1% of the total proteome of these species was predicted as PrLDs, spanning from 0.62% (*O. sativa*) to 1.57% (*A. thaliana*). In comparison with other studied model organisms (Gil-Garcia et al., 2021), this average is similar to the proportion of human, bacteria, and viruses PrLDs and lower than *S. cerevisiae* (~3%) (Figure 1; Table 1).

For predicted plant PrLDs, pCARs were identified with Waltz at the three designated thresholds 73.5, 80, and 85. As expected, the percentage of PrLDs CAR positive (containing at least one CAR) increased in all species as the threshold became less stringent, hence allowing the detection of an increasing number of pCARs of polar nature (Figure 2A). Remarkably, at threshold 73.5, more than 90% of the predicted PrLDs contained

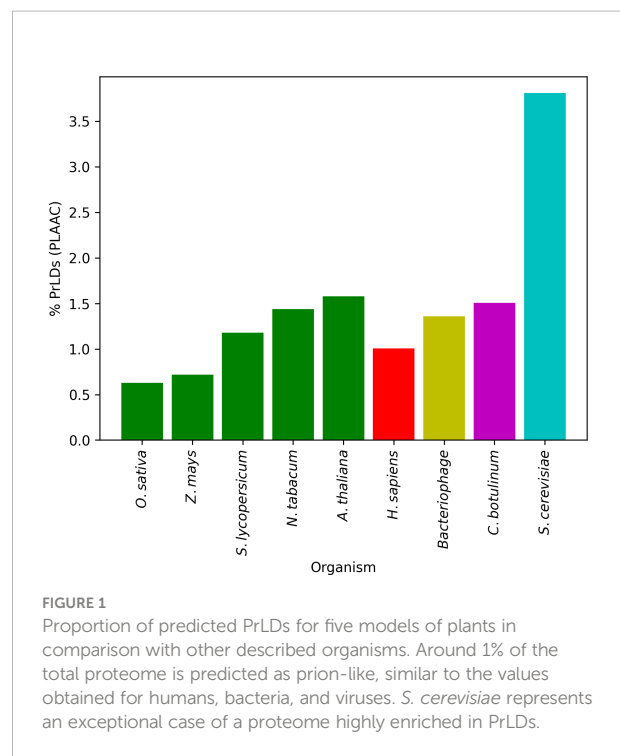


TABLE 1 Identification of pCARs from PrLDs for the five different plant model organisms.

Organism	Proteins	PrLDs	PrLDs (%)	Waltz threshold	CARs	PrLDs CAR+	PrLDs CAR + (%)	CARs/PrLD CAR+	Average CAR length
<i>Arabidopsis thaliana</i>	27474	432	1.57	85	679	303	70.14	2.24	11.6
				80	983	356	82.41	2.76	12.66
				73.5	1484	402	93.06	3.69	14.69
<i>Zea mays</i> (maize)	56926	404	0.71	85	667	302	74.75	2.21	12.58
				80	918	356	88.12	2.58	13.78
				73.5	1358	380	94.06	3.57	15.99
<i>Oryza sativa</i> (rice)	43672	270	0.62	85	429	190	70.37	2.26	11.48
				80	595	220	81.48	2.7	13.19
				73.5	864	253	93.7	3.42	15.52
<i>Solanum lycopersicum</i> (tomato)	34655	405	1.17	85	689	307	75.8	2.24	13
				80	957	354	87.41	2.7	14.2
				73.5	1366	377	93.09	3.62	15.88
<i>Nicotiana tabacum</i> (tobacco plant)	61673	883	1.43	85	1471	683	77.35	2.15	12.38
				80	1989	763	86.41	2.61	13.98
				73.5	2913	847	95.92	3.44	15.96

PrLDs were screened with PLAAC, and their corresponding CARs were predicted with Waltz at the three designated thresholds.

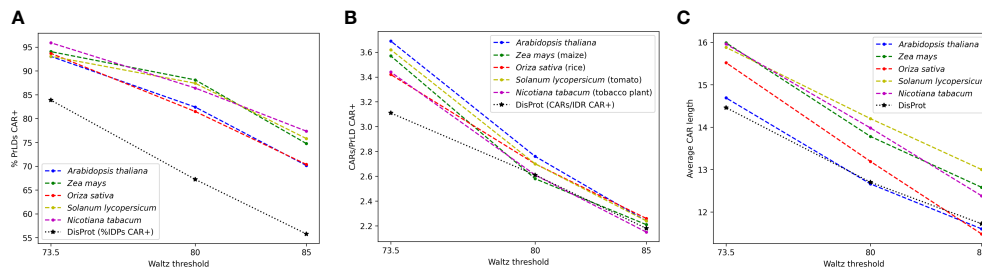


FIGURE 2

General statistics for pCARs from *A. thaliana* (blue), *Z. mays* (green), *O. sativa* (red), *S. lycopersicum* (ochre), and *N. tabacum* (purple), in comparison with general DisProt CARs (black). Most PrLDs are CAR-positive (A), often with more than one predicted pCARs (B). Lower Waltz thresholds contain more pCARs with longer average sequences (C).

such regions. According to the analysis, for the 5 tested plant species, PrLDs tend to display several pCARs, with an average length of >10 residues (Figures 2B, C). Comparing these metrics with generic intrinsically disordered regions (IDRs) extracted from DisProt (Santos et al., 2021), the positive rate of CARs in PrLDs is higher than in IDRs, as well as the number of CARs/LCD (low complexity domain, either PrLD or IDR) and their size. This indicates that CARs are more prevalent in PrLDs than in IDRs, likely because of the particular amino acid bias of prion LCD sequences.

As in IDRs (Santos et al., 2021), the widespread representation of CARs in plants is suggestive of a link with protein functionality. We compared the coincidence of pCARs from *A. thaliana* with linear interaction motifs (LIPs) as described in the MobiDB database (Piovesan et al., 2021), and observed a 63.5%, 65.6%, and 67.6% of pCAR residues overlapping with LIPs for the 73.5, 80 and 85 thresholds, respectively. This suggests that these regions might be conserved because of their contribution to establishing

functional interactions. Our Aggrescan algorithm (Conchillo-Sole et al., 2007) failed to identify 71.5% (703/983) of *A. thaliana*'s pCARs detected at the 80.0 threshold, indicating that, in the context of PrLDs, they do not confer to the containing protein a high risk for aggregation.

### 3.2 Amino acid composition of pCARs in *A. thaliana*

Once the presence of pCARs in all five plant species was confirmed and provided that the proportion of PrLDs and pCARs were similar, we selected the best-known plant model organism *A. thaliana* for subsequent analyses.

We compared the amino acid composition of PrLDs from *A. thaliana* with that of the complete proteome (Figure 3A). As expected from the encoded residue preferences in the employed PLAAC algorithm, in the PrLDs from *A. thaliana* are depleted of hydrophobic residues and enriched in  $\beta$ -breakers (GP), two

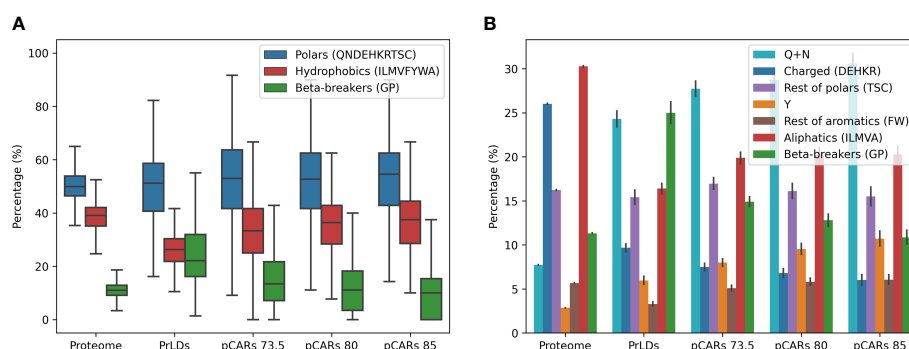


FIGURE 3

pCARs composition by threshold in *A. thaliana* compared with their PrLDs and its proteome. Compared with the proteome, PrLDs already have an intrinsic bias toward polar and  $\beta$ -breakers amino acids. In contrast, pCARs are enriched in hydrophobics and depleted of  $\beta$ -breakers (A). Most polar amino acids are QNs, and aromatic residues seem to modulate the amyloidogenic load of pCARs (B). Barplot draws error bars in the plot with 95% confidence interval.

features that play against the uncontrolled and generic aggregation of these disordered regions (Rousseau et al., 2006; Theillet et al., 2013). Indeed, glycines and prolines (G+P) alone equal the total content of hydrophobic residues, around 25%, whereas, in the total proteome, non-polar residues are threefold more abundant than G+P. Next, we compared pCARs composition against PrLDs. pCARs exhibit a higher percentage of hydrophobic residues and a lower content in  $\beta$ -breakers than PrLDs. As expected, the proportion of G+P in pCARs decreases as the Waltz threshold becomes more stringent. The percentage of hydrophilic residues in pCARs resembles that in PrLDs, being insensitive to the applied threshold, suggesting that they might provide background control to the equilibrium between disorder and assembly in these sequences.

To better dissect the relevant physicochemical features hidden in these amino acid groups, we split them into seven categories: Q+N, charged, rest of polars, tyrosine, rest of aromatics, aliphatics, and  $\beta$ -breakers. As observed in Figure 3B, only two residues (Q+N) account for almost 25% of the amino acids in PrLDs, a situation that contrasts with these residues' low contribution to this plant proteome composition (7.9%). Asp and Gln are known to play a critical role in prion conformation (Perutz et al., 2002; Derkatch et al., 2004; Jiang et al., 2004; Sabate et al., 2015b; Zambrano et al., 2015; Zhang et al., 2016). Interestingly enough, Q+N are more frequent in pCARs than in complete PrLDs, their number increasing with the stringency of the prediction, indicative that they would be important contributors to the homo- or heterotypic interactions these stretches might establish. The higher proportion of hydrophobic amino acids in pCARs, relative to its PrLDs containers, can be traced to increases in both aromatic and aliphatic residues. However, it seems that the presence of aromatic residues is a distinctive feature of pCARs, their proportion increasing with the detection threshold. Specifically, tyrosines (Y) account in all the cases for more than 50% of the total aromatics, likely due to its dual polar/apolar character and its unique hydrogen bonding capability. This bias toward Y is reminiscent of what has been observed in liquid-liquid phase separation (LLPS), where phenylalanine and tyrosine are not interchangeable (Wang et al., 2018), likely because Y best balances solubility and assembly potential (Diaz-Caballero et al., 2018).

Not surprisingly, regions between pCARs (interpCARs) also differ in residues composition from pCARs (Figure 4). Compared with pCARs, interpCARs are highly enriched in G+P and depleted in aromatic residues. Despite Q+N are also abundant in interpCARs there is a strong preference for Q over N in these regions, indicating that their properties are not entirely exchangeable. Another interesting observation is that the cationic arginine (R) and histidine (H) residues are preferred at interpCARs, but this bias does not apply to other charged residues like lysine, glutamate, and aspartate. The average length

of interpCARs is longer than that of pCARs; 20.4 and 12.7 at threshold 80.0, respectively.

In a way, the distribution and biased composition of pCARs within PrLDs is reminiscent of the proposed stickers and spacers model of LLPS (Wang et al., 2018; Martin et al., 2020; Holehouse et al., 2021; Bremer et al., 2022); with the Y+F+W+N enriched pCARs acting as stickers that facilitate protein interactions and the P+G+H+R enriched interpCARs as spacers, connecting stickers and providing them the adequate solubility and disorder context.

### 3.3 Functional enrichments of pCARs in *A. thaliana*

To explore the potential contribution of pCARs to the biological roles of *A. thaliana* prion-like proteins, we performed a functional enrichment analysis using the DAVID database (Sherman et al., 2022).

Regarding the molecular functions of pCARs, significant enrichments were found for regulative processes, especially for nucleic acids, including DNA, RNA and mRNA binding, but also protein binding (Figure 5; Supplementary Table 1). This is interesting, as CARs have already been described to mediate key PPIs in IDPs (Santos et al., 2021; Pintado-Grima et al., 2022a). It is not surprising that, provided the previously proposed regulatory function of CARs, pCARs are found enriched in biological activities of gene expression, transcription, and translation. Related enrichments have been described in amyloidogenic (Antonets and Nizhnikov, 2017) and prion-like

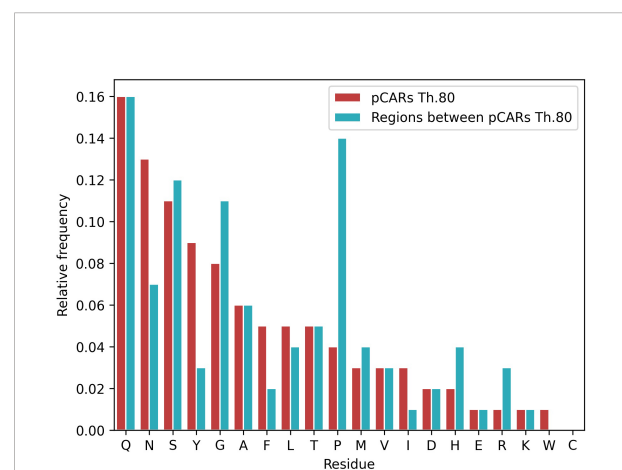


FIGURE 4  
Residue composition comparison between *A. thaliana* pCARs and interpCARs (regions between two pCARs in any PrLD sequence) at threshold 80. pCARs are enriched in aromatic residues and depleted of  $\beta$ -breakers (GP). An opposite tendency is observed for interpCARs, which seem to compensate the sequential amyloidogenic load by reducing the aromatics and increasing  $\beta$ -breakers.

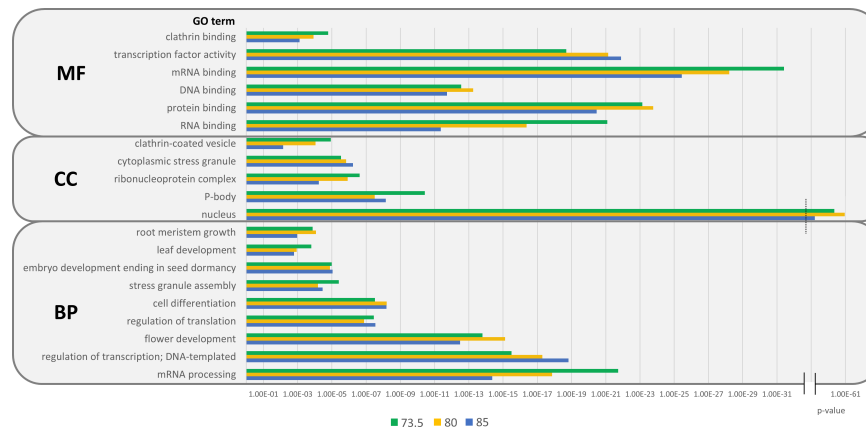


FIGURE 5

GO enrichments of pCARs from *A. thaliana* by threshold. Significant enrichments were found in cellular components (CC), biological processes (BP), and molecular functions (MF) at the three designated thresholds: 73.5 (green), 80 (yellow), and 85 (blue). pCARs are mainly associated with nuclear components and MLOs and are found in several regulatory processes and plant-specific functions.

(Garai et al., 2021) proteins from plants. The fact that a significant proportion of PrLDs is pCAR-positive makes it challenging to disentangle the contributions of the complete PrLD sequence and pCARs only in a given protein to the detected enrichments. However, even when the *A. thaliana* prionome is used as background for the GO analysis, significant enrichments in regulatory processes such as regulation of transcription (BP), transcription factor activity (MF), or nucleus (CC) were still observed (Supplementary Figure 1). *A. thaliana*'s pCARs are also enriched in plant-specific biological processes such as leaf and flower development or root meristem growth. In fact, there is already experimental evidence for plant PrLDs being involved in the autonomous flowering pathway, specifically for the luminidependens PrLD from *A. thaliana* (Chakrabortee et al., 2016).

pCARs are found to be associated with nuclear components, ribonuclear complexes, and nucleus, the latter with the lowest p-values observed ( $p < 10^{-61}$ ) among all GO terms analyzed. Certain membraneless organelles (MLOs), such as cytoplasmic stress granules or P-bodies were also enriched in pCARs. PrLDs are recruited to MLOs in response to different stimuli (Boncella et al., 2020; Iglesias et al., 2021), and they need to have some condensate driving element that facilitates the formation of liquid droplets. This would be especially relevant when the protein phase separates by itself, in the absence of additional co-factors (Pintado-Grima et al., 2022b); as described, pCARs might act as sticker regions facilitating PrLDs' coalescence and the stabilization of these cellular compartments (Tsoi et al., 2021). In general, RNA and mRNA binding and processing are associated with pCARs. This might reflect a situation in which many of the protein constituents of MLOs are originally nuclear

RNA binding proteins (Mann and Donnelly, 2021), with many of them bearing PrLDs (Harrison and Shorter, 2017; Gotor et al., 2020).

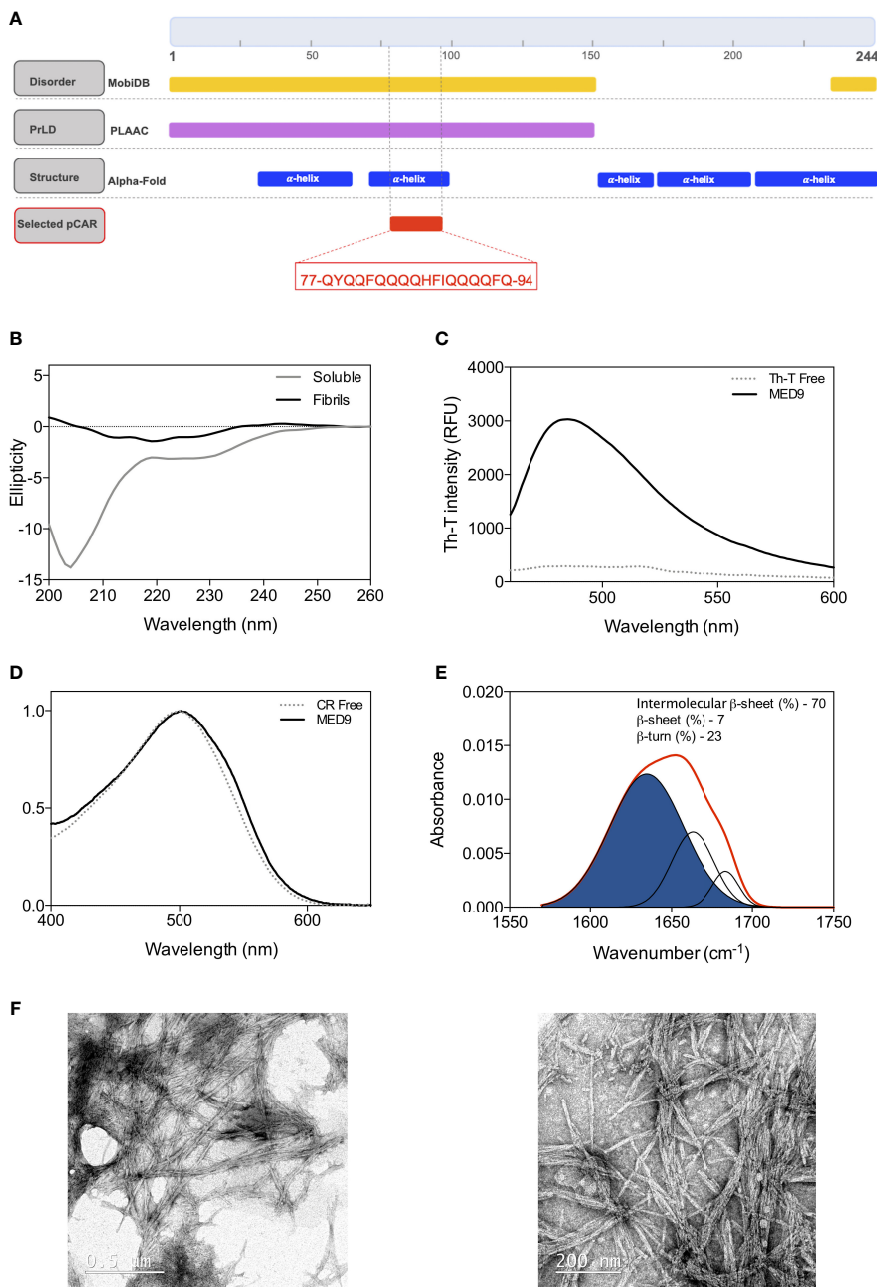
### 3.4 MED9 pCAR form amyloid fibrils

The concept of CARs has been gaining strength by means of the continuous experimental validation of predicted segments of hydrophilic nature that form amyloid fibrils *in vitro* (Santos et al., 2021; Pintado-Grima et al., 2022a). Our goal is to increase the body of evidence on the intrinsic amyloidogenic potential of these sequences, when they are disconnected from their interpCARs regions. Accordingly, we examined a new pCAR from *A. thaliana*'s MED9 protein (Q8RWA2) in this work.

MED9, the mediator of RNA polymerase II transcription subunit 9, is a nuclear protein that is part of the mediator complex. It acts as a coactivator involved in the transcriptional regulation of almost all RNA polymerase II-related genes. The N-terminal domain of MED9 (1-149) is predicted as a PrLD by PLAAC. Within this PrLD, an 18-residue pCAR comprising residues 77 to 94 (77-QYQQFQQQHFQQQQFQ-94) emerges as an interesting pCAR that we selected by its overall polar composition (67%), significant aromatic content (22%), low Waltz score (77.87), and for being undetectable by well-established aggregation and amyloid predicting algorithms like Aggrescan (Conchillo-Sole et al., 2007; De Groot et al., 2012), Tango (Fernandez-Escamilla et al., 2004) or ZipperDB (Thompson et al., 2006).

MobiDB (Piovesan et al., 2021) consistently predicts MED9 PrLD to belong to a long-disordered region (1-150) (Figure 6A). AlphaFold (Jumper et al., 2021; Varadi et al., 2022) predict that

### Mediator of RNA polymerase II transcription subunit - MED9



**FIGURE 6**

Experimental characterization of the predicted pCAR derived peptide in MED9 from *A thaliana*. **(A)** MED9 diagram showing the location of the predicted disordered regions (yellow), PrLD (violet), secondary structures (blue) and the selected pCAR (red). The sequence of pCAR is shown in the red box. **(B)** Far-UV circular dichroism spectra of the peptide, evolution from the initial disordered soluble state into mature amyloid fibrils at 50  $\mu$ M and 48 hours of incubation at 37  $^{\circ}$ C with continuous agitation at 100 RPM. **(C)** Thioflavin-T fluorescence emission spectrum when excited at 440nm in the absence (dashed line) and presence (solid line) of the aggregated sample; note the characteristic fluorescence enhancement at 480 nm when the dye is bound to amyloid-like aggregates. **(D)** CR spectral changes in the absence (dashed line) and in the presence (solid line) of the incubated peptide. **(E)** FTIR spectrum in the amide I region of the incubated sample. The red line corresponds to the absorbance spectra and the blue area indicates the inter-molecular  $\beta$ -sheet contribution to the total area upon Gaussian deconvolution. **(F)** Representative TEM micrographs illustrating the fibrils of the incubated peptide.



residues 32-64 and 70-97 might form two alpha-helices, which, in the absence of tertiary contacts, are probably only attained upon binding to a molecular partner (i.e., coiled-coil formation). The MED9 pCAR maps within the second predicted helix (Figure 6A), suggesting a potential implication in PPIs through a folding upon binding mechanism, as reported in a significant proportion of previously characterized CARs (Santos et al., 2021).

To assess the amyloidogenic potential of MED9 pCAR, we analyzed whether a peptide corresponding to its sequence was able to self-assemble into amyloid fibrils *in vitro*. We incubated the peptide at 50  $\mu$ M in 20 mM Tris and 100 mM NaCl pH 8 for 48 hours at 37 °C and 100RPM. After this time, the initial CD signal corresponding to the disordered soluble peptide was significantly reduced and shifted to a faint  $\beta$ -sheet signature, indicating protein aggregation (Figure 6B). The end-point sample was positive for amyloid dye binding (Th-T and CR), suggesting the formation of amyloid-like structures (Figures 6C, D). To assess the secondary structure content of the insoluble aggregates not accessible by CD, we used FTIR and recorded the amine I region of the infrared spectra (1700-1600  $\text{cm}^{-1}$ ). We observed a dominant contribution of the 1630  $\text{cm}^{-1}$  band (70% of the area) associated with the formation of amyloid-like intermolecular  $\beta$ -sheets (Figure 6E). A morphological analysis of the aggregates by TEM confirmed the fibrillar nature of the aggregates (Figure 6F).

Together, this data confirms that the MED9 pCAR has the inherent potential to aggregate into amyloid fibrils. Furthermore, the coincidence of predicted coiled-coil and amyloid propensities in the same sequence has been previously observed in a large number of human PrLDs and yeast prions (Fiumara et al., 2010; Batlle et al., 2021). Indeed, a human member of the mediator complex, MED15, is a *bona fide* prion-like protein whose PrLD transits between disordered, coiled-coil, and amyloid states (Batlle et al., 2021). As mentioned above, the MED9 pCAR is just one example of a conjunct of amyloid-competent sequences not detectable by classical aggregation predictors like our own Aggrescan (Conchillo-Sole et al., 2007).

## 4 Discussion

The prion phenomenon continues to attract high interest. In recent times, significant effort has been devoted to understanding the role PrLDs play in biological pathways across species, beyond those initially described in yeast prions. Although PrLDs seem to share a moderate amyloid propensity (Sabate et al., 2015b), the containing proteins do not always behave as *bona fide* prions, but rather exploit these LCDs to promote the functional interactions needed for their activity. Amyloid regions in PrLDs need to be milder and more polar

than the hydrophobic stretches often identified in the core of pathogenic amyloid fibrils; to keep the protein soluble but allow specific inter-molecular interactions. In this scenario, the concept of CARs, whose existence has already been validated in generic IDPs (Santos et al., 2021), is translated here to PrLDs, since these domains are, by definition, intrinsically disordered. In the PrLD context, pCARs would define the boundaries of the regions that would concentrate this mild amyloid potential, flanked by longer and highly soluble interpCARs segments, acting analogously to the way stickers and spacers do in LLPS (Martin et al., 2020).

pCAR composition suggests that these regions within PrLDs arise from a reduction of  $\beta$ -breakers and an increase in hydrophobic residues, especially of an aromatic character, relative to the rest of the domain. In addition, Tyr seems to play a unique role in pCARs, since it unites aromaticity and polarity, allowing interactions with water in the soluble state and pi-pi interactions in the assembled state. Notably, the content in polar residues of pCARs is equal to or higher than that of the complete PrLDs or even the whole proteome, which ensures that their amyloid propensity is framed in an overall soluble background. In this context, the polar Gln and specially Asn residues, which are highly enriched in PrLDs relative to the proteome but also in pCARs relative to these domains, would also contribute to control the amyloid load, since their side chains have a high tendency to form hydrogen bonds with the solvent and thus enable solubility, but at the same time, they can form very tight interactions within the dehydrated interfaces of  $\beta$ -strands (Balbirnie et al., 2001; Peccati et al., 2020). The preference for Asn relative to Gln in pCARs seems to make sense since Asn richness promotes the formation of benign assemblies, whereas Gln richness promotes the formation of toxic aggregates (Halfmann et al., 2011).

The present analysis uncovered a previously unexplored putative amyloid sequence space in plants PrLDs, relative to previous studies, including ours; where amyloid-prone regions identification was limited by rationally implemented but arbitrary requirements, including a fixed size, the absence of Pro or the belief that a single region would suffice to nucleate aggregation (Sabate et al., 2015a; Batlle et al., 2017b; Fernandez et al., 2017), essentially because the bias toward studying disease-associated amyloids made us wrongly thought that highly hydrophilic sequences, like pCARs, could not form amyloid structures autonomously. Importantly, the presence of these cryptic amyloidogenic regions seems to relate to RNA- and mRNA-related processes occurring in the nucleus, highlighting the significant role that PrLDs containing CARs might play in regulating plant-specific biochemical pathways.

Although the pCAR concept indeed requires extensive experimental validation before their presence can be established as a general principle modulating the activity of prion-like proteins, it is also certain that for most of the experimentally

validated CARs, structural analysis indicated that these initially disordered regions might fold upon interaction with a partner, or, alternatively, that they mediate multimerization. This suggests that its higher amyloidogenic load, relative to the rest of the PrLD sequence, cannot be purged out by natural selection because CARs are needed for functional interactions. Surprisingly, in many instances, the folded structures adopted by CARs in functional complexes correspond to  $\alpha$ -helices, often involved in coiled-coil dimerization (Batlle et al., 2021). Chameleon sequences adopt an  $\alpha$ -helix in the folded state of globular proteins but display a hidden propensity to form  $\beta$ -sheets when unfolded, and  $\alpha$ -helix/ $\beta$ -strand-discordant stretches were proposed to be associated with amyloid fibril formation already in 2001 (Kallberg et al., 2001). What makes special CARs, such as those of *A. thaliana* MED9 or human MED15, is that they might populate three different states, disordered in isolation,  $\alpha$ -helix in functional PPIs, and  $\beta$ -sheets if they eventually transition to the amyloid state; with the three propensities encoded in the same sequence but controlled by the protein microenvironment and the sequential context.

We believe the approach described in the present study may translate to the detection of pCARs in other interesting prionomes, to build a comprehensive repository of polar amyloidogenic regions for further experimental validation and benchmarking future predictive algorithms.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. The datasets are also available online at: <http://carsdb.ppmclab.com/prionplants>.

## Author contributions

CP-G, JS, and VI generated the dataset. CP-G and VI performed the analyses. ZM-A conducted the experimental validation. CP-G, IP, and SV conceptualized the project. CP-G, JS, and SV wrote the manuscript with the contributions of all authors. SV and IP acquired funding and reviewed the final

version of the manuscript. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was funded by European Union's Horizon 2020 research and innovation programme under GA 952334 (PhasAGE). SV was supported by the Spanish Ministry of Science and Innovation (PID 2019-105017RB-I00), by ICREA and ICREA Academia 2020. CP-G was supported by the Secretariat of Universities and Research of the Catalan Government and the European Social Fund (2021 FI\_B 00087). JS was supported by the Spanish Ministry of Science and Innovation via a doctoral Grant (FPU17/01157). VI was supported by the Spanish Ministry of Universities and the European Union-Next Generation EU (ruling 02/07/2021, Universitat Autònoma de Barcelona).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1060410/full#supplementary-material>

## References

- Afsar Minhas, F. U. A., Ross, E. D., and Ben-Hur, A. (2017). Amino acid composition predicts prion activity. *PLoS Comput. Biol.* 13, e1005465. doi: 10.1371/journal.pcbi.1005465
- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146–158. doi: 10.1016/j.cell.2009.02.044
- An, L., and Harrison, P. M. (2016). The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. *Biol. Direct.* 11, 32. doi: 10.1186/s13062-016-0134-5
- Antonets, K. S., and Nizhnikov, A. A. (2017). Predicting amyloidogenic proteins in the proteomes of plants. *Int. J. Mol. Sci.* 18. doi: 10.3390/ijms18102155

- Balbirnie, M., Grothe, R., and Eisenberg, D. S. (2001). An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2375–2380. doi: 10.1073/pnas.041617698
- Battle, C., Calvo, I., Iglesias, V., Lynch, C. J., Gil-Garcia, M., Serrano, M., et al. (2021). MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. *Commun. Biol.* 4, 414. doi: 10.1038/s42003-021-01930-8
- Battle, C., De Groot, N. S., Iglesias, V., Navarro, S., and Ventura, S. (2017a). Characterization of soft amyloid cores in human prion-like proteins. *Sci. Rep.* 7, 12134. doi: 10.1038/s41598-017-09714-z
- Battle, C., Iglesias, V., Navarro, S., and Ventura, S. (2017b). Prion-like proteins and their computational identification in proteomes. *Expert Rev. Proteomics* 14, 335–350. doi: 10.1080/14789450.2017.1304214
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Boncella, A. E., Shattuck, J. E., Cascarina, S. M., Paul, K. R., Baer, M. H., Fomicheva, A., et al. (2020). Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proc. Natl. Acad. Sci. U.S.A.* 117, 5826–5835. doi: 10.1073/pnas.1912723117
- Bremer, A., Farag, M., Borcherds, W. M., Peran, I., Martin, E. W., Pappu, R. V., et al. (2022). Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* 14, 196–207. doi: 10.1038/s41557-021-00840-w
- Chakrabortee, S., Kayatekin, C., Newby, G. A., Mendillo, M. L., Lancaster, A., and Lindquist, S. (2016). Luminidependens (LD) is an arabidopsis protein with prion behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6065–6070. doi: 10.1073/pnas.1604478113
- Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtov, S. G., and Liebman, S. W. (1995). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science* 268, 880–884. doi: 10.1126/science.7754373
- Choi, K. J., Quan, M. D., Qi, C., Lee, J. H., Tsoi, P. S., Zahabiyon, M., et al. (2022). NANOG prion-like assembly mediates DNA bridging to facilitate chromatin reorganization and activation of pluripotency. *Nat. Cell Biol.* 24, 737–747. doi: 10.1038/s41556-022-00896-x
- Conchillo-Sole, O., De Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESKAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinf.* 8, 65. doi: 10.1186/1471-2105-8-65
- De Groot, N. S., Castillo, V., Grana-Montes, R., and Ventura, S. (2012). AGGRESKAN: method, application, and perspectives for drug design. *Methods Mol. Biol.* 819, 199–220. doi: 10.1007/978-1-61779-465-0\_14
- Derkatch, I. L., Uptain, S. M., Outeiro, T. F., Krishnan, R., Lindquist, S. L., and Liebman, S. W. (2004). Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI+] prion in yeast and aggregation of Sup35 *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12934–12939. doi: 10.1073/pnas.0404968101
- Diaz-Caballero, M., Navarro, S., Fuentes, I., Teixidor, F., and Ventura, S. (2018). Minimalist prion-inspired polar self-assembling peptides. *ACS Nano* 12, 5394–5407. doi: 10.1021/acsnano.8b00417
- Dorone, Y., Boeynaems, S., Flores, E., Jin, B., Hateley, S., Bossi, F., et al. (2021). A prion-like protein regulator of seed germination undergoes hydration-dependent phase separation. *Cell* 184, 4284–4298 e4227. doi: 10.1016/j.cell.2021.06.009
- Fernandez, M. R., Battle, C., Gil-Garcia, M., and Ventura, S. (2017). Amyloid cores in prion domains: Key regulators for prion conformational conversion. *Prion* 11, 31–39. doi: 10.1080/19336896.2017.1282020
- Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306. doi: 10.1038/nbt1012
- Fiumara, F., Fioriti, L., Kandel, E. R., and Hendrickson, W. A. (2010). Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143, 1121–1135. doi: 10.1016/j.cell.2010.11.042
- Garai, S., Citu, Singla-Pareek, S. L., Sopory, S. K., Kaur, C., and Yadav, G. (2021). Complex networks of prion-like proteins reveal cross talk between stress and memory pathways in plants. *Front. Plant Sci.* 12, 707286. doi: 10.3389/fpls.2021.707286
- Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* 26, 326–332. doi: 10.1093/bioinformatics/btp691
- Gil-Garcia, M., Iglesias, V., Pallares, I., and Ventura, S. (2021). Prion-like proteins: from computational approaches to proteome-wide analysis. *FEBS Open Bio* 11, 2400–2417. doi: 10.1002/2211-5463.13213
- Gotor, N. L., Armaos, A., Calloni, G., Torrent Burgas, M., Vabulas, R. M., De Groot, N. S., et al. (2020). RNA-Binding and prion domains: the yin and yang of phase separation. *Nucleic Acids Res.* 48, 9491–9504. doi: 10.1093/nar/gkaa681
- Halfmann, R., Alberti, S., Krishnan, R., Lyle, N., O'donnell, C. W., King, O. D., et al. (2011). Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol. Cell* 43, 72–84. doi: 10.1016/j.molcel.2011.05.013
- Harrison, A. F., and Shorter, J. (2017). RNA-Binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438. doi: 10.1042/BCJ20160499
- Holehouse, A. S., Ginell, G. M., Griffith, D., and Boke, E. (2021). Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates. *Biochemistry* 60, 3566–3581. doi: 10.1021/acs.biochem.1c00465
- Iglesias, V., De Groot, N. S., and Ventura, S. (2015). Computational analysis of candidate prion-like proteins in bacteria and their role. *Front. Microbiol.* 6, 1123. doi: 10.3389/fmicb.2015.01123
- Iglesias, V., Paladin, L., Juan-Blanco, T., Pallares, I., Aloy, P., Tosatto, S. C. E., et al. (2019). In silico characterization of human prion-like proteins: Beyond neurological diseases. *Front. Physiol.* 10, 314. doi: 10.3389/fphys.2019.00314
- Iglesias, V., Santos, J., Santos-Suarez, J., Pintado-Grima, C., and Ventura, S. (2021). SGnn: A web server for the prediction of prion-like domains recruitment to stress granules upon heat stress. *Front. Mol. Biosci.* 8, 718301. doi: 10.3389/fmolb.2021.718301
- Jiang, Y., Li, H., Zhu, L., Zhou, J. M., and Perrett, S. (2004). Amyloid nucleation and hierarchical assembly of Ure2p fibrils. role of asparagine/glutamine repeat and nonrepeat regions of the prion domains. *J. Biol. Chem.* 279, 3361–3369. doi: 10.1074/jbc.M310494200
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kallberg, Y., Gustafsson, M., Persson, B., Thyberg, J., and Johansson, J. (2001). Prediction of amyloid fibril-forming proteins. *J. Biol. Chem.* 276, 12945–12950. doi: 10.1074/jbc.M010402200
- King, O. D., Gitler, A. D., and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* 1462, 61–80. doi: 10.1016/j.brainres.2012.01.016
- Lancaster, A. K., Nutter-Upham, A., Lindquist, S., and King, O. D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* 30, 2501–2502. doi: 10.1093/bioinformatics/btu310
- Mann, J. R., and Donnelly, C. J. (2021). RNA Modulates physiological and neuropathological protein phase transitions. *Neuron* 109, 2663–2681. doi: 10.1016/j.neuron.2021.06.023
- Martin, E. W., Holehouse, A. S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., et al. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* 367, 694–699. doi: 10.1126/science.aaw8653
- Martins, P. M., Navarro, S., Silva, A., Pinto, M. F., Sarkany, Z., Figueiredo, F., et al. (2020). MIRRAGGE - minimum information required for reproducible AGGREGATION experiments. *Front. Mol. Neurosci.* 13, 582488. doi: 10.3389/fnmol.2020.582488
- Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., et al. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* 7, 237–242. doi: 10.1038/nmeth.1432
- Michelitsch, M. D., and Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11910–11915. doi: 10.1073/pnas.97.22.11910
- Pallares, I., De Groot, N. S., Iglesias, V., Sant'anna, R., Biosca, A., Fernandez-Busquets, X., et al. (2018). Discovering putative prion-like proteins in plasmodium falciparum: A computational and experimental analysis. *Front. Microbiol.* 9, 1737. doi: 10.3389/fmicb.2018.01737
- Peccati, F., Diaz-Caballero, M., Navarro, S., Rodriguez-Santiago, L., Ventura, S., and Sodupe, M. (2020). Atomistic fibrillar architectures of polar prion-inspired heptapeptides. *Chem. Sci.* 11, 13143–13151. doi: 10.1039/D0SC05638C
- Perutz, M. F., Pope, B. J., Owen, D., Wanker, E. E., and Scherzinger, E. (2002). Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid beta-peptide of amyloid plaques. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5596–5600. doi: 10.1073/pnas.042681599
- Pintado-Grima, C., Barcenas, O., Manglano-Artunedo, Z., Vilaca, R., Macedo-Ribeiro, S., Pallares, I., et al. (2022a). CARs-DB: A database of cryptic amyloidogenic regions in intrinsically disordered proteins. *Front. Mol. Biosci.* 9, 882160. doi: 10.3389/fmolb.2022.882160
- Pintado-Grima, C., Barcenas, O., and Ventura, S. (2022b). In-silico analysis of pH-dependent liquid-liquid phase separation in intrinsically disordered proteins. *Biomolecules* 12. doi: 10.3390/biom12070974

- Piovesan, D., Necci, M., Escobedo, N., Monzon, A. M., Hatos, A., Micetic, I., et al. (2021). MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res.* 49, D361–D367. doi: 10.1093/nar/gkaa1058
- Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. *Science* 216, 136–144. doi: 10.1126/science.6801762
- Rousseau, F., Serrano, L., and Schymkowitz, J. W. (2006). How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.* 355, 1037–1047. doi: 10.1016/j.jmb.2005.11.035
- Sabate, R., Rousseau, F., Schymkowitz, J., Batlle, C., and Ventura, S. (2015a). Amyloids or prions? that is the question. *Prion* 9, 200–206. doi: 10.1080/19336896.2015.1053685
- Sabate, R., Rousseau, F., Schymkowitz, J., and Ventura, S. (2015b). What makes a protein sequence a prion? *PLoS Comput. Biol.* 11, e1004013. doi: 10.1371/journal.pcbi.1004013
- Sant'anna, R., Fernandez, M. R., Batlle, C., Navarro, S., De Groot, N. S., Serpell, L., et al. (2016). Characterization of amyloid cores in prion domains. *Sci. Rep.* 6, 34274. doi: 10.1038/srep34274
- Santos, J., Pallares, I., Iglesias, V., and Ventura, S. (2021). Cryptic amyloidogenic regions in intrinsically disordered proteins: Function and disease association. *Comput. Struct. Biotechnol. J.* 19, 4192–4206. doi: 10.1016/j.csbj.2021.07.019
- Serio, T. R., Cashikar, A. G., Moslehi, J. J., Kowal, A. S., and Lindquist, S. L. (1999). Yeast prion [psi<sup>+</sup>] and its determinant, Sup35p. *Methods Enzymol.* 309, 649–673. doi: 10.1016/S0076-6879(99)09043-6
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists, (2021 update). *Nucleic Acids Res.* 50 (W1), W216–W221. doi: 10.1093/nar/gkac194
- Sormanni, P., Aprile, F. A., and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* 427, 478–490. doi: 10.1016/j.jmb.2014.09.026
- Tetz, G., and Tetz, V. (2018). Prion-like domains in eukaryotic viruses. *Sci. Rep.* 8, 8931. doi: 10.1038/s41598-018-27256-w
- Theillet, F. X., Kalmar, L., Tompa, P., Han, K. H., Selenko, P., Dunker, A. K., et al. (2013). The alphabet of intrinsic disorder: I. act like a pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord. Proteins* 1, e24360. doi: 10.4161/idp.24360
- Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D., and Eisenberg, D. (2006). The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4074–4078. doi: 10.1073/pnas.0511295103
- Toombs, J. A., McCarty, B. R., and Ross, E. D. (2010). Compositional determinants of prion formation in yeast. *Mol. Cell Biol.* 30, 319–332. doi: 10.1128/MCB.01140-09
- Toombs, J. A., Petri, M., Paul, K. R., Kan, G. Y., Ben-Hur, A., and Ross, E. D. (2012). De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6519–6524. doi: 10.1073/pnas.1119366109
- True, H. L., and Lindquist, S. L. (2000). A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* 407, 477–483. doi: 10.1038/35035005
- Tsoi, P. S., Quan, M. D., Choi, K. J., Dao, K. M., Ferreon, J. C., and Ferreon, A. C. M. (2021). Electrostatic modulation of hnRNPA1 low-complexity domain liquid-liquid phase separation and aggregation. *Protein Sci.* 30, 1408–1417. doi: 10.1002/pro.4108
- UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi: 10.1093/nar/gkab1061
- Wang, J., Choi, J. M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., et al. (2018). A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 174, 688–699 e616. doi: 10.1016/j.cell.2018.06.006
- Wickner, R. B. (1994). [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* 264, 566–569. doi: 10.1126/science.7909170
- Yuan, A. H., and Hochschild, A. (2017). A bacterial global regulator forms a prion. *Science* 355, 198–201. doi: 10.1126/science.aai7776
- Zambrano, R., Conchillo-Sole, O., Iglesias, V., Illa, R., Rousseau, F., Schymkowitz, J., et al. (2015). PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res.* 43, W331–W337. doi: 10.1093/nar/gkv490
- Zhang, Y., Man, V. H., Roland, C., and Sagui, C. (2016). Amyloid properties of asparagine and glutamine in prion-like proteins. *ACS Chem. Neurosci.* 7, 576–587. doi: 10.1021/acscchemneuro.5b00337