Computational and Structural Biotechnology Journal 19 (2021) 4192–4206

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Cryptic amyloidogenic regions in intrinsically disordered proteins: Function and disease association

Jaime Santos *, Irantzu Pallarès, Valentín Iglesias, Salvador Ventura *

Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 1 April 2021
Received in revised form 23 July 2021
Accepted 23 July 2021
Available online 26 July 2021

Keywords:
Amyloid
Aggregation
Protein disorder
Intrinsically disordered proteins
Protein–protein interactions
Evolution

ABSTRACT

The amyloid conformation is considered a fundamental state of proteins and the propensity to populate it a generic property of polypeptides. Multiple proteome-wide analyses addressed the presence of amyloidogenic regions in proteins, nurturing our understanding of their nature and biological implications. However, these analyses focused on highly aggregation-prone and hydrophobic stretches that are only marginally found in intrinsically disordered regions (IDRs). Here, we explore the prevalence of cryptic amyloidogenic regions (CARs) of polar nature in IDRs. CARs are widespread in IDRs and associated with IDPs function, with particular involvement in protein–protein interactions, but their presence is also connected to a risk of malfunction. By exploring this function/malfunction dichotomy, we speculate that ancestral CARs might have evolved into functional interacting regions playing a significant role in protein evolution at the origins of life.

1. Introduction

The information for folding and activity is naturally encoded in protein primary sequences [1]. Sequences are constrained by evolution to preserve proteins' structure and function and prevent aberrant interactions. Despite this purifying selection, polypeptides retain a certain potential to self-assemble into amyloid fibrils that competes with folding and the attainment of native conformations [2–5]. Uncontrolled amyloid assembly has a detrimental impact on cell fitness, sequestering functional proteins into inactive complexes and/or generating cytotoxic species, a process connected with over 37 human pathologies including Alzheimer's and Parkinson's diseases [2,6].

Amyloidogenicity seems indeed to be a generic property of the polypeptide chain, intimately associated with the ability of proteins to fold and function [7]. Aggregation-prone regions often map at the hydrophobic core of globular proteins or in buried positions engaged in the cooperative interactions that sustain the native state [8–11]. Accordingly, recent studies propose that amyloidogenicity and protein stability are evolutionary coupled, since the impact of mutations in globular and amyloid states are thermodynamically correlated [12]. To a lesser extent, amyloidogenic regions are also found in solvent-exposed regions involved in functional activities such as binding or catalysis [13–16]. Thus, protein aggregation propensity is considered to be an unavoidable side effect associated to the emergence and evolution of globular proteins [12,17].

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) have been traditionally considered significantly depleted of aggregation-prone regions [12,18]. This stems from the lack of a hydrophobic core and their particular compositional bias that inherently confers them a high overall solubility and protects from aggregation [19,20]. α-Synuclein, IAPP and other IDPs whose aggregation is associated with human disease are the exception rather than the rule.

IDRs stablish multivalent protein–protein interactions (PPI) with structurally different partners, a central property of IDPs biology [21]. In these events, the entropic penalty of the interaction is balanced by a dense network of enthalpic contacts. Whereas it is now clear that in globular proteins the determinants for functional

Abbreviations: APR, Aggregation-prone region; CARs, Cryptic amyloidogenic regions; CD, Circular dichroism; CR, Congo red; FTIR, Fourier transform infrared; IDPs, Intrinsically disordered proteins; IDRs, Intrinsically disordered regions; PBS, Phosphate buffer saline; PPI, Protein-protein interactions; TEM, Transmission electron microscopy; Th-T, Thioflavin-T; Rb, Retinoblastoma associated proteins; RbC, Core region of Rb.

* Corresponding authors at: Institut de Biotecnologia i de Biomedicina, Parc de Recerca UAB, Mòdul B, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain.

E-mail addresses: jaime.santos@uab.es (J. Santos), salvador.ventura@uab.es (S. Ventura).

https://doi.org/10.1016/j.csbj.2021.07.019

binding overlap to a significant extent with the ones driving self-assembly into amyloid conformers [4,14]; this function-amyloid duality has not been reported for IDPs, essentially because state of the art aggregation predictors indicated that amyloidogenic regions are scarce in this protein subset [12,18].

Recent observations indicate that protein sequences consisting of more hydrophilic and less aliphatic residues than pathogenic amyloids can also assemble into fibrils [22–24]. Their low hydrophobic character would have impeded to identify these cryptic amyloidogenic regions (CARs) in previous proteome-wide analysis. Here, we implemented a computational strategy to explore the prevalence of CARs in IDRs. Our data indicate that >50 % of IDRs contain at least one region of mild, but significant, amyloid propensity. The data suggest that these CARs are associated with the establishment of PPI and protein complexes formation, and thus their presence is inherent to IDPs function.

The natural abundance of human proteins containing CARs seems to be inversely correlated to the effective concentrations of these stretches at their IDRs, suggesting that, aside from being functional, their excessive accumulation might be detrimental. Accordingly, proteins holding CARs are more associated with diseases like Parkinson's and Alzheimer's than those devoid of them and they appear to play a central role in cancer related pathways. We demonstrate that CARs display full amyloidogenic potential by evaluating the capacity to assemble into amyloid fibrils of two short peptides comprised within these regions in human p53 and the retinoblastoma-associated protein. Finally, we hypothesize that CARs might be a remanent of the amyloid origin of life [25–28].

2. Materials and methods

2.1. Data acquisition

The dataset of IDRs was obtained from the DisProt database (release 2019_09) [29]. Folded proteins were retrieved from the Astral SCOPe (Structural Classification of Proteins — extended) 2.07 genetic domain sequence subsets, and filtered for less than 40 % identity to each other [30]. Linear interacting peptides (LIPs) were extracted from MobiDB that recapitulates manually curated linear motives from other sources: ELM, DIBS, MFIB, DisProt and IDEAL (Data retrieval performed in September 2019) [31]. Protein ID mapping was performed using the UniProt ID mapping tool (http://www.uniprot.org/mapping/). Experimental data on peptides with amyloidogenic behavior were extracted from literature cited in the Table 1. Cancer associations were found in the following references [32–50].

2.2. CARs screening in DisProt

Disordered regions were obtained from the DisProt database (release 2019_09) and were curated for completely overlapped regions or regions already contained in longer annotations [29]. Regions shorter than 21 residues were pre-excluded from the analysis. Disordered regions were introduced as an input in the Waltz algorithm [51] and were scanned for positive peptides using different thresholds (73.5, 80.0 and 85.0), in accordance with the proposed criteria. Duplicated positive peptides arising for partial overlaps in disordered regions were manually curated to avoid duplications. The same procedure was applied when the dataset of folded proteins was analyzed.

2.3. Gene ontology annotation

Gene ontology terms were analyzed with the Functional Annotation Tool of DAVID v6.8 (Database for Annotation, Visualization and Integrated Discovery) [52]. The complete DisProt dataset was used as a background set to infer enrichment. The GO terms (GO_DIRECT categories) enriched with a p-value (Modified Fishers Exact p-value) < 0.1 were retrieved. Disease association was also assessed with DAVID (GAD_DISEASE).

2.4. Analysis of the overlap between CARs and LIPs

To calculate the degree of overlapping between predicted CARs and annotated LIPs we employed and in-house script to compare the beginning and end positions of each of the predicted CARs with those from the LIPs annotated in the same DisProt region and extract the overlapping residues. We considered a positive overlap if at least 1 residue coincided. The percentage of residues of the CAR overlapping with the LIPs was calculated as the number of overlapping residues in relation with the total residues of the CAR. To assess the degree of overlapping that could be expected by chance, we used a randomization script to generate a dataset of 285 random DisProt regions with randomly distributed 8-residue peptides (median of LIPs length) that mimicked LIPs. Overlapping was tested as previously described.

2.5. Protein abundance

Protein abundance data for the analyzed human proteins was taken from PaxDB [53]. The obtained values were then log10 transformed for the analysis. Abundance data was retrieved for 488 of the 525 human proteins. Protein solubilities were predicted using the CamSol method [54]. Statistical analyses were performed using a non-parametric *t*-test to compare cumulative distributions (Kolmogorov-Smirnov test).

2.6. Human network analysis

Data on protein–protein interactions of the human proteins in the 85-threshold dataset was retrieved from STRING [55]. PPI interaction network was constructed using Cytoscape (v 3.8.2) and the STRING plugin was exploited for visualization. Clustering was performed using the AutoAnnotate v1.3.3 Cytoscape app with the clusterMaker2 MCL algorithm. Network ontology annotations were performed with the STRING Cytoscape plugin.

2.7. Peptide preparation and aggregation

Peptides with sequences Ac-ILVSIGESFG-NH2 and Ac-EYFTLQIR-NH2 were purchased from Synpeptide (Shanghai, China) with a purity >95%. Peptides were N-terminal acetylated and C-terminal amidated to neutralize terminal charges and mimic the protein environment. Peptides were dissolved in hexafluoroisopropanol to a concentration of 1 mg / mL, divided into aliquots (200 μL per aliquot) and vacuum dried with a SpeedVac (Thermo Fisher Scientific, Waltham, USA). Peptide aliquots were stored at −80 °C until assayed. Right before each experiment, aliquots were reconstituted with 20 μL hexafluoroisopropanol and further diluted in phosphate buffer saline (PBS) at pH 7.4 to a concentration of 200 μM. Hexafluoroisopropanol traces were gently evaporated under a stream of N_2. For aggregation assays, samples were incubated for 2 days at 37 °C with continuous agitation at 100 RPM in a 96 wells plate (non-treated) (Sarstedt, Germany). Each well contained 150 μL solutions of 200 μM peptide in PBS buffer. Plates were incubated at 37 °C, 100 RPM in an orbital culture shaker.

2.8. Binding to amyloid dyes

The fluorescence spectra of thioflavin-T (Th-T) were recorded using a Spark plate reader (Tecan, Switzerland). Spectra acquisition

Table 1
List of amyloidogenic peptides predicted by Waltz in disease-associated amyloidogenic proteins (top) and experimentally validated amyloidogenic polar peptides (bottom). Predictions were performed using the "best overall performance" default threshold (92.0) and an adapted threshold (73.5). Peptides experimentally described as structural elements of amyloid fibrils are noted with a plus sign (+). Negative Waltz predictions are indicated with a negative sign (-). n.a: not analyzed; n.i: no information. The references from which we extracted the data are indicated adjacent to the protein/peptide name.

Protein/Peptide	Waltz best overall performance (92.0)			Adapted threshold (73.5)			Experimental validation
	Positive peptide	Sequence	Waltz score	Positive peptide	Sequence	Waltz score	
Disease-linked polypeptides							
Aß-42 [62]	16–21	KLVFFA	97.99	n.a.	n.a	n.a	+
	37–42	GGVVIA	92.30	n.a.	n.a	n.a	+
α-synuclein [62]	35–40	EGVLYV	97.32	n.a.	n.a	n.a	+
IAPP [62,63]	7–18	QVFLIVLSVALN	93.89	n.a.	n.a	n.a	+
	55–62	NFGAILSS	93.98	n.a.	n.a	n.a	+
Beta-2-microglobulin [62]	26–30	KIQVY	92.98	n.a.	n.a	n.a	+
	78–90	KDWSFYLLYYTEF	96.58	n.a.	n.a	n.a	+
Transthyretin [62]	111–117	AEVVFTA	94.31	n.a.	n.a	n.a	+
TAU [64]	591–596	KVQIIN	97.99	n.a.	n.a	n.a	+
Cystatin-C [65]	30–35	AVGEYN	92.31	n.a.	n.a	n.a	n.i.
	58–66	AGVNYFLDV	97.43	n.a.	n.a	n.a	+
	98–104	SFQIYAV	98.66	n.a.	n.a	n.a	+
Polar amyloidogenic peptides							
NY7 [22]	2–7	YNYNYN	97.32	1–7	NYNYNYN	94.79	+
QY7 [22]	–	–	–	1–7	QYQYQYQ	87.01	+
SY7 [22]	–	–	–	1–7	SYSYSYS	84.90	+
GY7 [22]	–	–	–	2–7	YGYGYG	74.25	+
GYNGFG [23]	–	–	–	1–6	GYNGFG	77.25	+
SYSGYS [23]	–	–	–	1–6	SYSGYS	79.93	+
SYSSYGQS [23]	–	–	–	1–8	SYSSYGQS	85.95	+
NFSIRY [24]	–	–	–	1–6	NFSIRY	79.93	+
EVYKWT [24]	–	–	–	1–6	EVYKWT	77.26	+
AEIEIE [24]	–	–	–	1–6	AEIEIE	73.91	+
IQVYSR [24]	–	–	–	1–6	IQVYSR	84.95	+
IGYNIK [24]	–	–	–	1–6	IGYNIK	86.62	+
MFSEFD [24]	–	–	–	1–6	MFSEFD	75.92	+

was recorded from 470 to 600 nm with an excitation wavelength of 440 nm, 5 nm of excitation bandwidth, 10 nm of emission bandwidth and 1 nm interval. Peptides were preincubated with Th-T at a final concentration of 40 μM for 10 min before measurement. PBS pH 7.4 buffer with 40 μM Th-T and without peptide was used as a control.

Congo Red (CR) binding to peptide fibrils was analyzed using a Cary 100 UV/Vis spectrophotometer (Varian, Palo Alto, CA, USA) in the 400–700 nm range using a 1 cm optical length quartz cuvette. Samples (100 μL) were preincubated in 900 μL of CR at a final concentration of 5 μM for 10 min at room temperature. PBS pH 7.4 buffer with 5 μM CR and without peptide was used as a control.

2.9. Transmission electron microscopy

For transmission electron microscopy analysis (TEM) aggregated peptides were diluted to 100 μM Milli-Q water, sonicated for 5 min at minimum intensity in an ultrasonic bath (VWR ultrasonic cleaner) and placed onto carbon-coated copper grids to adsorb for 1 min. The excess of sample was carefully absorbed using ashless filter paper. Then, the grids were washed with distilled water and negative stained with 2 % (w/v) uranyl acetate for 1 min. The excess of uranyl acetate was absorbed using ashless filter paper. A TEM JEM-1400 (JEOL, Peabody, USA) microscope was used operating at an accelerating voltage of 120 kV. Representative images of each grid were selected.

2.10. Far circular dichroism spectroscopy

Far-UV circular dichroism (CD) spectra of the different peptide solutions were recorded on a Jasco J-815CD spectrometer (Halifax, Canada) at 25 °C. CD signal was measured from 260 nm to 200 nm at 0.2 nm intervals, 2 nm bandwidth, 2 s of response time and a scan speed of 200 nm/min on a 0.1 cm quartz cell. 10 accumulations were recorded and averaged for each measurement.

2.11. Attenuated Total Reflectance Fourier transform infrared spectroscopy

Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy experiments were performed using a Bruker Tensor 27 FTIR spectrometer (Bruker Optics Inc) with a Golden Gate MKII ATR accessory. Samples were dried out under a N2 stream and measured at spectral resolution of 2 cm^{-1} within the 1800–1500 cm^{-1} range. Each spectrum consisted of 16 independent scans. All spectral data were acquired and normalized using the OPUS MIR Tensor 27 software. Data were deconvoluted using the Peak Fit 4.12 program (Systat Software Inc., San Jose, CA, USA).

3. Results

3.1. Defining conditions to screen for polar amyloidogenic sequences

The dissection of the sequential features driving amyloid formation in disease-associated proteins (i.e., Aβ42, α-synuclein, TAU, IAPP, Transthyretin, β2-microglobulin) led to the conventional definition of amyloidogenic regions as highly hydrophobic and β-sheet prone sequence stretches, ideally devoid of net charge, able to nucleate the assembly of amyloid fibrils [2,56]. However, evidence gathered in the last years indicates that amyloid propensity is not necessarily restricted to these conventional low-solubility/aggregation-prone stretches. Thus, physicochemically divergent regions of a more soluble nature have been reported to retain the capacity to self-assemble into amyloid fibrils, expanding the amyloid sequence space [22–24].

In globular proteins, destabilizing mutations that facilitate the exposure of hydrophobic aggregation-prone regions previously protected in the tertiary/quaternary structure cause amyloidosis [10,57]. Therefore, sequences of this kind cannot be frequent in IDPs, where they would be constitutively exposed to solvent. Thus, it is not surprising that current computational approaches, benchmarked against pathological amyloids [58,59], do not detect these regions in IDPs in general and, in particular, at IDRs involved in physiological binding, which are often polar. This has led to propose that the association between function and amyloidogenicity observed in globular proteins does not apply to IDPs. However, it could be simply that classical algorithms are not sensitive enough to the presence of weaker amyloidogenic regions of a less hydrophobic nature.

To explore this hypothesis, we used Waltz, a well-validated algorithm that employs an empirical position-specific scoring matrix derived from the analysis of a database of short amyloid-forming and non-forming sequences [51]. Waltz predictions arise from the experimentally determined propensity of a protein sequence to adopt an amyloid fold, being less biased by the physicochemical nature of amino acids than other aggregation prediction algorithms, like CamSol, Aggrescan, or Tango [54,60,61], and thus theoretically capable of working on a broader spectrum of amyloid sequences. Waltz's "Best overall performance" threshold (threshold = 92.0) is set up to identify conventional amyloids and, as shown in Table 1, when used to scan the sequences of disease-associated amyloidogenic proteins, it identifies one or several high scoring sequence-stretches within them. However, when this threshold is used to analyze a dataset of polar peptides experimentally shown to form amyloids, Waltz predicts them as non-amyloid sequences in the large majority of cases (Table 1). Making the threshold less stringent may allow identifying these less potent amyloids, and a 73.5 value is sufficient to detect them all. With the idea of covering the amyloid propensities ranging between the minimum required for polar amyloid identification (73.5) and the one established for conventional amyloids (92.0), we gradually increased the Waltz threshold from 73.5 to 85.0 in subsequent analysis.

3.2. Prevalence of cryptic amyloidogenic regions within IDPs

DisProt is a manually curated database that assembles data from the literature with experimental evidence on proteins' disordered states [29,66]. DisProt contains over 3000 annotated regions of 1411 different proteins that have been experimentally validated (release 2019_09). Among these proteins, 1135 have regions with more than 20 residues, which were considered of sufficient length to be included in the analysis, as per the standard in the field [67].

To systematically screen these disordered regions, we applied the Waltz algorithm using 73.5, 80.0, and 85.0 thresholds. Positively detected regions consisting of at least seven residues are further referred to as Cryptic Amyloid Regions or CARs. The results of each analysis are summarized in Table 2, and CARs containing proteins are listed in Supplementary Data 1. We considered positive proteins/regions those containing at least one CAR. We predicted 1492 to 3472 CARs depending on the applied threshold, which resulted in 56 to 84 % of the analyzed proteins containing at least one CAR in their IDRs (Table 2). Similar trends were observed when we focused on human proteins with 55 to 80 % of positive hits (Table 2).

We found an average of 2.2 to 3.3 CARs per disordered region, with most of the regions displaying between 1 and 4 CARs (Table 2, Fig. 1A). By analyzing the length of positive IDRs, we observed that the most significant differences between thresholds occur in the shortest disordered regions in the dataset (21 to 50 residues) (Fig. 1B). Since a seven residue CAR would correspond to 14 to 33 % of these shorter stretches, we conjecture that weak amyloidogenic regions would be preferentially allowed there, those already detected with the 73.5 threshold, whereas longer disordered sequences could buffer more potent CARs. We noticed a significant correlation between the number of CARs and the length of the IDRs, suggesting that these amyloidogenic regions are, on average, homogeneously distributed across the sequence (Fig. 1C). This correlation is strongly dependent on the regions' amyloid potential, as we appreciated a clear difference between the 73.5 and the other threshold datasets. We found an average of 3.1, 2.4, and 2.0 CARs each 100 residues, which for the 80- and 85-threshold datasets corresponds to roughly 1 CAR per 50 residues, coincident with the drop of those CARs in < 50 residues long regions (Table 2 and Fig. 1D). Of note, in globular proteins, the prevalence of aggregation-prone regions (APRs) is near this value, with an average of 1 APR every 50 residues [12,18].

We then compared the prevalence of amyloidogenic regions at the three proposed Waltz thresholds in a database of folded proteins (Astral SCOPe 2.07 [30]), consisting of 14,323 different sequences based on PDB SEQRES records and with less than 40 % of identity. We found that 99.83, 97.81, and 95.32 % of the analyzed proteins (for the 73.5, 80.0, and 85.0 thresholds, respectively) contain at least one positive region. These results fit with the already described evolutionary association between amyloidogenicity and globular protein stability [12] and suggest that an intrinsic amyloidogenic potential is present in virtually every folded protein. Thus, the amyloidogenic load that a given protein sequence can accommodate seems to depend on the structural context strongly. This reinforces the need for dedicated and distinct amyloid prediction schemes when dealing with folded proteins and IDPs.

Taken together, the results indicate that IDPs possess a significant amyloidogenic load embedded within their polar sequences. This trait is observed for a significant fraction of the analyzed disordered regions, suggesting it has a ubiquitous nature. We note that these regions' prevalence depends on their amyloidogenic potential and the length of their hosting IDRs, which suggests a certain evolutive pressure to avoid uncontrolled spanning of CARs in disordered, and thus exposed, protein segments.

3.3. Amino acid composition of cryptic amyloidogenic regions

We studied the compositional traits of CARs. First, we asked whether these regions could accommodate charged residues. It turned out that 88.5, 84.0 and 81.3 % of the CARs contain at least one charged residue (Asp, Arg, Glu or Lys), with 74.5, 71.5 and 69.4 % of them expected to be charged at physiological pH (for the 73.5, 80.0, and 85.0 thresholds, respectively) (Table 3). The eminently charged character of CARs is not completely surprising since most IDPs are polyampholytes, with sequences that include both positively and negatively charged residues [68], but it strongly contrasts with that of the amyloidogenic regions of globular proteins, which being buried in the native fold, tend to be devoid of charges. Interestingly, acidic CARs are >2-fold more abundant than cationic ones, independently of the considered threshold (Table 3). This prevalence of negatively over positively charged residues might respond to a strategy to modulate the amyloidogenic potential of these sequences, since acidic amino acids seem to be intrinsically less amyloid-prone than cationic ones, owing both to their shorter sidechain lengths and the specific characteristics of the charge-carrying carboxyl moiety [69].

Next, we compared the composition of the identified CARs with that of DisProt disordered regions (Table 4). Hydrophilic amino acids (polar + ionizable) account for roughly 50 % of the residues in both CARs and IDRs, with polar and ionizable residues, being slightly over and underrepresented in CARs, respectively. This is expected, since despite the eminent charged nature of CARs, an

Table 2
Identification and prevalence of CARs in DisProt. The database was screened with Waltz, considering three thresholds, and positive proteins/regions were defined as those containing at least one CAR. The percentage of positive proteins for each threshold was calculated relative to the total number of proteins and human proteins with IDRs longer than 20 residues.

Threshold		Proteins	Proteins with regions > 20 residues	Number of CARs	Positive regions	Positive proteins	Negative proteins (regions > 20 residues)	Average length	CARs per region	CARs per protein	CARs each 100 residues	Positive proteins (%)
73.5	Total	1411	1135	3472	1075	952	183	115	3.11	3.64	3.1	83.88
	Human	525	422	1278	388	339	83	119	3.29	3.77	3.09	80.33
80.0	Total	1411	1135	2188	836	763	372	134.66	2.61	2.86	2.45	67.22
	Human	525	422	776	303	275	147	139.1	2.57	2.82	2.36	65.16
85.0	Total	1411	1135	1492	683	633	502	147.7	2.18	2.35	2.025	55.77
	Human	525	422	553	253	231	191	150.88	2.18	2.39	1.99	54.74

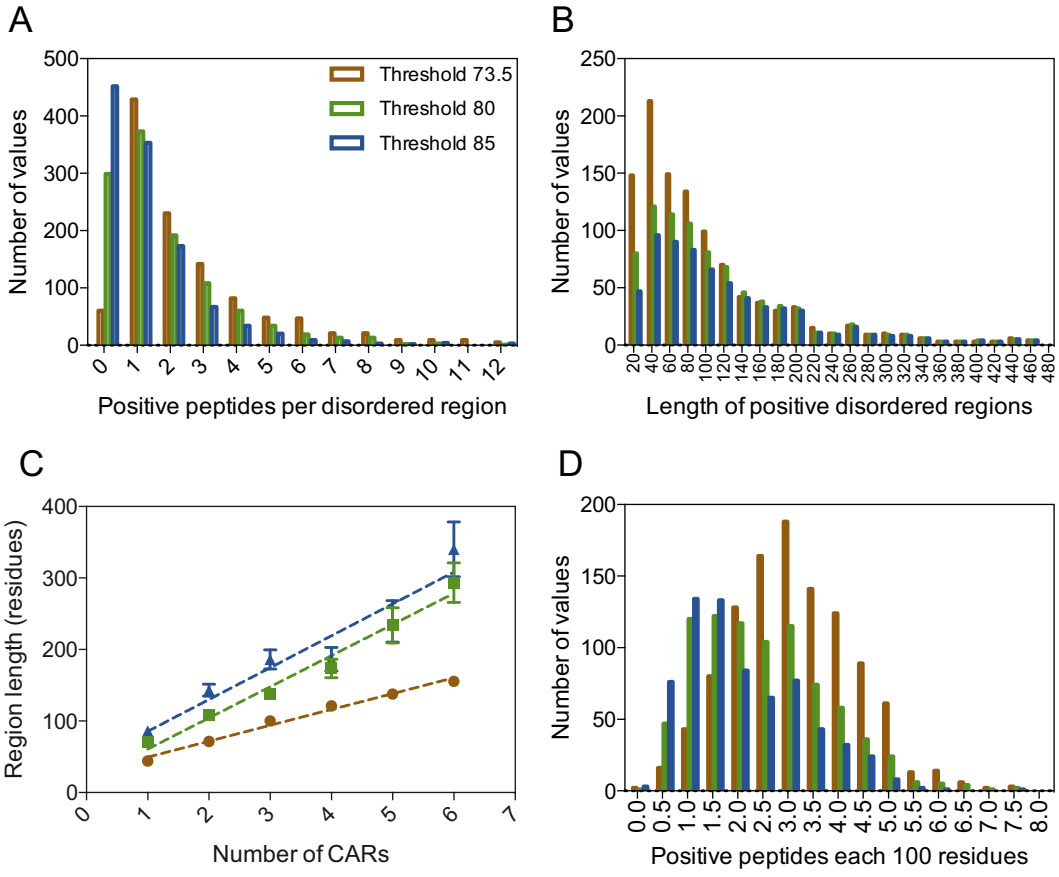

Fig. 1. CARs prevalence and distribution in disordered regions. Disprot was screened with Waltz, considering three thresholds, and positive/regions were defined as those containing at least one CAR. Histograms illustrate the number of positive peptides per disordered region (A) and the length of disordered regions (B). (C) Correlation between the number of CARs and the length of the hosting disordered region. (D) The histogram shows the CARs' density each 100 residues at the different thresholds. Color legend in panel A.

Table 3
Analysis of the net charge of predicted CARs for each of the three analyzed thresholds. The percentage of CARs with anionic, cationic or neutral character are shown. CARs with a net charge of 0 were divided in two classes depending on whether they have charged residues or not. Histidines were not considered as charged at physiological pH.

Threshold	Negatively charged CARs (%)	Positively charged CARs (%)	CARs with compensated charges (%)	CARs without charged residues (%)
73.5	53.02	21.43	14.06	11.49
80.0	51.96	19.52	12.52	16.00
85.0	52.07	17.29	11.93	18.70

excess of charge would necessarily prevent the assembly. Aliphatic and especially aromatic residues are enriched in CARs relative to IDRs, whereas Gly + Pro are significantly underrepresented in CARs. This bias is consistent with the amyloid-promoting character of aromatic residues [70] and the β-breaker properties of Gly and Pro [8,71].

The composition of CARs differs from that of the protein segments identified in the "amylome" by Eisenberg and co-workers using the 3D profile method [72], which were enriched in Val, Ile and Ala and depleted in Asp, Glu and Arg. This result is expected since for their analysis they used complete genomes, including globular and disordered proteins, and they reported that the

Table 4
Compositional biases observed in the CARs with the three thresholds. The compositional bias of the screened disordered regions is also included.

Threshold	Polar (NQTS) %	Ionizable (DEHKR) %	Aliphatic (ILMV) %	Aromatic (FYW) %	β-breakers (GP) %	Rest (AC) %
73.5	27.79	23.01	22.41	8.67	8.68	9.45
80.0	27.56	20.91	23.98	11.27	7.46	8.82
85.0	26.86	19.59	24.95	13.04	6.63	8.93
Disordered regions	23.23	30.79	16.73	5.39	15.08	8.78

identified amyloidogenic regions were 3-times less frequent in disordered regions.

3.4. Cryptic amyloidogenic regions are involved in the establishment of protein–protein interactions.

Amyloidogenicity is assumed to be associated with a detrimental effect on protein homeostasis. Accordingly, it has been shown that evolution tends to purge out amyloid stretches from globular protein sequences, unless they are required for their structure, stability, or biological function [3,12,14,73–75]. Therefore, it seems reasonable to speculate that CARs are allowed within IDRs because they are associated with IDPs function.

To explore the biological relevance of CARs, we first assessed whether CARs-containing proteins were associated with defined biological functions. The high proportion of CARs-positive proteins in DisProt (56 – 84 %) makes it difficult to dissociate potential CAR-mediated roles from those generally associated with IDRs. Despite that, human CARs-containing proteins show moderate enrichments in specific gene ontology categories (p-value < 0.1), compared with the DisProt background (Fig. 2). Enrichments gradually increase as Waltz detection thresholds rise. Our results suggest that this subset of proteins is involved in the establishment of intermolecular interactions, as shown by their enrichment in molecular binding functions (e.g., protein binding, enzyme binding, calmodulin binding), and they participate in biological processes where the formation of complexes is important (i.e., cell adhesion and transcription initiation and regulation). The cellular compartment ontology also reflects their implication in protein complex formation. Of note, negative proteins for the 85-threshold, which correspond to 44 % of the analyzed proteins, were only enriched in growth factor activity (Molecular Function) and positive regulation of tumor necrosis factor production (Biological Process) and not in any of the above mentioned for CAR-positive proteins (not shown). This indicates segregation of protein functions according to the amyloidogenic load of the analyzed IDRs.

The IDPs capacity to establish PPIs is encoded in their primary sequence in the form of modular interaction units, referred to as short linear motifs or linear interacting peptides (LIPs) [76–78]. Given the enrichment of CAR-positive IDPs in binding functions, we speculated that a certain number of CARs might be associated with regions involved in functional PPI, similarly to what has been described for globular proteins [13,14]. To test our hypothesis, we retrieved a database of manually curated LIPS from MobiDB [78] (see Methods section for details), which mapped to 354 unique DisProt regions, of which 285 were at least 21 residues long. 368, 243, and 175 (73.5, 80, and 85 thresholds) of the CARs predicted in the previous section overlapped with these LIPs, with an average of ≈ 90 % of the residues in the overlapping CAR being also part of the LIP (Table 5). Thence, 71.57, 49.47, and 33.05 % of the analyzed LIPs were found to overlap with CARs.

To explore whether this overlap may be an artifact of the high prevalence of CARs in IDR, we assessed the degree of overlap expected by random chance. To that aim, we generated a testing dataset with 285 arbitrary DisProt regions in which we randomly defined 8-residue peptides (the median length of LIPs) and assessed their coincidence with CARs. We found that 57.89, 35.79, and 24.21 % (for the 73.5, 80.0, and 85.0 thresholds, respectively) of these randomly defined peptides overlapped with CARs but only 46.07, 48.98, and 47.3 % of the CARs residues were contained within the LIP. Taken together, these results suggest a significant association at the residue level between LIPs and CARs that cannot be attributed to a spurious overlap. Indeed, according to our analysis, independently of the selected threshold, the probability of any CAR residue to map into a LIP is ≥2.4-fold higher than the one expected by chance.

Overall, these analyses point towards a functional origin of CARs, associated with the facilitation of PPI.

3.5. Protein abundance is connected with IDPs amyloid load in a solubility independent manner

In contrast to protein folding, aggregation and amyloid formation is a second- or higher-order reaction and sensitive to protein concentration [79,80]. This implies that highly abundant proteins with a particular aggregation propensity are more prone to aggregate than low abundant proteins with the same aggregation potential. Indeed, multiple in silico analyses revealed an anti-correlation between gene-expression or protein abundance and aggregation [81–84]. This observation led to the "living on the edge" hypothesis that suggests that protein abundance is tightly regulated to an optimal level, sufficient for proteins to develop their biological function but low enough to maintain their solubility [84]. Proteins exceeding their critical concentration are considered supersaturated and constitute a metastable subproteome, often associated with diseases [85–88].

By analyzing the abundance of human proteins in the datasets of positive and negative proteins for the three Waltz thresholds analyzed (73.5, 80.0, 85.0), we found that CAR-positive proteins tend to be less abundant than negative ones, a trend that correlates with their amyloidogenic potential (Fig. 3A). This is clear when comparing the cumulative frequencies of protein abundance between the positive and negative protein sets with the 85-threshold (Fig. 3B) (p-value = 0.0172, Kolmogorov-Smirnov test). This connection responds to the proteins' amyloid load and not to their global solubility/aggregation propensity since no significant trends could be detected when the same analysis was performed employing CamSol [54] (Fig. 3C and Supplementary Fig. 1) (p-value = 0.9828, Kolmogorov-Smirnov test). These results indicate that the amyloidogenicity of IDRs constrains to some extent the IDPs abundance, which ultimately implies a deleterious side-effect of CARs, whose presence and potency need to be balanced to favor their functional contribution.

3.6. Higher IDRs amyloid loads are associated with disease pathways

Structural disorder is especially abundant in proteins associated with human disease. This feature is not restricted to amyloid-related pathologies, since the majority of proteins associated with cancer and cardiovascular diseases contain long IDRs [89,90]. It is also well known that amyloidogenic regions might establish aberrant interactions that lead to loss of protein functions and/or the formation of toxic assemblies [6,7]. Driven by the widespread presence of CARs in IDRs and their potential deleterious nature, we

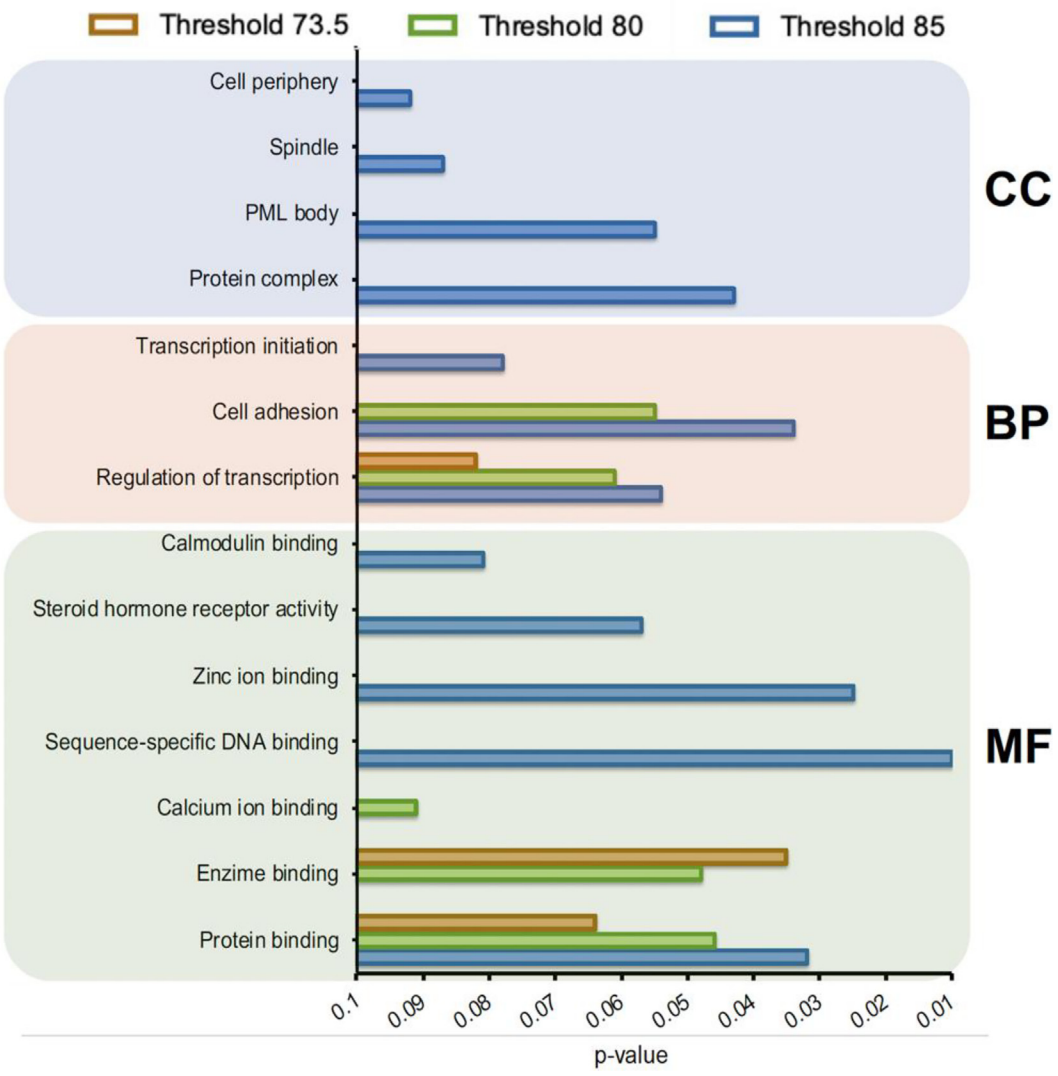

Fig. 2. CAR-positive proteins GO enrichment analysis. GO enrichment was performed and segregated into its three ontologies: cellular component (CC), biological process (BP) and molecular function (MF). Orange, green and blue bars correspond to the 73.5, 80 and 85-threshold datasets respectively (same color legend as in Fig. 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Overlap between predicted CARs and experimentally validated LIPs.

Threshold	DisProt regions > 20 residues with LIPs	Total CARs that overlap with LIPs	Percentage of CAR residues overlapping with LIP (%)	Regions with LIPs and overlapping CARs	Proteins with LIPs and overlapping CARs	LIPs that overlap with CARs (%)
73.5	285	409	89.14	204	192	71.57
80.0	285	243	92.64	141	139	49.47
85.0	285	175	91.86	117	115	33.05

explored if some of the human pathologies associated with protein disorder might be somehow connected to the presence of these amyloid segments.

We first analyzed the disease enrichment of the positive datasets in comparison with the DisProt background. We found five enriched categories for the 85-threshold dataset (Fig. 3D), whereas only Parkinson's disease was enriched in the 80-threshold dataset. The enrichments in Alzheimer's and Parkinson's disease-related proteins are in agreement with the amyloid nature of CARs and include their primary aggregating proteins (APP and α-synuclein). Other proteins implicated in biochemical pathways associated with these conditions were also found to contain CARs and contribute to this enrichment, such as p53, c-Fos, TDP-43, Calpastatin or POU2F1 [91–96].

We also found enrichment in three cancer-associated protein subsets, with the highest association found in breast cancer (Fig. 3D). Previous computational analysis showed that around 79 % of cancer-associated proteins contain IDRs of 30 residues or longer [97]. This association was attributed to the fact that IDRs are often involved in gene regulation, cell-cycle regulation, and molecular recognition, acting as central hubs in signaling networks [89,90]. These functions entail binding to multiple targets and participation in PPI networks which, according to our analysis, would imply the presence of CARs, thus justifying the observed

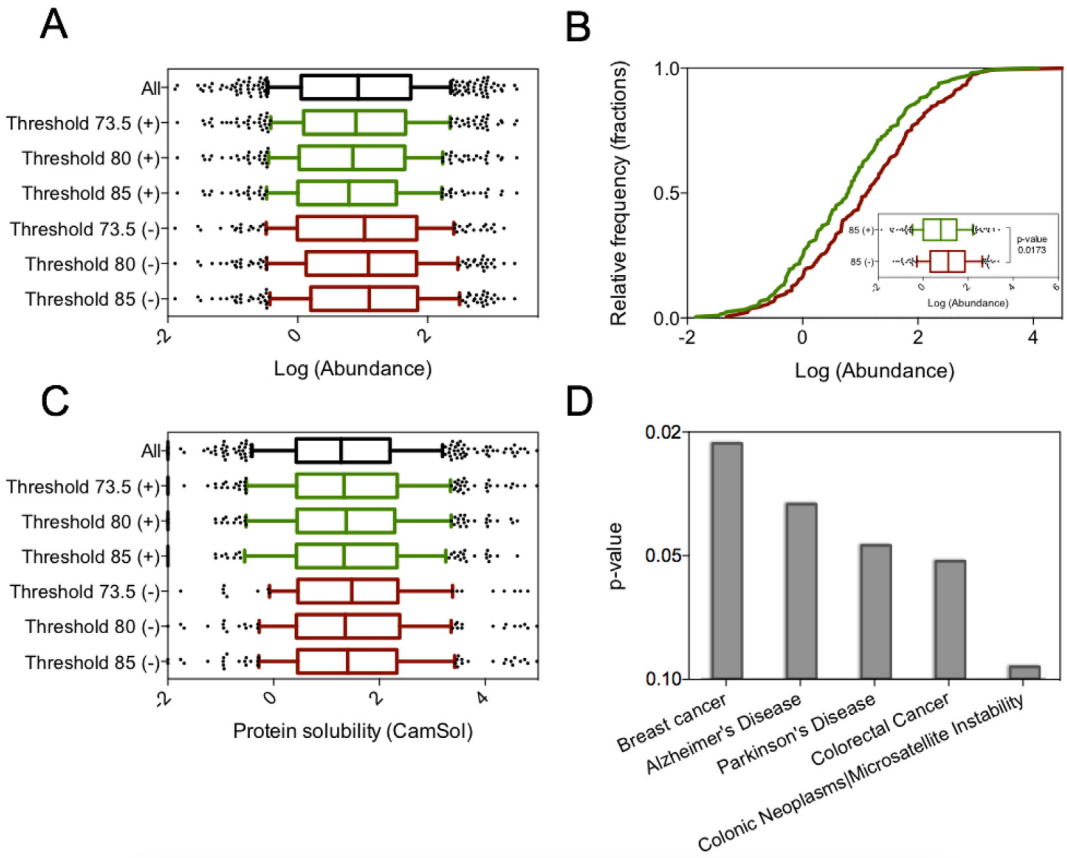

Fig. 3. The presence of CARs constrains protein abundance and is linked with disease pathways in humans. (A) Box-plots depict the abundance of the datasets of CARs positive (+, green) and negative (-, red) proteins for the different thresholds. (B) Cumulative frequency analysis comparing the abundance of positive (+, green) and negative (-, red) proteins for the 85.0 threshold. The inset shows the equivalent box-plot with the statistical p-value (Kolmogorov-Smirnov test). (C) Box-plots comparing the solubility of the different datasets calculated with CamSol. The legend is equivalent to panel (A). (D) Ontology enrichment in disease associated pathways (GAD_disease) of the positive proteins for the 85.0-threshold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

enrichments. Recent reports exploring the role of protein aggregation in tumor formation suggest that it may be an underestimated mechanism of loss of tumor-suppressing activity and oncogenic gain-of-function [98–102].

To get further insight into cancer pathways associated with CAR-positive proteins, we inspected the interaction network of the 85-threshold positive dataset. After a clustering analysis, we identified a central hub comprising 76 proteins densely connected with the surrounding clusters (Supplementary Fig. 2, proteins listed in Supplementary Data 2), which was significantly enriched in cancer-associated ontologies when compared with the complete 85-threshold positive dataset (p-values < 0.001 Supplementary Fig. 3). This hub included several well-known cancer-associated proteins such as apc, erbB-2, c-fos, c-myc, Rb, p53 (Fig. 4 and Table 6). An analysis of this subnetwork revealed that several of these proteins had ontology annotations associated with those found enriched in the previous section; protein binding, gene expression, transcription factor binding (in yellow, green, and blue in Fig. 4). We note that 29 of these proteins were directly associated with cancer pathways (in red in Fig. 4).

Overall, we speculate that CAR-positive proteins hold an amyloidogenic load that makes them intrinsically sensitive to malfunction. Deleterious interactions would be efficiently prevented and reversed in healthy conditions; however, under stress, aging, or disease, the protein quality control machinery functionality is reduced, and aberrant contacts might occur and accumulate. In this scenario, proteins containing CARs would constitute an IDP metastable subproteome susceptible to establish inactivating non-functional interactions. To date, this phenomenon has been primarily linked with protein solubility and aggregation propensities; yet our data suggest a veiled contribution of cryptic amyloidogenicity to it. Of note, our analysis indicates that the previously reported cancer association of IDRs is especially significant in CAR-positive proteins with higher Waltz scores, including a cluster of highly interconnected proteins with some of the most paradigmatic cancer-related proteins.

Our results suggest a three-way connection between the establishment of PPI, the presence of CARs, and the susceptibility to disease in IDPs. The data suggest that CARs are functional regions in IDRs, which, if unregulated, constitute vulnerable spots in the human proteome that might lead to pathology.

3.7. Cryptic amyloidogenic regions hold the potential to assemble into amyloid fibrils

To provide experimental evidence for the above computational data, we decided to characterize experimentally two different CARs. Because of their connection with PPI, we decided to focus on CAR-derived peptides present in experimentally validated LIPs and for which there was direct structural evidence for the interaction (i.e. bounded state present in the PDB). This would allow us to obtain further information on the role of CARs on the interaction and its implication in events of folding-upon-binding.

Due to their relevance in human disorders, specifically in cancer, we selected a first peptide involved in complex formation in

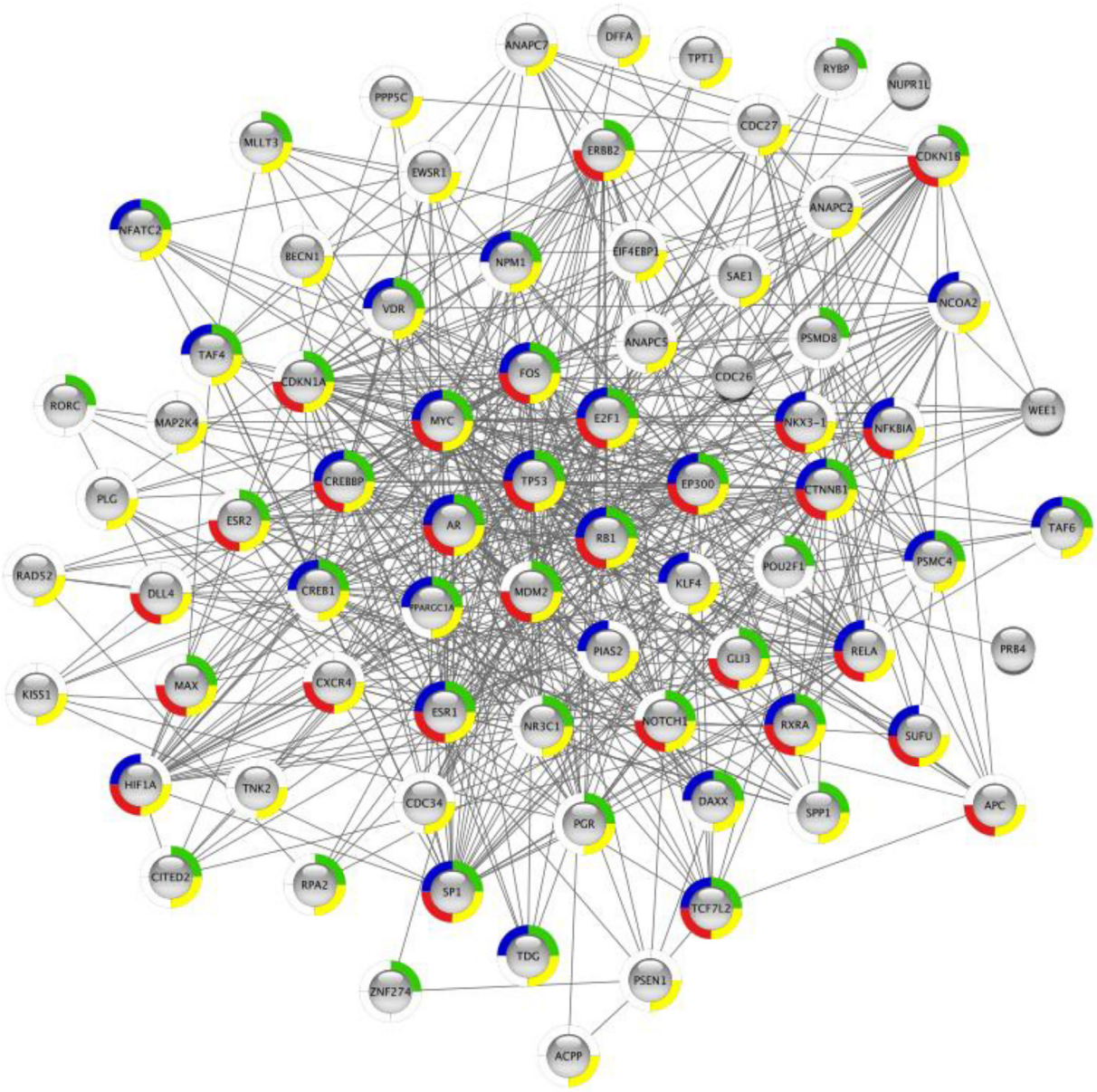

Fig. 4. Protein-protein interaction network of the cancer-enriched central hub of the 85-threshold positive dataset. Color code is associated with the ontology annotation of each protein: red (cancer pathways), yellow (protein binding), green (gene expression), blue (transcription factor binding). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
List of some well-known cancer associated proteins found in the cancer-enriched central hub of the 85-threshold positive dataset.

Gene name	Uniprot ID	Protein name	Pathology	References (PMID)
APC	P25054	Adenomatous polyposis coli protein	Familial adenomatous polyposis 1, Desmoid disease, Medulloblastoma, Gastric cancer, Hepatocellular carcinoma	1651563, 8940264, 10666372, 1338691, 16897741
ERBB2	P04626	Receptor tyrosine-protein kinase erbB-2	Breast, Lung, Ovarian, Gastric and Esophageal cancers	2992089, 15457249, 17471238
FOS	P01100	Proto-oncogene c-Fos	Epithelial ovarian carcinoma, Head and neck squamous cell carcinoma, Osteosarcomas, Hepatocellular carcinomas	18854825, 27965308, 19760594
MYC	P01106	Myc proto-oncogene protein	Burkitt lymphoma, Multiple myeloma, Generic in multiple cancers	6961453, 10618400, 20164920
RB1	P06400	Retinoblastoma-associated protein (Rb)	Childhood cancer retinoblastoma, Generic in multiple cancers	2594029, 12204530
TP53	P04637	Cellular tumor antigen p53	TP53 is frequently mutated or inactivated in about 50% of cancers.	21779514, 17311302, 24212651

the retinoblastoma-associated protein (Rb) and a second peptide localized in the oligomerization domain of p53.

Rb binding to E2F transcription factors represses G1-S transitions, ensuring that cell proliferation strictly depends on growth-factor signals [103]. Deregulation of this pathway is a molecular hallmark in several cancer cases. The C-terminal domain of Rb is unstructured in solution and folds upon binding, forming a structured complex with the E2F1-DP1 heterodimer (Fig. 5A) [103]. A core region of Rb (RbC) comprising residues 829 to 864 is sufficient to mediate this interaction and indispensable for growth suppression. RbC is unstructured in solution but folds upon binding into a strand-loop-helix structure on top of the intermolecular β-sandwich formed by the E2F1-DP1 heterodimer (Fig. 5A).

We predicted a long CAR between residues 831 and 856 (score 84.0) with a shorter core comprising residues 831-ILVSIGESFG-840, that lies just below the default 92 Waltz threshold (score 91.6), and overlaps with the β-strand and loop regions of RbC that interacts with the E2F1-DP1 heterodimer and forms the fifth β-strand of the intermolecular β-sandwich upon binding (in red in Fig. 5A). This β-strand is involved in the formation of an intermolecular antiparallel β-sheet when interacting with E2F1. We analyzed whether a peptide corresponding to this sequence

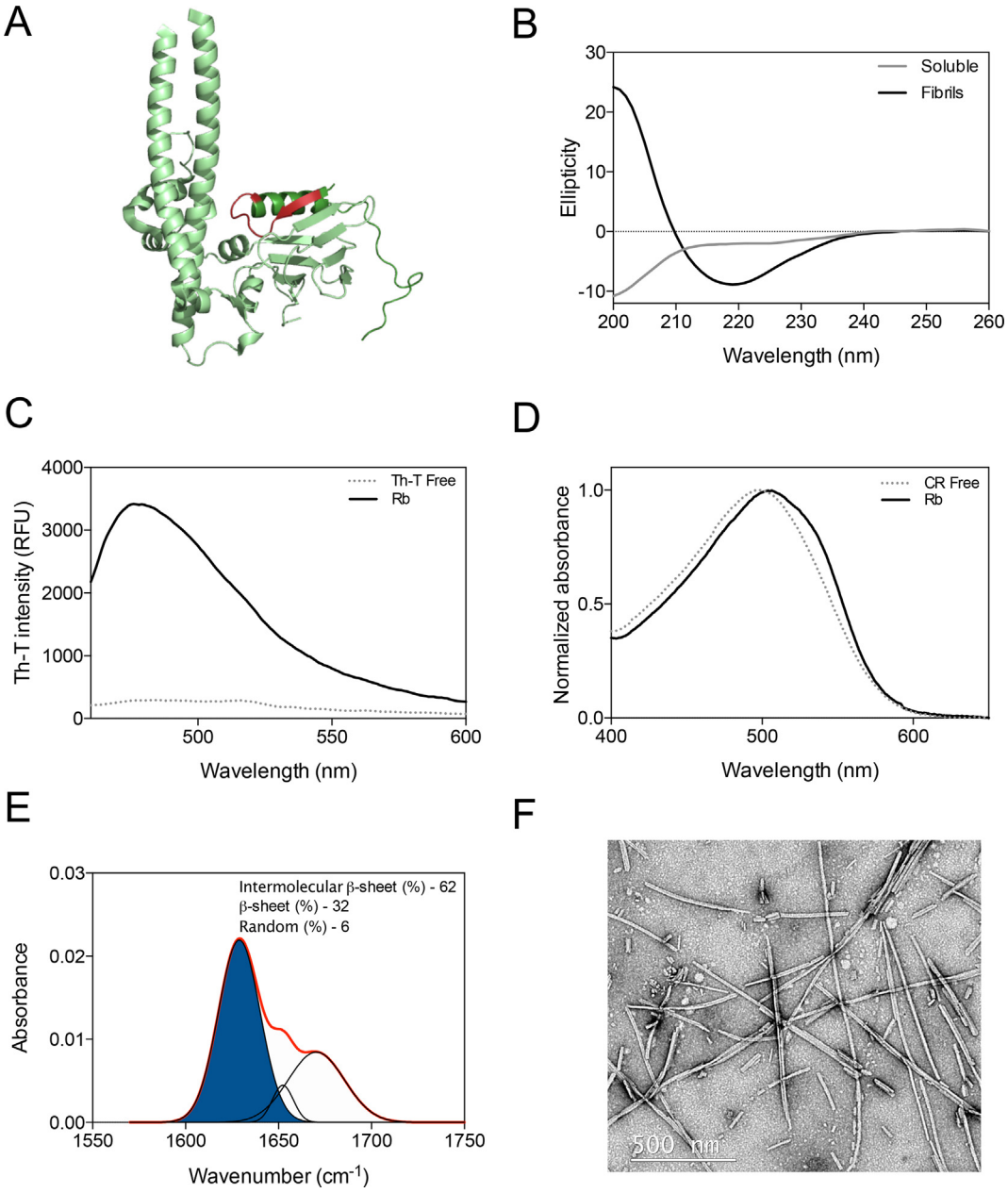

Fig. 5. Experimental validation of a CAR identified in the RbC region (831-ILVSIGESFG-840) of Rb. (A) The RbC region of Rb folds upon binding when is bound to the E2F1-DP1 heterodimer (PDB: 2AZE). Rb polypeptide chain is in dark green and the analyzed peptide is highlighted in red. (B) CD spectra in the far-UV region of 200 μM of the peptide before and after a 2 days incubation at 37 °C with continuous agitation at 100 RPM. (C) Fluorescence emission spectra of Th-T in the absence (dashed line) and presence (solid line) of the incubated peptide sample. (D) CR absorbance spectrum 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. Red line corresponds to the acquired spectra, blue area indicates the band corresponding to the inter-molecular β-sheet contribution to the total area upon Gaussian deconvolution. (F) Representative electron microscopy micrograph of the Rb peptide fibrils. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Ac-ILVSIGESFG-NH2) was able to self-assemble amyloid fibrils. We incubated the peptide at 200 μM, 37 °C, and 100 RPM for 48 h. We observed a disorder to order transition by CD in which the initial intrinsically disordered peptide acquires a characteristic β-sheet signature (Fig. 5B). The incubated peptide was positive for binding to Th-T and CR, suggesting that it self-assembles into amyloid-like structures (Fig. 5C and D). We used FTIR spectroscopy, recording in the amide I region of the spectrum (1700–1600 cm^{-1}), to analyze the assemblies' secondary structure content. The incubated peptide spectrum was dominated by a signal at 1630 cm^{-1} (62 % of the area) attributed to the formation of amyloid-like intermolecular β-sheets (Fig. 5E). The morphological analysis of the sample by transmission electron microscopy (TEM) confirmed the assemblies' fibrillar nature, forming straight and unbranched nanorods that eventually associate laterally (Fig. 5F).

To date, there is no direct experimental evidence for the aggregation of full-length Rb. Nevertheless, Rb's aberrant localization has been reported in Parkinson's disease, including localization to Lewy bodies with a punctate pattern characteristic of aggregation processes [104].

p53 acts as a key regulator of several signaling and cell-fate-decision pathways, being a hub in PPI networks [105–107]. p53 plays a central role in tumor suppression via apoptosis and is mutated in about 50 % of human cancers [50]. p53 can be dissected into three regions: the N-terminal transactivation domain, the DNA binding domain, and the C-terminal domain, in charge of p53 regulation and tetramerization. Both the N- and C-terminal domains are intrinsically disordered and together mediate the 71 % of the p53 interactions [108]. p53 tetramerization is fundamental for site-specific DNA binding and transcription activation [109]. Tetramerization is driven by a short-disordered region comprising residues 320 to 355 that folds upon binding in a tetramer, described as a dimer of dimers (Fig. 6A) [110]. This region is negative for Waltz's using the default 92-threshold but is predicted to contain a CAR of variable length in the 323–334 segment depending on the used threshold. We selected for experimental characterization a short peptide in this region (326-EYFTLQIR-333) with a score of 80.6 that overlaps with the β-strand (327–332) formed upon binding in the p53 tetramer (in red in Fig. 6A). This β-strand is fundamental for dimerization, forming an inter-subunit antiparallel β-sheet.

To assess the peptide's assembling properties (Ac-EYFTLQIR-NH2), we prepared and incubated the p53 peptide in the conditions described for Rb. The CD analysis reveals that before incubation, the peptide already samples two populations, with two minimums at 200 and 216 nm associated with disorder and β-sheet, respectively (Fig. 6B). This β-sheet contribution may result from peptide dimerization, as it occurs in the tetrameric structure (Fig. 6A). In the incubated peptide, we observed a negative band at ≈ 225 nm. The red-shift of the β-sheet band has been observed previously [111] and seems to be associated to the fibrils' particular architecture. The incubated peptide was positive for Th-T and CR amyloid dyes suggesting that it aggregates into amyloid-like assemblies (Fig. 6C and D). Deconvolution of the FTIR spectra revealed a strong band at 1630 cm^{-1} (58 % of the area), indicating the presence of intermolecular β-sheet structures in the incubated peptide (Fig. 5E). TEM micrographs confirmed that the p53 peptide self-assembles into fibrillar supramolecular structures (Fig. 5F). These fibrillar aggregates are straight and unbranched with a tendency to establish lateral associations.

Remarkably, the complete tetramerization domain of p53 has already been reported to aggregate into amyloid fibrils [112], a property that is enhanced by some cancer-associated mutations [113]. p53 amyloid fibrils have been detected in Alzheimer's and cancer patients [114–116], and induction of intracellular p53 amyloids results in cellular transformation and tumor induction in mice [99].

4. Discussion

The development of predictive tools with the capacity to analyze large protein datasets has been of paramount importance to build up our knowledge on the prevalence and implications of amyloidogenic sequences at the proteome level [59]. However, most computational analyses have focused on screening amyloidogenic regions that are too strong to be permitted in solvent-exposed regions and tend to be buried in the core of globular proteins [8,9,12]. Milder amyloidogenic regions -here named CARs- intrinsically exposed to solvent in IDRs went unnoticed in these analyses, essentially due to their polar nature. Thus, our current understanding of amyloid biology is biased since it results from evaluating only a fraction of the amyloid sequence space. This study finds that a large proportion of IDRs bear amyloidogenic regions cryptically encoded in their primary sequences.

We have shown that CARs are linked with function and are consistently found in regions involved in PPI, suggesting that they are the subject of a positive selection in association with defined biological functions. Conversely, the data support a negative pressure constraining the abundance of CAR-containing proteins and controlling CARs prevalence in IDRs. This observation suggests a deleterious contribution of these regions to physiology that might explain the association of IDRs bearing CARs with cancer, Alzheimer's, and Parkinson's diseases. Thus, it appears that the evolution of IDRs functions, especially those involving PPIs, entails the concomitant appearance and fixation of amyloidogenic regions that, as a side effect, put them at risk of non-functional interactions, aggregation, and malfunction. Thus, CARs can either facilitate the switch of disordered conformations towards the formation of functional complexes or mis-functional assemblies, depending on the cellular context. The present study expands the "living on the edge" hypothesis to the area of IDPs. It highlights how the cryptic amyloid propensity of disordered regions should be considered when evaluating the interplay between amyloidogenicity, function, and protein evolution.

Driven by these observations, we characterized in vitro two of the identified CARs and confirmed their amyloid nature. Notably, both segments play a significant role in establishing protein–protein interactions, central to the respective proteins' biological function. The Rb CAR mediates a heterotypic interaction with the E2F1-DP1 complex on which depends Rb activity (Fig. 5A). Of note, E2F1 and DP1 are disordered when isolated, and E2F1 also contains a predicted CAR (278-SSENFQISLKS-288) in the same intermolecular β-sandwich where RbC binds (Supplementary Fig. 4). The characterized p53 amyloid peptide is a fundamental element for homodimerization, which folds upon binding into a β-strand and forming an antiparallel double-stranded sheet with the other monomer (Fig. 6A). Mutations that map at this region significantly impact the tetramer stability, compromising p53 function and being associated with different cancers [110]. In the two studied cases, the oligomerization of a disordered region displaying a cryptic amyloid nature is critical for the functional structured complexes' constitution. In our view, they constitute two illustrative examples explaining why CARs are likely to be persistent in IDRs, the ultimate reason being that PPI and amyloidogenicity sequence codes overlap significantly. They also exemplify how CARs expand the range of protein sequences compatible with an amyloid fold to those with low intrinsic structural propensities.

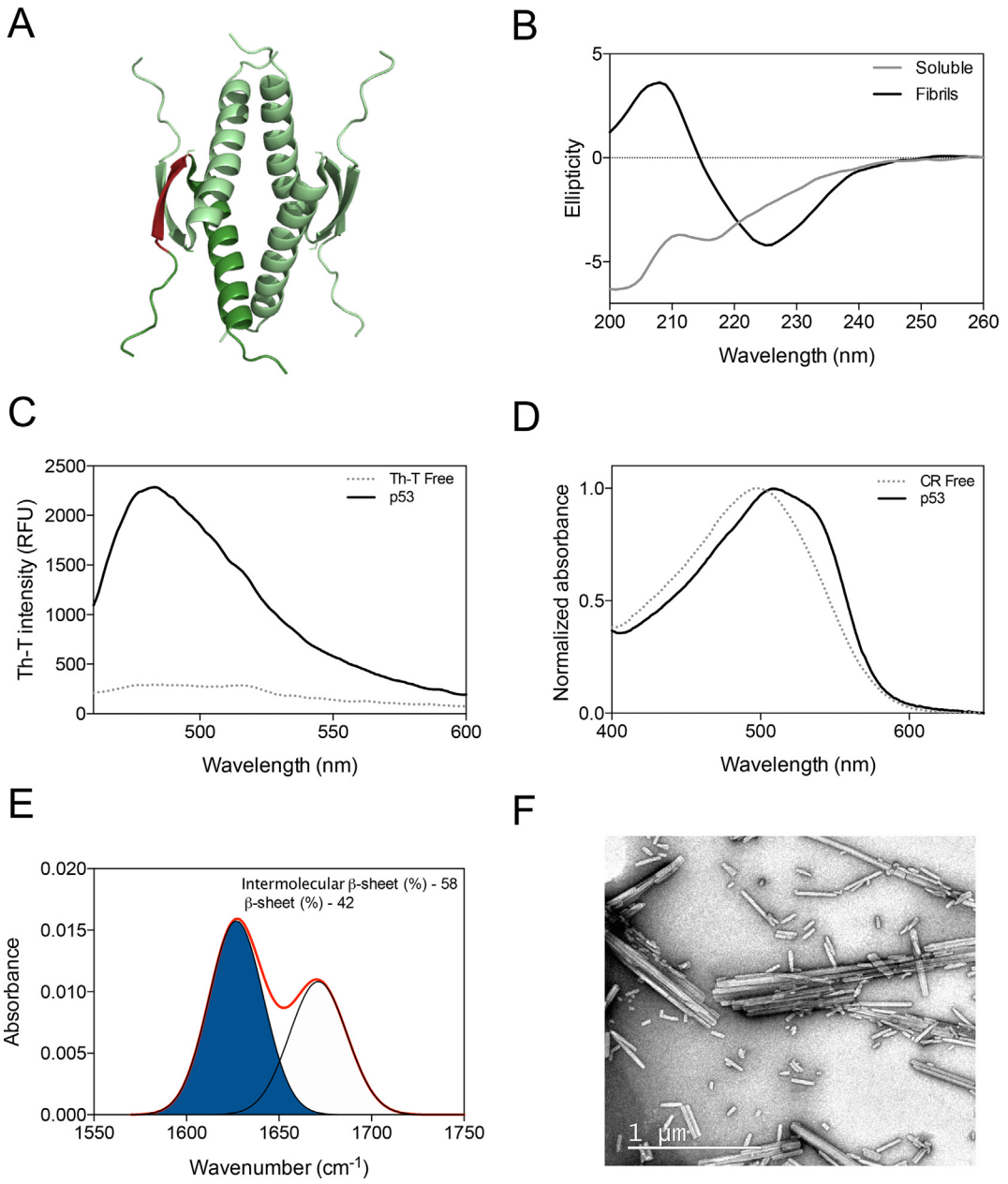

Fig. 6. Experimental validation of a CAR identified in the tetramerization domain (326-EYFTLQIR-333) of p53. (A) Upon oligomerization the p53 tetramerization domain folds in a strand-loop-helix conformation (PDB: 3SAK). The analyzed peptide is highlighted in red. (B) CD spectra in the far-UV region of 200 μM of the peptide before and after a 2 days incubation at 37 °C with continuous agitation at 100 RPM. (C) Fluorescence emission spectra of Th-T in the absence (dashed line) and presence (solid line) of the incubated peptide sample. (D) CR absorbance spectrum 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. Red line corresponds to the acquired spectra, blue area indicates the band corresponding to the inter-molecular β-sheet contribution to the total area upon Gaussian deconvolution. (F) Representative electron microscopy micrograph of the p53 peptide fibrils. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

According to the amyloid-driven origin of life hypothesis, short amyloid peptides are candidates to be the first functional protein folds in the prebiotic era and the first molecules able to self-replicate, transmit information and catalyze reactions [26–28]. Based on this hypothesis, it has been proposed that globular proteins may have gradually emerged from these early amyloidogenic peptides, thus justifying the thermodynamic coupling of amyloid and globular structures as a relict of this evolution [12]. Nevertheless, this structural leap is still poorly understood. In light of our results, we propose that a first evolutive step in the prebiotic amyloid-world would be the conversion of short unstructured amyloidogenic peptides into interacting regions that allow for the establishment of functional homotypic and heterotypic interactions. These contacts would have been instrumental for forming the first globular complexes, establishing cooperative networks, and early coacervation. In a way, the early globular assemblies could be a kind of functional amyloid oligomers. The cryptic amyloidogenicity identified in this study might well be a reminiscence of such an evolutionary pathway.

5. Conclusions

By expanding previous sequence amyloid propensity large-scale analysis to the realm of disordered proteins, we provide a novel perspective on the role of polypeptides' intrinsic amyloidogenicity

in protein function, malfunction, and evolution that complements and reinforces the conceptual framework established for globular proteins. Overall, it appears that the propensities of amyloidogenic regions in present proteins depend on both their structural and functional contexts, their ubiquitous presence responding to the ancestral need of proteins for intra- and intermolecular interactions to exert their biological functions.

CRediT authorship contribution statement

Jaime Santos: Conceptualization, Investigation, Formal analysis, Writing - review & editing. **Irantzu Pallarès:** Writing - review & editing, Visualization, Supervision. **Valentín Iglesias:** Software. **Salvador Ventura:** Conceptualization, Supervision, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) BIO2016-78310-R to S.V and by ICREA, ICREA-Academia 2015 and 2020 to S.V. J.S. was supported by the Spanish Ministry of Science and Innovation via a doctoral grant (FPU17/01157). We thank the members of the Microscopy Services of the UAB for their assistance. We also thank Salvador Bartolome from the Laboratori de Luminiscència i Espectroscòpia de Biomolècules, UAB, for insightful discussion.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.07.019.

References

[1] Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science 2012;338(6110):1042–6.
[2] Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. Annu Rev Biochem 2017;86(1):27–68.
[3] Pastore A, Temussi PA. The two faces of Janus: functional interactions and protein aggregation. Curr Opin Struct Biol 2012;22(1):30–7.
[4] Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc Natl Acad Sci U S A 2009;106 (25):10159–64.
[5] Temussi PA, Tartaglia GG, Pastore A. The seesaw between normal function and protein aggregation: How functional interactions may increase protein solubility. BioEssays 2021;43:e2100031.
[6] Winklhofer KF, Tatzelt J, Haass C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. EMBO J 2008;27 (2):336–49.
[7] Dobson CM, Knowles TPJ, Vendruscolo M (2020) The Amyloid Phenomenon and Its Significance in Biology and Medicine. Cold Spring Harb Perspect Biol 12.
[8] Rousseau F, Serrano L, Schymkowitz JWH. How evolutionary pressure against protein aggregation shaped chaperone specificity. J Mol Biol 2006;355 (5):1037–47.
[9] De Baets G, Reumers J, Delgado Blanco J, Dopazo J, Schymkowitz J, Rousseau F, et al. An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. PLoS Comput Biol 2011;7(6):e1002090.
[10] Castillo V, Espargaro A, Gordo V, Vendrell J, Ventura S. Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria. Proteomics 2010;10:4172–85.
[11] Fraga H, Graña-Montes R, Illa R, Covaleda G, Ventura S. Association between foldability and aggregation propensity in small disulfide-rich proteins. Antioxid Redox Signal 2014;21(3):368–83.
[12] Langenberg T, Gallardo R, van der Kant R, Louros N, Michiels E, Duran-Romaña R, et al. Thermodynamic and evolutionary coupling between the native and amyloid state of globular proteins. Cell Rep 2020;31(2):107512. https://doi.org/10.1016/j.celrep.2020.03.076.
[13] Masino L, Nicastro G, Calder L, Vendruscolo M, Pastore A. Functional interactions as a survival strategy against abnormal aggregation. FASEB J 2011;25(1):45–54.
[14] Castillo V, Ventura S, Nussinov R. Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. PLoS Comput Biol 2009;5(8):e1000476.
[15] Buck PM, Kumar S, Singh SK, Iakoucheva LM. On the role of aggregation prone regions in protein evolution, stability, and enzymatic catalysis: insights from diverse analyses. PLoS Comput Biol 2013;9(10):e1003291.
[16] Grana-Montes R, Sant'anna de Oliveira R, Ventura S. Protein aggregation profile of the human kinome. Front Physiol 2012;3:438.
[17] Ganesan A, Siekierska A, Beerten J, Brams M, Van Durme J, De Baets G, et al. Structural hot spots for the solubility of globular proteins. Nat Commun 2016;7(1). https://doi.org/10.1038/ncomms10816.
[18] Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol 2004;342:345–53.
[19] Dyson HJ. Making sense of intrinsically disordered proteins. Biophys J 2016;110(5):1013–6.
[20] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 2000;41(3):415–27.
[21] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 2015;16(1):18–29.
[22] Díaz-Caballero M, Navarro S, Fuentes I, Teixidor F, Ventura S. Minimalist prion-inspired polar self-assembling peptides. ACS Nano 2018;12 (6):5394–407.
[23] Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, et al. Atomic structures of low-complexity protein segments reveal kinked beta sheets that assemble networks. Science 2018;359:698–701.
[24] Louros N, Orlando G, De Vleeschouwer M, Rousseau F, Schymkowitz J. Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities. Nat Commun 2020;11:3314.
[25] Greenwald J, Kwiatkowski W, Riek R. Peptide amyloids in the origin of life. J Mol Biol 2018;430(20):3735–50.
[26] Greenwald J, Riek R. On the possible amyloid origin of protein folds. J Mol Biol 2012;421(4-5):417–26.
[27] Maury CPJ. Amyloid and the origin of life: self-replicating catalytic amyloids as prebiotic informational and protometabolic entities. Cell Mol Life Sci 2018;75(9):1499–507.
[28] Greenwald J, Friedmann MP, Riek R. Amyloid aggregates arise from amino acid condensations under prebiotic conditions. Angew Chem Int Ed Engl 2016;55(38):11609–13.
[29] Hatos A, Hajdu-Soltesz B, Monzon AM, Palopoli N, Alvarez L, et al. DisProt: intrinsic protein disorder annotation in 2020. Nucl Acids Res 2020;48: D269–76.
[30] Chandonia JM, Fox NK, Brenner SE (2019) SCOPe: classification of large macromolecular structures in the structural classification of proteins-extended database. Nucleic Acids Res 47: D475-D481.
[31] Necci M, Piovesan D, Clementel D, Dosztanyi Z, Tosatto SCE. MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavours in proteins. Bioinformatics 2020.
[32] Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, et al. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science 1991;253(5020):665–9.
[33] Eccles DM, van der Luijt R, Breukel C, Bullman H, Bunyan D, et al. Hereditary desmoid disease due to a frameshift mutation at codon 1924 of the APC gene. Am J Hum Genet 1996;59:1193–201.
[34] Huang H, Mahler-Araujo BM, Sankila A, Chimelli L, Yonekawa Y, Kleihues P, et al. APC mutations in sporadic medulloblastomas. Am J Pathol 2000;156 (2):433–7.
[35] Nakatsuru S, Yanagisawa A, Ichii S, Tahara E, Kato Yo, Nakamura Y, et al. Somatic mutation of the APC gene in gastric cancer: frequent mutations in very well differentiated adenocarcinoma and signet-ring cell carcinoma. Hum Mol Genet 1992;1(8):559–63.
[36] Katoh H, Shibata T, Kokubu A, Ojima H, Kosuge T, Kanai Y, et al. Genetic inactivation of the APC gene contributes to the malignant progression of sporadic hepatocellular carcinoma: a case report. Genes Chromosomes Cancer 2006;45(11):1050–7.
[37] King C, Kraus M, Aaronson S. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. Science 1985;229(4717):974–6.
[38] Stephens P, Hunter C, Bignell G, Edkins S, Davies H, Teague J, et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. Nature 2004;431 (7008):525–6.
[39] Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene 2007;26(45):6469–87.
[40] Mahner S, Baasch C, Schwarz J, Hein S, Wölber L, Jänicke F, et al. C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. Br J Cancer 2008;99(8):1269–75.
[41] Muhammad N, Bhattacharya S, Steele R, Phillips N, Ray RB. Involvement of c-Fos in the promotion of cancer stem-like cell properties in head and neck squamous cell carcinoma. Clin Cancer Res 2017;23(12):3120–8.

[42] Durchdewald M, Angel P, Hess J. The transcription factor Fos: a Janus-type regulator in health and disease. Histol Histopathol 2009;24:1451–61.
[43] Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc Natl Acad Sci U S A 1982;79(24):7824–7.
[44] Shou Y, Martelli ML, Gabrea A, Qi Y, Brents LA, Roschke A, et al. Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor progression in multiple myeloma. Proc Natl Acad Sci U S A 2000;97(1):228–33.
[45] Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463(7283):899–905.
[46] Yandell DW, Campbell TA, Dayton SH, Petersen R, Walton D, Little JB, et al. Oncogenic point mutations in the human retinoblastoma gene: their application to genetic counseling. N Engl J Med 1989;321(25):1689–95.
[47] Sherr CJ, McCormick F. The RB and p53 pathways in cancer. Cancer Cell 2002;2(2):103–12.
[48] Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis. Genes Cancer 2011;2(4):466–74.
[49] Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat 2007;28(6):622–9.
[50] Ozaki T, Nakagawara A. Role of p53 in cell death and human cancers. Cancers (Basel) 2011;3(1):994–1013.
[51] Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, Reumers J, et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 2010;7(3):237–42.
[52] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4 (1):44–57.
[53] Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics 2012;11(8):492–500.
[54] Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol 2015;427(2):478–90.
[55] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S et al. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47: D607-D613.
[56] Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003;424(6950):805–8.
[57] Espargaró A, Castillo V, de Groot NS, Ventura S. The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: the SH3 case. J Mol Biol 2008;378(5):1116–31.
[58] Santos J, Pujols J, Pallarès I, Iglesias V, Ventura S. Computational prediction of protein aggregation: advances in proteomics, conformation-specific algorithms and biotechnological applications. Comput Struct Biotechnol J 2020;18:1403–13.
[59] Santos J, Iglesias V, Ventura S. Computational prediction and redesign of aberrant protein oligomerization. Prog Mol Biol Transl Sci 2020;169:43–83.
[60] Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinf 2007;8(1):65. https://doi.org/10.1186/1471-2105-8-65.
[61] Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004;22(10):1302–6.
[62] Sanchez de Groot N, Pallares I, Aviles FX, Vendrell J, Ventura S. Prediction of "hot spots" of aggregation in disease-linked polypeptides. BMC Struct Biol 2005;5:18.
[63] Tenidis K, Waldner M, Bernhagen J, Fischle W, Bergmann M, Weber M, et al. Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. J Mol Biol 2000;295 (4):1055–71.
[64] Seidler PM, Boyer DR, Rodriguez JA, Sawaya MR, Cascio D, Murray K, et al. Structure-based inhibitors of tau aggregation. Nat Chem 2018;10(2):170–6.
[65] Tsiolaki PL, Louros NN, Hamodrakas SJ, Iconomidou VA. Exploring the 'aggregation-prone' core of human Cystatin C: a structural study. J Struct Biol 2015;191(3):272–80.
[66] Piovesan D, Tabaro F, Micetic I, Necci M, Quaglia F, et al. DisProt 7.0: a major update of the database of disordered proteins. Nucl Acids Res 2017;45: D219–27.
[67] Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, et al. What's in a name? Why these proteins are intrinsically disordered: why these proteins are intrinsically disordered. Intrinsically Disord Proteins 2013;1(1): e24157. https://doi.org/10.4161/idp.24157.
[68] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A 2013;110(33):13392–7.
[69] Houben B, Michiels E, Ramakers M, Konstantoulea K, Louros N, et al. Autonomous aggregation suppression by acidic residues explains why chaperones favour basic residues. EMBO J 2020;39:e102864.
[70] Bemporad F, Taddei N, Stefani M, Chiti F. Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase. Protein Sci 2006;15:862–70.
[71] Parrini C, Taddei N, Ramazzotti M, Degl'Innocenti D, Ramponi G, Dobson CM, et al. Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. Structure 2005;13(8):1143–51.
[72] Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A 2010;107(8):3487–92.
[73] Monsellier E, Ramazzotti M, Taddei N, Chiti F, Nussinov R. Aggregation propensity of the human proteome. PLoS Comput Biol 2008;4(10): e1000199.
[74] de Groot NS, Ventura S. Protein aggregation profile of the bacterial cytosol. PLoS ONE 2010;5:e9383.
[75] Castillo V, Chiti F, Ventura S, Rezaei H. The N-terminal helix controls the transition between the soluble and amyloid states of an FF domain. PLoS ONE 2013;8(3):e58297.
[76] Pancsa R, Fuxreiter M. Interactions via intrinsically disordered regions: what kind of motifs? IUBMB Life 2012;64:513–20.
[77] Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. Mol Biosyst 2012;8(1):268–81.
[78] Piovesan D, Necci M, Escobedo N, Monzon AM, Hatos A, et al. MobiDB: intrinsically disordered proteins in 2021. Nucl Acids Res 2021;49:D361–7.
[79] Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci U S A 2012;109 (50):20461–6.
[80] Meisl G, Kirkegaard JB, Arosio P, Michaels TCT, Vendruscolo M, Dobson CM, et al. Molecular mechanisms of protein aggregation from global fitting of kinetic models. Nat Protoc 2016;11(2):252–72.
[81] Tartaglia GG, Vendruscolo M. Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. Mol Biosyst 2009;5(12):1873. https://doi.org/10.1039/b913099n.
[82] Vecchi G, Sormanni P, Mannini B, Vandelli A, Tartaglia GG, Dobson CM, et al. Proteome-wide observation of the phenomenon of life on the edge of solubility. Proc Natl Acad Sci U S A 2020;117(2):1015–20.
[83] Castillo V, Graña-Montes R, Ventura S. The aggregation properties of Escherichia coli proteins associated with their cellular abundance. Biotechnol J 2011;6(6):752–60.
[84] Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem Sci 2007;32(5):204–6.
[85] Ciryam P, Kundra R, Morimoto RI, Dobson CM, Vendruscolo M. Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. Trends Pharmacol Sci 2015;36(2):72–7.
[86] Ciryam P, Tartaglia G, Morimoto R, Dobson C, Vendruscolo M. Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. Cell Rep 2013;5(3):781–90.
[87] Ciryam P, Antalek M, Cid F, Tartaglia GG, Dobson CM, Guettsches A-K, et al. A metastable subproteome underlies inclusion formation in muscle proteinopathies. Acta Neuropathol Commun 2019;7:197.
[88] Ciryam P, Kundra R, Freer R, Morimoto RI, Dobson CM, Vendruscolo M. A transcriptional signature of Alzheimer's disease is associated with a metastable subproteome at risk for aggregation. Proc Natl Acad Sci U S A 2016;113(17):4753–8.
[89] Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, et al. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 2009;10(Suppl 1):S7. https://doi.org/10.1186/1471-2164-10-S1-S7.
[90] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008;37 (1):215–46.
[91] Szybińska A, Leśniakx W. P53 dysfunction in neurodegenerative diseases – the cause or effect of pathological changes? Aging Dis 2017;8(4):506. https://doi.org/10.14336/AD.2016.1120.
[92] Gillardon F, Skutella T, Uhlmann E, Holsboer F, Zimmermann M, et al. Activation of c-Fos contributes to amyloid beta-peptide-induced neurotoxicity. Brain Res 1996;706:169–72.
[93] McAleese KE, Walker L, Erskine D, Thomas AJ, McKeith IG, Attems J. TDP-43 pathology in Alzheimer's disease, dementia with Lewy bodies and ageing. Brain Pathol 2017;27(4):472–9.
[94] Zhu L, Gong Li, Yang T, Xiao X, Guo Y. Calpastatin mediates development of Alzheimer's disease in diabetes. J Alzheimers Dis 2019;68(3):1051–9.
[95] Diepenbroek M, Casadei N, Esmer H, Saido TC, Takano J, et al. Overexpression of the calpain-specific inhibitor calpastatin reduces human alpha-Synuclein processing, aggregation and synaptic impairment in [A30P]alphaSyn transgenic mice. Hum Mol Genet 2014;23:3975–89.
[96] Taguchi K, Yamagata HD, Zhong W, Kamino K, Akatsu H, Hata R, et al. Identification of hippocampus-related candidate genes for Alzheimer's disease. Ann Neurol 2005;57(4):585–8.
[97] Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002;323 (3):573–84.
[98] Xu J, Reumers J, Couceiro JR, De Smet F, Gallardo R, Rudyak S, et al. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. Nat Chem Biol 2011;7(5):285–95.

[99] Navalkar A, Pandey S, Singh N, Patel K, Datta D, et al. Direct evidence of cellular transformation by prion-like p53 amyloid infection. J Cell Sci 2021;134(11). https://doi.org/10.1242/jcs.258316.
[100] De Smet F, Saiz Rubio M, Hompes D, Naus E, De Baets G, Langenberg T, et al. Nuclear inclusion bodies of mutant and wild-type p53 in cancer: a hallmark of p53 inactivation and proteostasis remodelling by p53 aggregation. J Pathol 2017;242(1):24–38.
[101] Anvarian Z, Nojima H, van Kappel EC, Madl T, Spit M, Viertler M, et al. Axin cancer mutants form nanoaggregates to rewire the Wnt signaling network. Nat Struct Mol Biol 2016;23(4):324–32.
[102] Claes F, Maritschnegg E, De Baets G, Siekierska A, Rubio MS, et al. The tumor suppressor protein PTEN undergoes amyloid-like aggregation in tumor cells. BioRxiv 2020. 2020.2011.2030.402115.
[103] Rubin SM, Gall A-L, Zheng N, Pavletich NP. Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release. Cell 2005;123(6):1093–106.
[104] Jordan-Sciutto KL, Dorsey R, Chalovich EM, Hammond RR, Achim CL. Expression patterns of retinoblastoma protein in Parkinson disease. J Neuropathol Exp Neurol 2003;62(1):68–74.
[105] Levine AJ, Oren M. The first 30 years of p53: growing ever more complex. Nat Rev Cancer 2009;9(10):749–58.
[106] Collavin L, Lunardi A, Del Sal G. p53-family proteins and their regulators: hubs and spokes in tumor suppression. Cell Death Differ 2010;17(6):901–11.
[107] Zhao R, Gish K, Murphy M, Yin Y, Notterman D, et al. Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. Genes Dev 2000;14:981–93.
[108] Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 2008;9(Suppl 1):S1. https://doi.org/10.1186/1471-2164-9-S1-S1.
[109] Pietenpol JA, Tokino T, Thiagalingam S, el-Deiry WS, Kinzler KW, Vogelstein B. Sequence-specific transcriptional activation is essential for growth suppression by p53. Proc Natl Acad Sci U S A 1994;91(6):1998–2002.
[110] Chène P. The role of tetramerization in p53 function. Oncogene 2001;20 (21):2611–7.
[111] Iyer A, Roeters SJ, Kogan V, Woutersen S, Claessens M, et al. C-terminal truncated alpha-synuclein fibrils contain strongly twisted beta-sheets. J Am Chem Soc 2017;139:15392–400.
[112] Lee AS, Galea C, DiGiammarino EL, Jun B, Murti G, Ribeiro RC, et al. Reversible amyloid formation by the p53 tetramerization domain and a cancer-associated mutant. J Mol Biol 2003;327(3):699–709.
[113] Higashimoto Y, Asanomi Y, Takakusagi S, Lewis MS, Uosaki K, Durell SR, et al. Unfolding, aggregation, and amyloid formation by the tetramerization domain from mutant p53 associated with lung cancer. Biochemistry 2006;45(6):1608–19.
[114] Farmer KM, Ghag G, Puangmalai N, Montalbano M, Bhatt N, Kayed R. P53 aggregation, interactions with tau, and impaired DNA damage response in Alzheimer's disease. Acta Neuropathol Commun 2020;8(1). https://doi.org/10.1186/s40478-020-01012-6.
[115] Lasagna-Reeves CA, Clos AL, Castillo-Carranza D, Sengupta U, Guerrero-Muñoz M, Kelly B, et al. Dual role of p53 amyloid formation in cancer; loss of function and gain of toxicity. Biochem Biophys Res Commun 2013;430 (3):963–8.
[116] Ano Bom APD, Rangel LP, Costa DCF, de Oliveira GAP, Sanches D, Braga CA, et al. Mutant p53 aggregates into prion-like amyloid oligomers and fibrils: implications for cancer. J Biol Chem 2012;287(33):28152–62.