A Computational Approach Reveals the Ability of Amyloids to Sequester RNA: the Alpha Synuclein Case

Jakob Rupert[1,2,*], Michele Monti[1,*], Elsa Zacco[1] and Gian Gaetano Tartaglia[1,3,**]

1 Centre for Human Technologies (CHT), Istituto Italiano di Tecnologia (IIT), Via Enrico Melen, 83, 16152, Genova, Italy
2 Department of Biology 'Charles Darwin', Sapienza University of Rome, P.le A. Moro 5, Rome, 00185, Italy
3 Catalan Institution for Research and Advanced Studies, ICREA, Passeig Lluís Companys 23, 08010, Barcelona, Spain

* contributed equally
**correspondence to gian@tartaglialab.com

ABSTRACT

Nucleic acids can act as potent modulators of protein aggregation, and RNA is able to either hinder or facilitate protein assembly depending on the molecular context. Here we used a computational approach to characterize the physico-chemical properties of regions involved in amyloid aggregation. In different experimental datasets we observed that, while the core is hydrophobic and highly ordered, external regions, more disordered, display a distinct tendency to interact with nucleic acids.

To validate our predictions, we performed aggregation assays with α-synuclein (aS140), a non-nucleic acid binding amyloidogenic protein, and a mutant truncated at the acidic C-terminus (aS103) that is predicted to sequester RNA. For both aS140 and aS103 we observed acceleration of the aggregation upon RNA addition with a significantly stronger effect for aS103. Due to the favorable electrostatics, we observed enhanced nucleic-acid sequestration ability for aS103 that entrapped a larger amount of RNA.

Overall, our research suggests that RNA sequestration is a rather common phenomenon linked to protein aggregation and constitutes a gain-of-function mechanism to be further investigated.

STATEMENT OF SIGNIFICANCE

Our study indicates that aggregation confers RNA-binding ability to non-RNA-binding proteins such as alpha synuclein. The sequestration of RNA upon protein aggregation might alter RNA homeostasis and impact multiple biochemical cascades.

INTRODUCTION

The order and structural robustness of amyloids render them a powerful platform for attracting other cellular factors and macromolecules[1,2]. Nucleic acids are embroiled in the modulation of protein phase transitions[3,4], as recent experiments show that they can be isolated from both physiological condensates and pathological inclusions[5,6]. RNA in particular might affect aggregate accumulation rate and nature[7]. In support of this observation, numerous proteins found aggregated in neurological disorders are RNA-binding and form ribonucleoprotein assemblies[4], and their ability to aggregate appears to be at least in part determined by the type and amount of RNA found in their proximity[1]. Yet, only a few studies have focused on understanding the ability of proteins to interact with nucleic acids after the transition towards amyloid species has taken place, or whether such transition might confer the ability to attract nucleic acids to proteins not classified as RNA/DNA-binding[8].

Many canonical RNA-binding proteins (RBPs) have a modular structure with one or more RNA-binding domains (RBDs) and at least one prion-like domain able to drive phase transitions[9]. Prion regions represent the core of the amyloid fibril structure upon aggregation[10,11] and the RBDs, usually found at least partially folded in the aggregated species, are external to the core and might retain RNA-binding ability[12]. This is the case of hnRNPDL-2, for which a cryo-EM structure of the aggregate shows two RBDs in a disordered, solvent-exposed portion of fibrils outside of the amyloid core[12]– the “fuzzy coat”. hnRNPDL-2 fibrils have a comparable affinity towards oligonucleotides both in the monomeric and the aggregated form[12], confirming that canonical RBP aggregates can retain their functionality even upon partial refolding into different conformations.

For non-canonical RBPs and for proteins that do not display RNA binding activity the functional consequences of amyloid formation are less clear. In the case of microtubule-associated protein tau, it has been shown that fibrils can sequester RNA both in cell[6] and *in vitro*[13], implying the protein conformational changes upon aggregation can promote RNA binding. This observation is further strengthened by the fact that tau fibrils, grown *in vitro* and without any additional cofactors, can sequester RNA[14]. Tau is not known as a canonical RBP, however it has been shown that RNA co-localises with this protein[15] and promotes its aggregation[16] into morphologically distinct fibril strains[17]. A recent cryo-EM structure has shown that RNA can also trigger the formation of a previously undetected tau aggregate morphology with a central core spanning residues E391 to A426[13]. In previously solved tau fibril structures, this specific segment has been reported to be part of the fuzzy coat, implying that RNA can actively change the fibril morphology. Moreover, electron density showed an unknown density that could correspond to the actual RNA molecules being sequestered in the fibrils.

In this work, we use a computational approach to investigate nucleic acid sequestration in protein aggregates by considering the physico-chemical properties of their amyloid cores,

hidden within the fibers and occupied in cross-β bonds, and their exterior, available to establish novel interactions and exert new functions. We validated our model on alpha synuclein (aS) that represents a particularly suitable case, since its phase transition is abundantly studied and does not bind RNA in the monomeric state under physiological conditions. aS is an intrinsically disordered protein involved in synaptic vesicle trafficking and machinery assembly, mitochondrial homeostasis and DNA repair[18–20]. It has been identified as the primary component of Lewy bodies[21], pathological aggregates found in certain neurodegenerative conditions such as Parkinson's Disease (PD) and Multiple System Atrophy (MSA)[21–23]. aS has a modular organization, with a basic N-terminus (residues 1-95), containing the amyloid central region (NAC, residues 61-95)[24], and an acidic, intrinsically disordered C-terminus (residues 96-140), shown to negatively affect fibrillation and oligomerization[25–27]. The C-terminal domain can undergo several post-translational modifications such as truncations at various residues[28,29], which occur naturally in the cell, presumably as incomplete protein degradation products. However, certain disease-related C-terminal truncations also boost the protein fibrillation rates and result in diverse aggregate morphologies, leading to an increase in cellular toxicity[26,30–32]. In the full length protein, long-range electrostatic interactions between the N-terminus and C-terminus influence the partially folded intermediary states of the protein and are highly affected by changes in pH and ionic strength[27,33–35]. These factors, as well as the presence of various charged polymers, strongly affect the aggregation rates, highlighting the role of electrostatic charge in aS misfolding[36–39]. In the absence of the acidic C-terminus, such in the case of disease-related truncations, it is possible that polyanions such as RNA could have an even more significant impact on aS aggregation.

Based on our calculations, we hypothesize that amyloid-forming proteins, such as aS, can acquire the ability to sequester RNA upon aggregation. We propose that RNA binding is an acquired ability of aS aggregates and that this is the crucial contribution recapitulating the experimental trend in aggregation rate modulation. More in general, our work indicates that misfolding upon aggregation has the potential to vary the RNA-binding ability of any given protein, potentially enabling a non-RNA-binding protein to acquire the ability to bind and sequester RNA in its aggregated form.

RESULTS

Calculations recapitulate experimental results on the ability of tau fibrils to attract RNA.

We started by assessing the robustness of our Zygregator-based calculations on the known case of tau. A cryo-EM structure of tau fibrils (PDB 7SP1), grown in the presence of total RNA, shows that RNA is trapped in a pocket on the external region of the aggregate[13]. The binding occurs through cation-π interactions between an arginine residue and the RNA phosphate backbone, as well as polar contacts between the RNA bases and the histidine residue of tau[13]. The 36-amino acid region of the C-terminal disordered domain (residues 391 - 426) is identified in the cryo-EM study as the fibril core[13] and another work indicates that the third microtubule-binding repeat (residues 306 – 335)[59,60] contributes to tau aggregation. Using the Zyggregator algorithm (**Materials and Methods**), which predicts aggregation-prone elements in proteins[41], we are able to successfully identify both these regions (**Fig. 1A**).

More in detail, the cryo-EM structure of aggregated tau reveals that residues R406 and H407 on the tau fibril surface directly interact with sequestered RNA and that residues E391 - S396, V398, S400, D402, S404, D421, P423, I425, A426 are exposed to the solvent. In agreement with this finding, we predict that the surface regions identified experimentally are less hydrophobic, more disordered and have an increased RNA-binding ability (**Fig. 1B; Materials and Methods**).

We note that residues 175-220 and 230-260, corresponding to the proline-rich region and first microtubule-binding repeat, have been reported to be directly involved in interactions with RNA[61] and identified as such by our approach (**Fig. 1C**). Although these disordered regions are not captured by cryo-EM, as not in the stable amyloid core, it cannot be excluded that they play a role in sequestering RNA in the course of aggregation, as also remarked by the authors of the study[13].

Thus, this initial result indicates that regions outside the core of tau amyloid fibrils have the potential to interact with RNA and we are able to identify them in great detail using our computational approaches. This finding also suggests that other aggregating proteins might behave as tau. For a number of other disordered proteins, including calcitonin[42,43], amylin[44], glucagon[45], Aβ42[46,47], α-synuclein[48–50], hnRNPDL-2[12] and tau[13], Zyggregator shows excellent performances in identifying the amyloid core of the aggregates, with an optimal Zyggregator threshold in the range 0.5-1.5 (**Fig. 1D**).

Internal and external regions of amyloids have different biophysical properties

To further investigate the general trend that regions able to bind RNA are in the external parts of amyloids, we retrieved a dataset of amyloids available from a recent publication[62] and analyzed their physico-chemical properties (**Supplementary Table 1**) using the Zyggregator algorithm to identify regions forming the amyloid core (**Materials and Methods)**.

As expected, our calculations show that the internal part of an amyloid is more hydrophobic than the external one **(Fig. 2A**). There is a significant shift of the RNA-binding propensity and structural disorder in favor of the external part (**Fig. 2B)** while the electrostatic charge distribution does not differ substantially (**Supplementary Fig. 1A,B**). Notably, the trend is observed with different physico-chemical propensity predictors (**Supplementary Fig. 2**) Given the overall negative charge of the phosphate backbone of RNA and no significant difference in the aggregate charge distribution, we find strong RNA-binding propensity exclusively for positively charged protein regions.

To better visualize this trend we chose the structure of the amyloid formed by *Podospora anserina* protein HET-s (PDB 2RNM) as an example (**Fig. 2D-F**). The structure was determined using solid-state NMR and included the flexible external part of the amyloid. Using different color scales, the distribution trends are clearly visible, especially in the hydrophobic amyloid core (**Fig. 2D**). The RNA-binding propensity is considerably higher for the unstructured external portion of the fibers, similarly to the structural disorder (**Fig. 2D-F, Supplementary Fig. 2B**). Overall, the analysis of the HET-s amyloid structure shows excellent agreement with the predicted results from the computational analysis.

Within the set of amyloids (**Fig 2A-C),** we analyzed proteins that have RNA and DNA binding domains annotated in Uniprot (**Supplementary Tables 2 and 3**). Consistent with our calculations of RNA-binding abilities, the nucleic acid-binding domains are predicted to be external to the aggregate (**Fig. 2G**). This is well exemplified by hnRNPDL-2, a protein in which residues 224-280 were found by cryoEM to constitute the core of the fibril (PDB 7ZIR), while two RNA-binding domains are external[12]. The aggregation propensity predicted by Zyggregator captures the inner part of the aggregate very well, and the RNA-binding domains are indeed predicted to be more exposed (**Fig. 2H**). Further investigating this phenomenon at the human proteome level, we carried out predictions of prion and RNA-binding domains respectively using PLAAC[51] and *cat*RAPID signature[63]. Proportionally to the overall aggregation propensity computed with Zyggregator, we observed that prion domains are in the core of the fibril while RNA-binding domains tend to be external (**Fig. 2 I-K, Supplementary Table 4**). In agreement with our analysis of hnRNPDL-2, this finding indicates that prion domains are internal to the amyloid core, while RNA-binding domains are external.

Physico-chemical properties of alpha synuclein aggregates

We next analyzed the aggregation propensity of α-synuclein (aS or aS140), a protein with a well-characterized amyloid core[41]. We computed both the overall score of the wild type sequence (**Fig. 3A**) and performed a mutational analysis with 10 000 random single point variations[64]. In accordance with the data shown in the literature[65], the central region of the protein is crucial for protein aggregation, with a peak coinciding with the NAC region (residues 61-95) and individual peaks corresponding to the three-repeat region, as shown by Doherty and colleagues[24] (**Fig. 3A**).

Mutations in the NAC region display a strong impact on aS140 aggregation (**Fig. 3B**) and the external regions of aS140 aggregates are predicted to be disordered with low overall

hydrophobicity and able to interact with RNA molecules (**Fig. 3C**). Amino acids 67-82 appear to be particularly sensitive to change and dramatically affect the propensity of aS to aggregate, in agreement with previous experimental work[66]. Random single mutations in the C-terminus have instead negligible effects. Yet, 14 out of the 24 negatively charged residues in aS140 (18 Glu and 6 Asp) occur in this region, indicating that the truncated form, lacking all amino acids in positions 103-140 (aS103), could behave differently compared to the wild type protein in terms of aggregation propensity as well as RNA-binding ability.

Computing the scores of the physico-chemical properties for both aS140 and aS103 reveals a strong difference between the two (**Fig. 3D**). The removal of the 12 negatively charged residues in the C-terminal portion results in a net charge increment (**Fig. 3D**). Accordingly, the external regions of aS103 and aS140 show a lower degree of hydrophobicity compared to the average of the dataset and an increase in RNA-binding ability (**Fig. 3D)** due to the abundance of charged and polar residues outside the NAC region. We note that removing the C-term slightly affects the predicted structural disorder, which is consistent with the overall unfolded nature of the whole aS.

We used the *cat*RAPID approach[56,58] to estimate the ability of aS140 and aS103 to interact with RNA (**Materials and Methods**). While the C-terminus and NAC show low interaction propensity, the N-terminal domain is predicted to be external to the aggregate and able to weakly interact with RNA (**Fig. 3E**). Based on this observation, we hypothesize that aS might be able to sequester RNA upon aggregation. As we will show in the following sections, this effect could shape the dynamic of the aggregation propensity, tuning the concentration of the RNA in solution.

RNA affects the aggregation of alpha synuclein isoforms in different ways

To study experimentally the effect that RNA might have on the aggregation kinetics of aS140 and aS103, we performed aggregation assays in the presence of a fluorescent aggregate intercalator. Both aS103 and aS140 were incubated with increasing concentrations of total RNA, from 0 to 500 ng/µL, for 24 h (**Fig. 4A; Supplementary Fig. 3A-B**). Since the experiments were performed in the presence of a reporter dye, visualization of the aggregates at the end of the experiments with a fluorescence microscope confirmed the presence of protein aggregates in all samples and in every tested condition (**Supplementary Fig. 4A**).

As previously reported in literature, aS103 aggregates significantly faster than aS140, irrespective of the concentration of RNA (**Fig. 3A and B**)[67], in agreement with our predictions . In the presence of RNA, the elongation phase of aS103 is significantly faster than that of aS140 (**Fig. 4A**) and is accompanied by a consistently shorter lag phase (**Fig. 4B**). Our data analysis shows that the estimated aggregation rate of aS103 is approximately two-fold higher in the absence of RNA and up to 10-fold higher in the presence of 500 ng/µL total RNA extract. Importantly, increasing concentrations of RNA lead to varied effects on the aggregation rate of both proteins. For aS140, the aggregation rate in the presence of RNA initially scales exponentially, increasing ca. five-fold at a concentration 250 ng/µL. Then the rate seems to reach a plateau and does not increase further (**Fig. 4A**). By contrast, incremental additions of RNA monotonically increases the aggregation rate of aS103 over ten-fold, reaching a maximum at 500 ng/µL with no evidence of reaching a plateau (**Fig. 4A**). RNA affects the lag time in the opposite way, with increasing concentrations effectively decreasing

the lag time of aS103 until reaching a stable plateau at 1.6 h. The opposite is true for aS140, where higher concentrations only start drastically decreasing the lag time with RNA concentrations higher than 250 ng/μL (**Fig. 4B**). The actual maximum value of fluorescence at the end of the exponential phase however does not differ much between aS103 and aS140 (**Fig. 4C**). RNA incrementally increases the level of maximum fluorescence by additions up to 250 ng/μL when the value stabilizes. These results suggest not only that RNA has a profound effect on aS aggregation, it also affects aS140 and aS103 in different and dynamic ways as our computational analysis predicted.

Having observed the effect RNA has on the aggregation kinetics of aS, we then set to test the original hypothesis that aS can sequester the RNA upon aggregation. First we verified that the monomeric forms of both proteins do not interact with RNA (**Supplementary Fig. 4B**).

We then quantified the RNA before and after the aggregation assays in the soluble and insoluble fractions as described (**Materials and Methods**). The concentration was measured individually in each well and normalized according to the initial one. Because of differences in the volumes in different fractions (soluble and insoluble), concentrations were recalculated into RNA mass and the data presented as the percentage of total RNA mass after aggregation in different fractions (**Fig. 4D and E**). The results clearly indicate that both aS140 and aS103 are able to sequester RNA through aggregation. Furthermore, there is a clear difference between the amount of RNA sequestered by individual protein constructs. As predicted by the computational analysis, aS103 aggregates can sequester RNA to a much higher degree than aS140 (**Fig. 4F**). The difference is less noticeable at lower RNA concentrations, but becomes more pronounced at higher concentrations. As the starting concentration of the protein is always 50 μM, it is safe to assume the difference is due to the higher sequestration capacity of aS103 aggregates formed. This is further supported by the fact that less than 10 % of the starting amount of RNA is found soluble at the concentration of 50 ng/μL (**Fig. 4D**). This sequestration capacity seems to be saturated at 250 ng/μL for aS103 while it keeps increasing for aS140 (**Fig. 4F**).

To verify whether the different distribution of RNA among the soluble and insoluble fractions is specific to protein aggregation, we repeated the experiments in the same conditions with bovine serum albumin, a non-aggregating, non-RNA-binding protein. The quantification of RNA in both conditions showed comparable values to the control values of total RNA without any protein added (**Supplementary Fig. 5A**). The wells in both cases were inspected by fluorescence microscopy and no aggregates were detected (**Supplementary Fig. 4A).**

We repeated the aggregation of aS103 and aS140 in the presence of total cell DNA extract, to determine whether what we observed is an RNA-specific effect. Since DNA and RNA degrade at different rates, we adjusted the normalization to account for it (**Supplementary Fig. 5B**). aS103 still shows the ability to sequester DNA at lower initial DNA concentrations, however higher concentrations apparently limit the binding ability (**Supplementary Fig. 5C**). aS140 barely sequesters any DNA in the aggregates, with less than 10% extracted from the insoluble part (**Supplementary Fig. 5D).**

These results show that both aS140 and aS103 have the ability to sequester nucleic acids upon aggregation. The grade of incorporation is protein sequence-dependent and confirms the computational analysis predicting aS103 aggregates as the higher-propensity binders. The

results further emphasize that protein aggregation is a dynamic process, and that aggregated species not only display different biophysical features but could also acquire new functional properties distinct from the ones attributed to the monomers.

DISCUSSION

Aggregates are able to sequester other molecules, including proteins[1] and several metabolites[68]. Here, we tested the hypothesis that RNA could be trapped by amyloid aggregates. Our hypothesis was formulated considering that aggregation-prone regions of proteins are mostly hydrophobic and buried in the core of fibrils[41], while external regions, prevalently hydrophilic and structurally disordered, can interact with nucleic acids, if their electrostatics are favorable[69]. Several lines of evidence, both computational and experimental, seemed to confirm our idea. Regions outside the core of amyloids are predicted to interact with RNA and annotated RNA-binding domains have a strong propensity to be external with respect to the amyloid core. The recently reported cryoEM structures of tau and hnRNPDL-2 show that RNA can be trapped on the surface of the aggregate[13] and the RNA-binding domains are outside the amyloid core[12].

Intriguingly, a number of artificial nucleic acid molecules, RNA and DNA aptamers, have been designed to impact protein aggregation[4,70]. For instance, an aptamer called T-SO517, developed by Ikebukuro and colleagues against oligomeric αS140, bound not only to the target but also to Aβ40 oligomers[71]. This finding confirms our hypothesis that aggregates can bind to nucleic acids, although it indicates that the binding may be non-specific. Thus, it is possible to speculate that sequestration of RNA material in the cellular context could contribute to the toxicity of amyloid aggregates, a hypothesis that has been previously formulated for proteins[1].

Here, we reasoned that the formation of the amyloid core of aS fibrils, involving the most hydrophobic residues (i.e., the NAC region), leaves the charged amino acids (i.e., N-terminal and C-terminal) available for interactions with other molecules such as RNA. This is corroborated by a number of experiments such as site-directed spin labeling coupled with EPR, hydrogen/deuterium exchange, and limited proteolysis[34,48–50], indicating that the central part of the protein is engaged in forming the core of the aggregates while the rest is exposed to the solvent. We predicted that removal of the negative charges corresponding to amino acids 104-140, not included in the core of the aggregate, would sharply increase the RNA-binding propensity of the regions outside the NAC. Thus, we tested the hypothesis of RNA-binding gain-of-function upon aggregation.

Our experimental results confirm the computational observation of acquired RNA sequestration ability and identify it as a way of modulating aS aggregation process by directly affecting the aggregation rates. Sequestration is protein-sequence dependent: aS103 shows a higher sequestration propensity compared to wild type aS140, likely due to the increased electrostatic attraction. RNA also modifies the aggregation rate of both aS variants.

As previously reported, the key to modulating aS aggregation propensity lies in the long-range intra and intermolecular interactions between the N-terminal and C-terminal domains of the protein, which shield the NAC region and prevent its self-assembly[34,35]. RNA likely enters in contact with the N-terminal, thus interfering with these interactions. The truncated aS103 lacks the C-terminal domain and *a priori* cannot engage in the shielding effect of the NAC region. It is plausible that, through transient interactions via the N-terminal domain, RNA acts as a molecular bridge between protein molecules, bringing the NAC regions into closer contacts. As C-terminal truncations of aS are commonly found in brain

samples from patients and aS103 is prevalent in the medial temporal lobe in Lewy body dementia, we speculate that RNA might play a role in the onset of the disease[28].

Very recently, the combined solid-state NMR - cryoEM approach to study the fibrils of aS by Zhang and colleagues has confirmed our observations regarding the outer disorder of aS fibrils [72]. The structure (PDB 7YK2) however also highlighted the ordering and interaction between the core of the fibril and the G14 - G25 region in the N-terminal part. This suggests that additional ordering of the structure might be present in the N-terminus upon aggregation. It is important to note, however, that the fibrils in the study were grown in the absence of RNA, which could potentially prevent these additional ordering to occur upon misfolding by shielding the N-terminus.

We note that RNA concentrations in our experiments are near-to-physiological. Indeed, we considered a mammalian cell has a cell volume between 1,000 and 10,000 μm^3 (i.e., 10^{-6} µL -10^{-5} µL [73]) and an RNA amount of 10 - 30 pg (i.e., 0.01 - 0.03 ng [74]), which indicates that 500 ng/µL of RNA are in the expected normal range (i.e., 0.01 ng in 10^{-5} µL, which is 1000 ng/µL). Monomeric, physiological aS was found associated with nucleic acids[20,75], nucleosomes[76] and processing bodies[77]. However, no proof of direct and functional RNA binding by the cytosolic, aggregated aS has ever been discovered. While Siegert and colleagues have shown RNA has little effect and can actually abrogate aS phase separation[78], the molecular aspects of the effect of RNA on aS liquid-to-solid phase transition was not characterized. It should be mentioned that Cohlberg and colleagues have previously reported that negatively charged molecules such as heparin display a stronger effect on aS aggregation compared to 1µg/µL RNA[38]. Nevertheless, it must be noted that the authors used thioflavin T to quantify aS aggregation. It has been reported that this dye can interact also with RNA[79], interfering with the correct quantification of protein aggregation. For this reason, we selected a different aggregation reporter, Proteostat™, that binds to RNA only weakly[80], obtaining significantly different results. An intriguing aspect is the pathological implication of RNA sequestration in synucleinopathies.

Overall our study shows that aggregation and the misfolding associated with it can lead to different biophysical and functional properties compared to the monomeric soluble protein forms. aS, a non-RNA-binding protein, can acquire the ability to sequester RNA in a sequence-dependent manner, bringing further emphasis on the dynamic structural and functional transitions between different protein states.

ACKNOWLEDGMENTS

The authors would like to thank the 'RNA initiative' at IIT and all the members of Tartaglia's lab at CRG, Sapienza and IIT. Our research was supported by the ERC ASTRA_855923 (GGT) and H2020 projects IASIS_727658 and INFORE_825080 (GGT). EZ received funding from the Newton fellowship scheme and the MINDED fellowship of the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 754490.

MATERIALS AND METHODS

Amyloid Dataset

Amyloid proteins (**Supplementary Table 1)** were taken from the PRALINE database[40]. The internal and external parts of amyloids were computed using Zyggregator[41]. In **Fig. 1**, we report the comparison of predicted and experimentally validated internal and external regions of calcitonin[42,43], amylin[44], glucagon[45], Aβ42[46,47], α-synuclein[48–50], hnRNPDL-2[12] and tau[13]. The optimal threshold is in the range 0.5-1.5 and corresponds to an Area under the Roc Curve of 0.84.

Analysis of the inner and outer part of amyloids

For each residue in a given sequence (**Supplementary Table 1)**, we used the Zyggregator that predicts the individual contribution to aggregation based on predictions of hydrophobicity, secondary structure and [41]. Our previous work indicates that the residues with high aggregation propensities are buried in the fibril, while those with low aggregation propensity are exposed to the solvent[41].

RNA/DNA binding domains and aggregation

We exploited annotations of RNA/DNA binding domains available from UniProt to characterize their physico-chemical properties in relation to the aggregation propensity (**Supplementary Tables 2 and 3, Fig. 2**). Specifically, we computed the fraction of RNA/DNA domains belonging to the predicted external part of the amyloids. The same analysis was carried out for prion domains predicted using PLAAC[51], as in a previous work[9] (**Fig. 2**).

Random mutagenesis

We performed an *in silico* random mutation analysis to define the role of each residue of aS sequence in the aggregation propensity of the protein. We selected random positions in the sequence and replaced them, one by one, with amino acids from an uniform distribution. We used the Zyggregator algorithm[41] to compute the aggregation propensity of the protein before and after the mutation since, for each residue in a given sequence, the method predicts the individual contribution to aggregation[41]. We simulated 10000 random mutations and calculated for each one the difference with the wild type in terms of Zyggregator score, expressed as a percentage.

Physico-chemical properties of external regions

Calculations of hydrophobicity[52], disorder[53] and RNA-binding ability[54] have been carried out through the CleverMachine approach that predicts physico-chemical properties of protein sequences[55].

In the Supporting Information (**Supplementary Fig. 1 and 2**), we show that the particular scale used is not strictly important, as the behavior of the histograms of the chemical physical property did not change with the scale change.

*cat*RAPID predictions of protein-RNA interactions

In *cat*RAPID, the interaction propensity between a protein-RNA pair is computed using secondary structure properties, van der Waals and hydrogen bonding potentials[56]. The algorithm is able to separate interacting versus non-interacting pairs with an area under the curve (AUC) receiver operating characteristic (ROC) curve of 0.78 (with false discovery rate (FDR) significantly below 0.25 when the Z-score values are >2)[57]. In this work, the N-terminus of aS (residues 1-25) is predicted to attract RNA, while the C-terminal region (residues 104-140) shows much poorer propensity. In our calculations[58], we used the human transcriptome and computed the overall interaction propensity (>0.4 corresponding to low binding affinity) for different aS regions. TDP43, TRA2A and SRSF3 are used as a positive control. aS103 and aS140 are not predicted to interact as monomers.

Protein production

The cDNA for the aS140 was inserted in a pET21a vector under the control of the T7 promoter. The aS103 cDNA was produced by deletion of the appropriate sequence from the full length construct with specific PCR primers (aS103for: 5'-TAAGCGGCCGCACTCGAGCAC-3' and aS103rev: 5'-ATTCTTGCCCAACTGGTCCTTTTTGACAAAGC-3').
Competent *E.coli* BL21 [DE3] cells were transformed with both constructs. 2 L of fresh LB broth supplemented with 50 µg/mL ampicillin were inoculated with overnight cultures with a 100:1 ratio. Cultures were grown at 37 °C up to an optical density of ca. 0.7 at 600 nm. Protein expression was induced with 0.5 mM IPTG and the cultures were shaken for 4 hours. Cells were harvested by centrifugation at 4500 g at 4 °C for 30 min, afterwards the pellet was resuspended on ice in 20 mM potassium phosphate buffer pH 7.2 and frozen immediately.
For protein purification the cells were thawed in ice-cold water and lysed by boiling at 95 °C for 30 min. The boiled suspension was centrifuged at 18000 g at 4 °C for 30 min, then streptomycin sulfate was added to the supernatant up to a concentration of 10 mg/mL. The solution was centrifuged at 18000 g at 4 °C for 30 min and the pellet containing precipitated impurities was discarded. The protein was slowly precipitated by addition of crystalline ammonium sulfate up to 360 mg/mL and the suspension was centrifuged again at 18000 g at 4 °C for 30 min. The pellet was resuspended in 20 mM Tris-HCl pH 7.4 with 400 mM KCl and dialysed against 20 mM Tris-HCl pH 7.4 overnight at 4 °C.
The proteins were purified using heparin affinity chromatography on a HiTrap™ Heparin column to remove remaining nucleic acids. Proteins were eluted with heparin elution buffer (20 mM potassium phosphate pH 7.2, 800 M NaCl) and loaded onto a HiLoad™ 16/600 Superdex™ 75 gel filtration column equilibrated in 20 mM potassium phosphate pH 7.2, 100 mM KCl. The eluted proteins were pooled and checked for concentration with BCA assay and purity with spectropolarimetry and SDS PAGE, before being stored aliquoted at - 80 °C.

Protein aggregation assays

A frozen protein aliquote of aS140 or aS103 was quickly thawed and filtered with a 0.22 µm syringe filter immediately before each assay. Total yeast RNA (totRNA, Roche) was resuspended in aggregation assay buffer (20 mM potassium phosphate pH 7.2, 100 mM KCl, 5 mM $MgCl_2$). For aggregation assays in the presence of DNA, total DNA from herring sperm (Sigma Merck) was resuspended in sterile, nuclease-free water for the stock solution. Protein, totRNA and DNA were diluted to the final concentration in the aggregation assay buffer with a 1000-time diluted Proteostat™ detection dye. For the BSA control, ultra-pure, DNA- and RNA-free BSA was used (Sigma Merck) and diluted to the final concentration in the sample buffer as described above for aS. The samples were distributed in 6 replicates on a 96-well black plate with transparent bottom and a single borosilicate glass bead was added to each well to ensure sample homogeneity and reproducibility. The excitation wavelength was set to 505 nm and the emission to 590 nm. Plates were incubated at 37 °C with constant double-orbital shaking at 200 rpm. Reads were taken every 10 min for aS140 and every 5 min for aS103 for 24 hours, with each read represented as an average of 10 scans.
All protein samples were aggregated in parallel to total RNA alone at the same starting concentrations and conditions for RNA degradation control and the values reported have already been accounted for degradation.

Aggregation data analysis

Data from the aggregation assays, obtained with Proteostat™ fluorescence dye, were fitted into a sigmoid curve using the Hill function (1):

$$h\,(x) \;=\; A\,\frac{x^h}{x^h + K^h}\;(1)$$

where A, h and K are the fitting parameters.

To compute the aggregation rate, we determined the slope of the fitted curves before plateau. With the previously set parameters and a Taylor expansion of $h(x)$ around K, we calculated the slope of the aggregation α curve as (2):

$$\alpha \;\equiv\; h\,(x \;=\; K) \;=\; h\,(K) \;+\; \frac{\partial h(x)}{\partial x};\, x \;=\; K$$

$$\alpha \;=\; A\frac{h}{2K}\;(2)$$

We define the "delay time" as the intersection of the fitting line with slope α centered in $x \;=\; K$ and $y \;=\; \frac{A}{2}$. The delay time τ is defined resolving the linear system (3):

$$y(x) \;=\; \alpha x \;+\; q$$

$$y(\tfrac{K}{2}) = \alpha\,\frac{K}{2} + \;q \;=\; \frac{A}{2} \qquad \rightarrow \qquad q \;=\; \frac{A - \alpha\,K}{2}$$

$$\alpha\tau \;+\; \frac{A - \alpha\,K}{2} \;=\; 0 \qquad \rightarrow \qquad \tau \;=\; \frac{\alpha\,K - A}{2\,\alpha}$$

$$\tau \;=\; K\,\frac{(h-1)}{h}$$

Nucleic acid quantification and extraction from protein aggregates

After aggregation, individual replicates were transferred from the 96-well plate to Eppendorf tubes and centrifuged at 4 °C, 18.000 g for 30 min. The soluble fraction was transferred to another tube, RNA found in solution was quantified using Qubit RNA BR, and DNA with Qubit DNA BR kit (Thermo Fisher Scientific, USA).
For RNA extraction, the insoluble pellet was resuspended in 25 µL of digestion buffer (2 M urea, 100 µg/mL proteinase K, 3 mM DTT) and incubated at room temperature for 15 min. 125 µL of 6 M guanidine hydrochloride was added to a final concentration of 5 M and the sample was incubated at room temperature for further 30 min. 1 mL of Trizol™ was added and RNA was extracted using standard protocol by the manufacturer.
DNA was extracted from the insoluble protein pellet with the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer protocol for DNA extraction from blood. The final elution was performed in 50 µL of buffer AE. Extracted nucleic acids were again quantified using the Qubit fluorescent dyes.

BIBLIOGRAPHY

1. Olzscha, H. *et al.* Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. *Cell* **144**, 67–78 (2011).
2. Gallardo, R., Ranson, N. A. & Radford, S. E. Amyloid structures: much more than just a cross-β fold. *Fold. Bind. ● Proteins* **60**, 7–16 (2020).
3. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
4. Zacco, E. *et al.* Probing TDP-43 condensation using an in silico designed aptamer. *Nat. Commun.* **13**, 3306 (2022).
5. Shmookler Reis, R. J. *et al.* "Protein aggregates" contain RNA and DNA, entrapped by misfolded proteins but largely rescued by slowing translational elongation. *Aging Cell* **20**, e13326 (2021).
6. Lester, E. *et al.* Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron* **109**, 1675-1691.e9 (2021).
7. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11**, 777–788 (2010).
8. Hallegger, M. *et al.* TDP-43 condensation properties specify its RNA-binding and regulatory repertoire. *Cell* **184**, 4680-4696.e22 (2021).
9. Gotor, N. L. *et al.* RNA-binding and prion domains: the Yin and Yang of phase separation. *Nucleic Acids Res.* **48**, 9491–9504 (2020).
10. Li, Q., Babinchak, W. M. & Surewicz, W. K. Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat. Commun.* **12**, 1620 (2021).
11. Lee, M., Ghosh, U., Thurber, K. R., Kato, M. & Tycko, R. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. *Nat. Commun.* **11**, 5735 (2020).
12. Garcia-Pardo, J. *et al.* Cryo-EM structure of hnRNPDL-2 fibrils, a functional amyloid associated with limb-girdle muscular dystrophy D3. *Nat. Commun.* **14**, 239 (2023).
13. Abskharon, R. *et al.* Cryo-EM structure of RNA-induced tau fibrils reveals a small C-terminal core that may nucleate fibril formation. *Proc. Natl. Acad. Sci.* **119**, e2119952119 (2022).
14. Chakraborty, P. *et al.* Co-factor-free aggregation of tau into seeding-competent RNA-sequestering amyloid fibrils. *Nat. Commun.* **12**, 4231 (2021).
15. Zhang, X. *et al.* RNA stores tau reversibly in complex coacervates. *PLOS Biol.* **15**, e2002183 (2017).
16. Dinkel, P. D., Holden, M. R., Matin, N. & Margittai, M. RNA Binds to Tau Fibrils and Sustains Template-Assisted Growth. *Biochemistry* **54**, 4731–4740 (2015).
17. Zwierzchowski-Zarate, A. N. *et al.* RNA induces unique tau strains and stabilizes Alzheimer's disease seeds. *J. Biol. Chem.* **298**, 102132 (2022).
18. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat. Rev. Neurosci.* **14**, 38–48 (2013).
19. Nguyen, M., Wong, Y. C., Ysselstein, D., Severino, A. & Krainc, D. Synaptic, Mitochondrial, and Lysosomal Dysfunction in Parkinson's Disease. *Trends Neurosci.* **42**, 140–149 (2019).
20. Schaser, A. J. *et al.* Alpha-synuclein is a DNA binding protein that modulates DNA repair with implications for Lewy body disorders. *Sci. Rep.* **9**, 10919 (2019).
21. Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. *Proc. Natl. Acad. Sci.* **95**, 6469–6473 (1998).
22. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
23. Grazia Spillantini, M. *et al.* Filamentous α-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. *Neurosci. Lett.* **251**, 205–208 (1998).
24. Doherty, C. P. A. *et al.* A short motif in the N-terminal region of α-synuclein is critical for both aggregation and function. *Nat. Struct. Mol. Biol.* **27**, 249–259 (2020).
25. Hoyer, W., Cherny, D., Subramaniam, V. & Jovin, T. M. Impact of the Acidic C-Terminal Region Comprising Amino Acids 109−140 on α-Synuclein Aggregation in Vitro. *Biochemistry* **43**, 16233–16242 (2004).
26. Levitan, K. *et al.* Conserved C-Terminal Charge Exerts a Profound Influence on the Aggregation Rate of α-Synuclein. *J. Mol. Biol.* **411**, 329–333 (2011).
27. Farzadfard, A. *et al.* The C-terminal tail of α-synuclein protects against aggregate replication but is critical for oligomerization. *Commun. Biol.* **5**, 123 (2022).
28. Sorrentino, Z. A. & Giasson, B. I. The emerging role of α-synuclein truncation in aggregation and

disease. *J. Biol. Chem.* **295**, 10224–10244 (2020).
29. Pancoe, S. X. *et al.* Effects of Mutations and Post-Translational Modifications on α-Synuclein In Vitro Aggregation. *J. Mol. Biol.* **434**, 167859 (2022).
30. Iyer, A. *et al.* C-Terminal Truncated α-Synuclein Fibrils Contain Strongly Twisted β-Sheets. *J. Am. Chem. Soc.* **139**, 15392–15400 (2017).
31. van der Wateren, I. M., Knowles, T. P. J., Buell, A. K., Dobson, C. M. & Galvagnion, C. C-terminal truncation of α-synuclein promotes amyloid fibril amplification at physiological pH. *Chem Sci* **9**, 5506–5516 (2018).
32. Delenclos, M. *et al.* Cellular models of alpha-synuclein toxicity and aggregation. *J. Neurochem.* **150**, 566–576 (2019).
33. Uversky, V. N., Li, J. & Fink, A. L. Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation*. *J. Biol. Chem.* **276**, 10737–10744 (2001).
34. Roeters, S. J. *et al.* Evidence for Intramolecular Antiparallel Beta-Sheet Structure in Alpha-Synuclein Fibrils from a Combination of Two-Dimensional Infrared Spectroscopy and Atomic Force Microscopy. *Sci. Rep.* **7**, 41051 (2017).
35. Stephens, A. D. *et al.* Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nat. Commun.* **11**, 2820 (2020).
36. Munishkina, L. A., Henriques, J., Uversky, V. N. & Fink, A. L. Role of Protein−Water Interactions and Electrostatics in α-Synuclein Fibril Formation. *Biochemistry* **43**, 3289–3300 (2004).
37. Antony, T. *et al.* Cellular Polyamines Promote the Aggregation of α-Synuclein*. *J. Biol. Chem.* **278**, 3235–3240 (2003).
38. Cohlberg, J. A., Li, J., Uversky, V. N. & Fink, A. L. Heparin and Other Glycosaminoglycans Stimulate the Formation of Amyloid Fibrils from α-Synuclein in Vitro. *Biochemistry* **41**, 1502–1511 (2002).
39. Kam, T.-I. *et al.* Poly(ADP-ribose) drives pathologic α-synuclein neurodegeneration in Parkinson's disease. *Science* **362**, eaat8407 (2018).
40. Vandelli, A. *et al.* The PRALINE database: protein and Rna humAn singLe nucleotIde variaNts in condEnsates. *Bioinforma. Oxf. Engl.* **39**, btac847 (2023).
41. Tartaglia, G. G. *et al.* Prediction of Aggregation-Prone Regions in Structured Proteins. *J. Mol. Biol.* **380**, 425–436 (2008).
42. Moriarty, D. F., Vagts, S. & Raleigh, D. P. A role for the C-terminus of calcitonin in aggregation and gel formation: a comparative study of C-terminal fragments of human and salmon calcitonin. *Biochem. Biophys. Res. Commun.* **245**, 344–348 (1998).
43. Reches, M., Porat, Y. & Gazit, E. Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin. *J. Biol. Chem.* **277**, 35475–35480 (2002).
44. Luca, S., Yau, W.-M., Leapman, R. & Tycko, R. Peptide Conformation and Supramolecular Organization in Amylin Fibrils: Constraints from Solid State NMR. *Biochemistry* **46**, 13505–13522 (2007).
45. Pedersen, J. S., Dikov, D. & Otzen, D. E. N- and C-terminal hydrophobic patches are involved in fibrillation of glucagon. *Biochemistry* **45**, 14503–14512 (2006).
46. Török, M. *et al.* Structural and dynamic features of Alzheimer's Abeta peptide in amyloid fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **277**, 40810–40815 (2002).
47. Lührs, T. *et al.* 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17342–17347 (2005).
48. Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. Biochemical characterization of the core structure of alpha-synuclein filaments. *J. Biol. Chem.* **277**, 19213–19219 (2002).
49. Der-Sarkissian, A., Jao, C. C., Chen, J. & Langen, R. Structural Organization of α-Synuclein Fibrils Studied by Site-directed Spin Labeling*. *J. Biol. Chem.* **278**, 37530–37535 (2003).
50. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W. & Woods, V. L. Structure and properties of α-synuclein and other amyloids determined at the amino acid level. *Proc. Natl. Acad. Sci.* **102**, 15477–15482 (2005).
51. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinforma. Oxf. Engl.* **30**, 2501–2502 (2014).
52. Roseman, M. A. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200**, 513–522 (1988).
53. Campen, A. *et al.* TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **15**, 956–963 (2008).
54. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).

55. Klus, P. *et al.* The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinforma. Oxf. Engl.* **30**, 1601–1608 (2014).
56. Bellucci, M., Agostini, F., Masin, M. & Tartaglia, G. G. Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445 (2011).
57. Lang, B., Armaos, A. & Tartaglia, G. G. RNAct: Protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res.* **47**, D601–D606 (2019).
58. Armaos, A., Colantoni, A., Proietti, G., Rupert, J. & Tartaglia, G. G. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. *Nucleic Acids Res.* **49**, W72–W79 (2021).
59. Hernández, F. *et al.* Tau Aggregation. *Neuroscience* (2022) doi:10.1016/j.neuroscience.2022.04.024.
60. von Bergen, M. *et al.* Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure. *Proc. Natl. Acad. Sci.* **97**, 5129–5134 (2000).
61. Wang, X. *et al.* The proline-rich domain and the microtubule binding domain of protein tau acting as RNA binding domains. *Protein Pept. Lett.* **13**, 679–685 (2006).
62. Vendruscolo, M. & Fuxreiter, M. Sequence Determinants of the Aggregation of Proteins Within Condensates Generated by Liquid-liquid Phase Separation. *J. Mol. Biol.* **434**, 167201 (2022).
63. Livi, C. M., Klus, P., Delli Ponti, R. & Tartaglia, G. G. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinforma. Oxf. Engl.* **32**, 773–775 (2016).
64. Monti, M., Armaos, A., Fantini, M., Pastore, A. & Tartaglia, G. G. Aggregation is a Context-Dependent Constraint on Protein Evolution. *Front. Mol. Biosci.* **8**, (2021).
65. Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* **13**, 1939–1941 (2004).
66. Waxman, E. A., Mazzulli, J. R. & Giasson, B. I. Characterization of Hydrophobic Residue Requirements for α-Synuclein Fibrillization. *Biochemistry* **48**, 9427–9436 (2009).
67. Zhang, Z. *et al.* Asparagine endopeptidase cleaves α-synuclein and mediates pathologic activities in Parkinson's disease. *Nat. Struct. Mol. Biol.* **24**, 632–642 (2017).
68. Anand, B. G., Prajapati, K. P. & Kar, K. Aβ1-40 mediated aggregation of proteins and metabolites unveils the relevance of amyloid cross-seeding in amyloidogenesis. *Biochem. Biophys. Res. Commun.* **501**, 158–164 (2018).
69. Tartaglia, G. G. & Caflisch, A. Computational analysis of the S. cerevisiae proteome reveals the function and cellular localization of the least and most amyloidogenic proteins. *Proteins* **68**, 273–278 (2007).
70. Murakami, K., Izuo, N. & Bitan, G. Aptamers targeting amyloidogenic proteins and their emerging role in neurodegenerative diseases. *J. Biol. Chem.* **298**, 101478 (2022).
71. Tsukakoshi, K., Abe, K., Sode, K. & Ikebukuro, K. Selection of DNA aptamers that recognize α-synuclein oligomers using a competitive screening method. *Anal. Chem.* **84**, 5542–5547 (2012).
72. Zhang, S. *et al.* Conformational Dynamics of an α-Synuclein Fibril upon Receptor Binding Revealed by Insensitive Nuclei Enhanced by Polarization Transfer-Based Solid-State Nuclear Magnetic Resonance and Cryo-Electron Microscopy. *J. Am. Chem. Soc.* **145**, 4473–4484 (2023).
73. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* **14**, e1002533 (2016).
74. Russo, J. & Russo, I. H. *Techniques and Methodological Approaches in Breast Cancer Research*. (Springer, 2014). doi:10.1007/978-1-4939-0718-2.
75. Hegde, M. L. & Rao, K. S. J. DNA induces folding in α-synuclein: Understanding the mechanism using chaperone property of osmolytes. *Arch. Biochem. Biophys.* **464**, 57–69 (2007).
76. Pinho, R. *et al.* Nuclear localization and phosphorylation modulate pathological effects of alpha-synuclein. *Hum. Mol. Genet.* **28**, 31–50 (2019).
77. Hallacli, E. *et al.* The Parkinson's disease protein alpha-synuclein is a modulator of processing bodies and mRNA stability. *Cell* **185**, 2035-2056.e33 (2022).
78. Siegert, A. *et al.* Interplay between tau and α-synuclein liquid–liquid phase separation. *Protein Sci.* **30**, 1326–1336 (2021).
79. Sugimoto, S., Arita-Morioka, K., Mizunoe, Y., Yamanaka, K. & Ogura, T. Thioflavin T as a fluorescence probe for monitoring RNA metabolism at molecular and cellular levels. *Nucleic Acids Res.* **43**, e92–e92 (2015).
80. Zacco, E. *et al.* RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *J. Mol. Biol.* **431**, 1671–1688 (2019).

Figures Legends

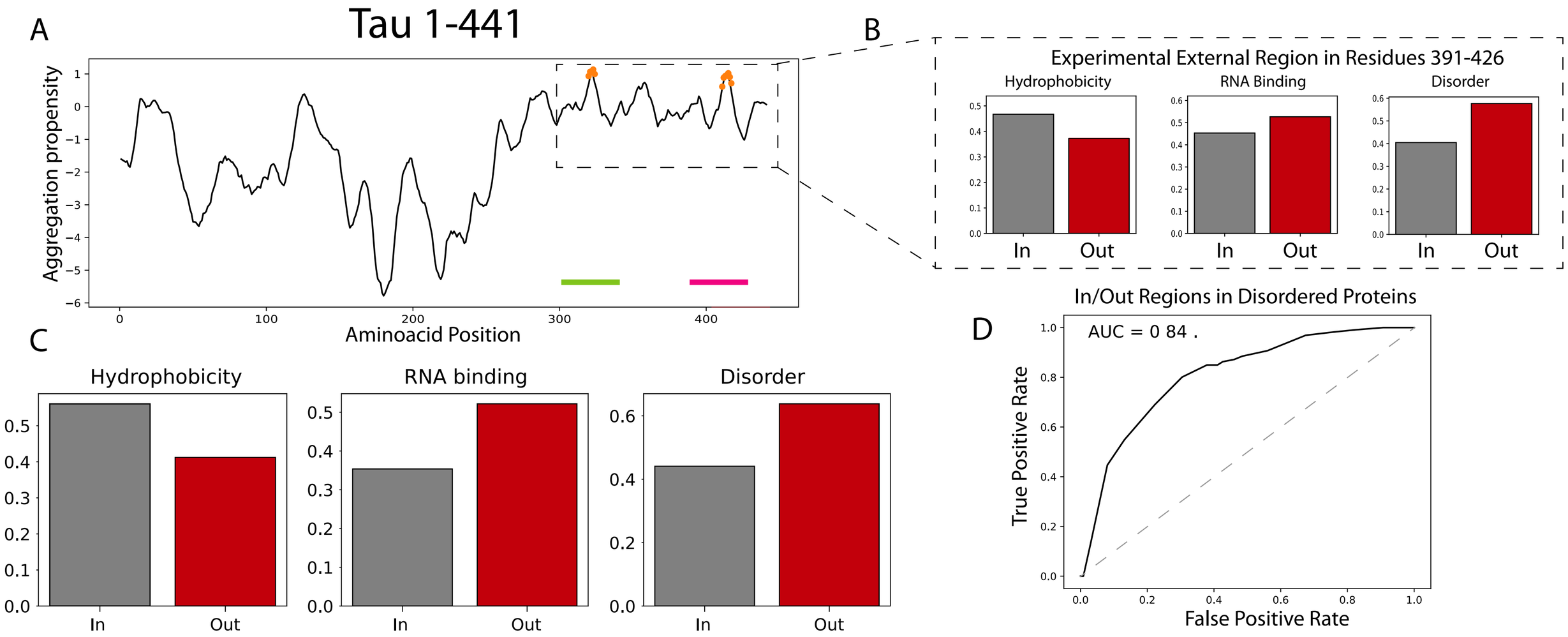
Figure 1. *Tau aggregation profile and prediction of physico-chemical properties*. **A)** Aggregation propensity profile of tau, with the predictions of the filament internal part in orange, while green and red bars indicate the regions found to be internal experimentally. **B)** Physico-chemical properties of residues 391 - 441, with the internal and external properties of the region marked in black (in) and red (out), respectively. **C)** Hydrophobicity, disorder, and RNA binding propensity of the entire sequence. **D)** ROC curve computed to determine the optimal Zyggregator threshold for finding internal and external residues based on experimental data from various proteins. The optimal Zyggregator threshold is in the range 0.5-1.5.

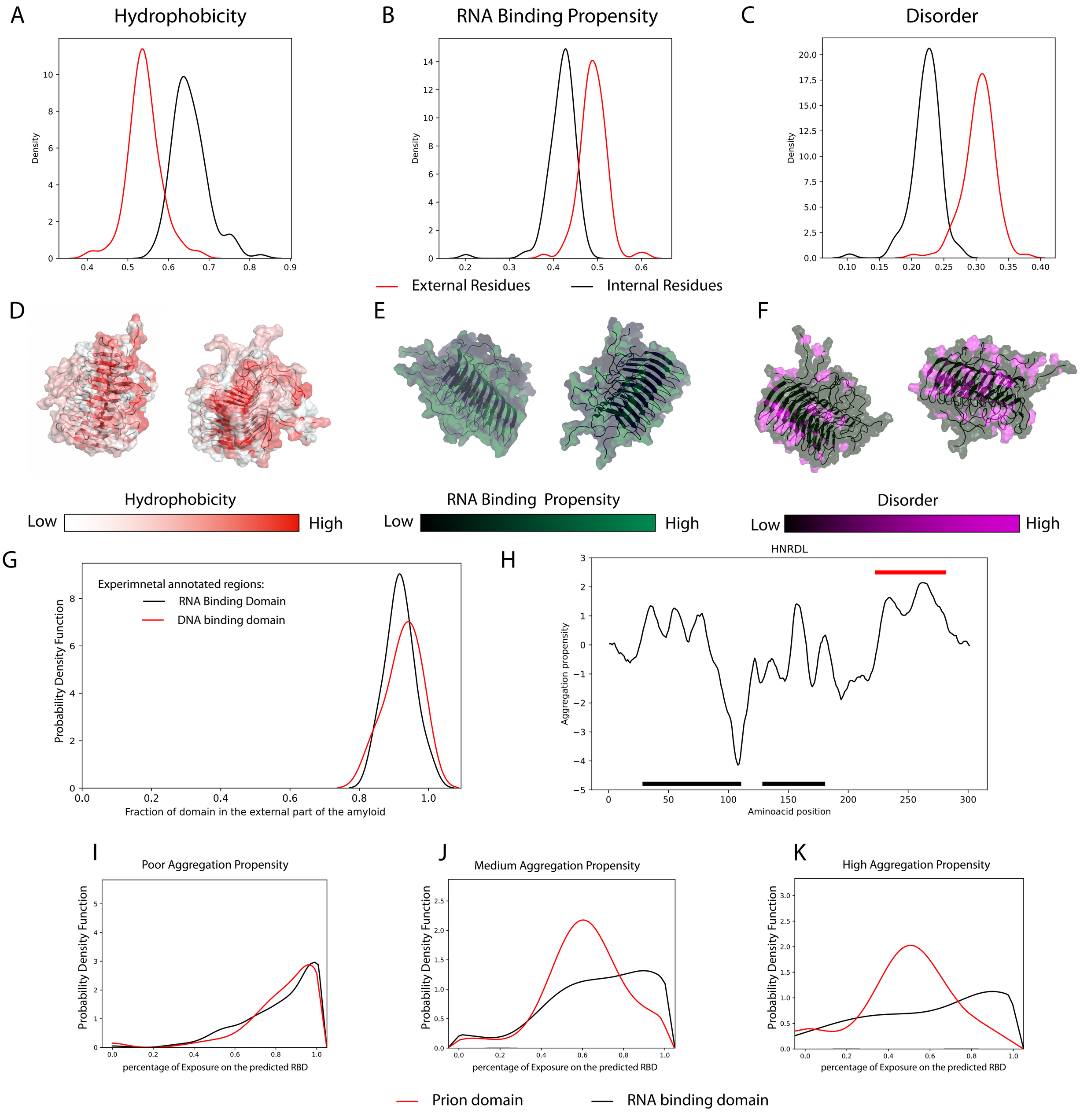
Figure 2. *Physico-chemical properties of internal and external regions of amyloids***. A-C)** Distribution score of the amyloid dataset using the Zyggregator method for internal and external portions of the amyloids (the curves are significantly different with p-value<0.05; Kolmogorov Smirnov test).. **D-F)** Molecular structure of the HET-s prion protein (PDB 2RNM) solved by solid-state NMR, depicted by different color scales for hydrophobicity, RNA-binding propensity, and disorder, respectively. **G)** Percentage of the RNA and DNA binding domains belonging to the external part of the fibril. **H)** Zyggregator profile of hnRNPDL-2. Black lines mark the RBDs and the red line the internal residues found experimentally. **I-K)** Percentage of residues exposed for the RNA-binding and prion domains for the human proteome at different aggregation propensity scores (see also **Supplementary Table 4**).

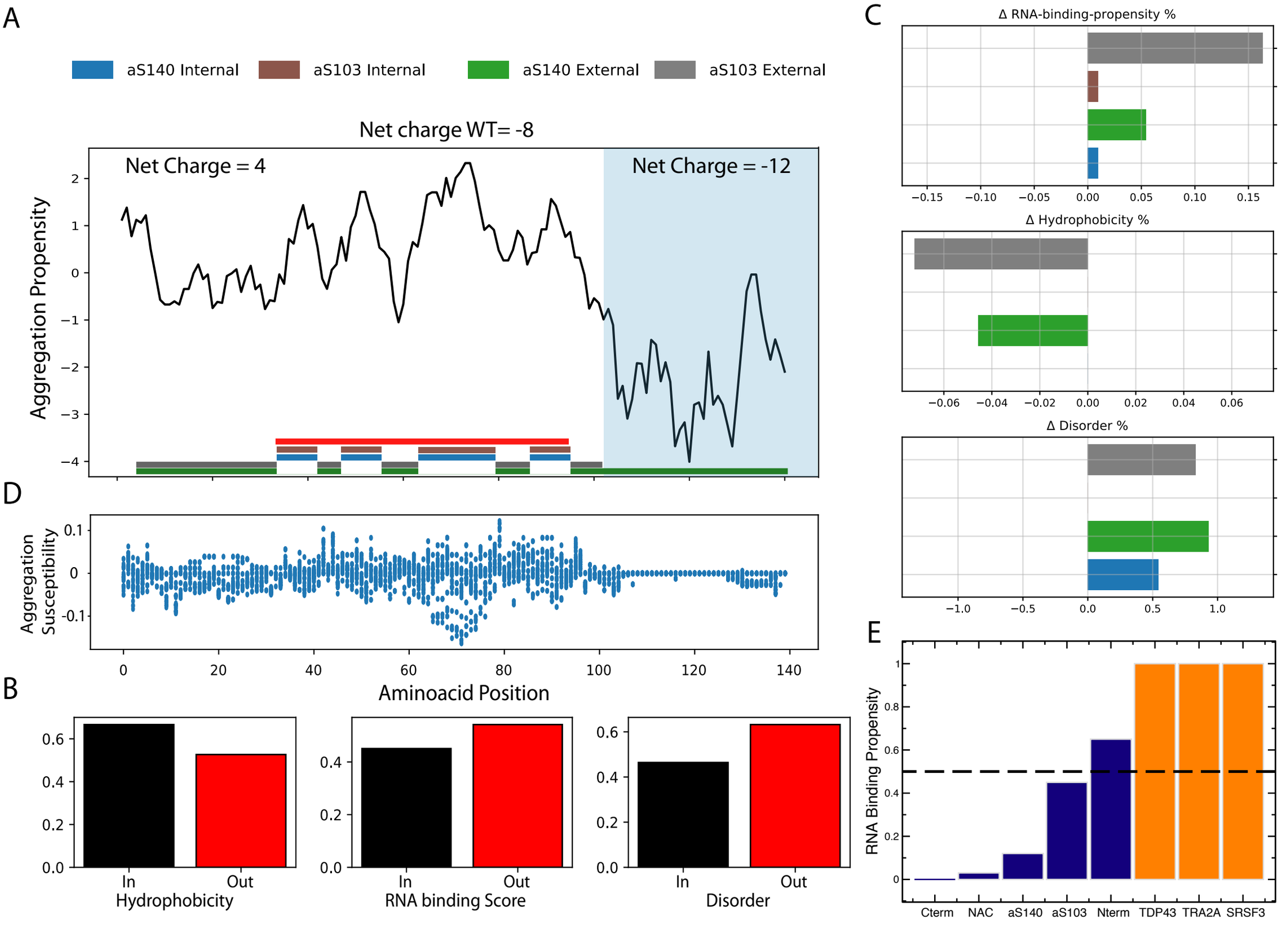
Figure 3. *Aggregation propensity of aS140 and aS103.* **A)** Zyggregator profile of aS140, with colored bars indicating predicted internal and external parts. The experimentally determined amyloid core of aS is marked with a red bar. **B)** Physico-chemical properties of aS140 and aS103. **C)** Random mutagenesis analysis shows that the NAC region and N-terminus are susceptible to significant changes in aggregation propensity. **D)** Hydrophobicity, RNA-binding propensity, and disorder of aS relative to its internal and external parts. Differences with respect to the mean of the amyloid histograms are shown (**Fig. 2A-C**). **E)** RNA-binding predictions of different regions of aS140 and aS103, compared with canonical RBP (scores > 0.5 indicate non-negligible interaction propensities).

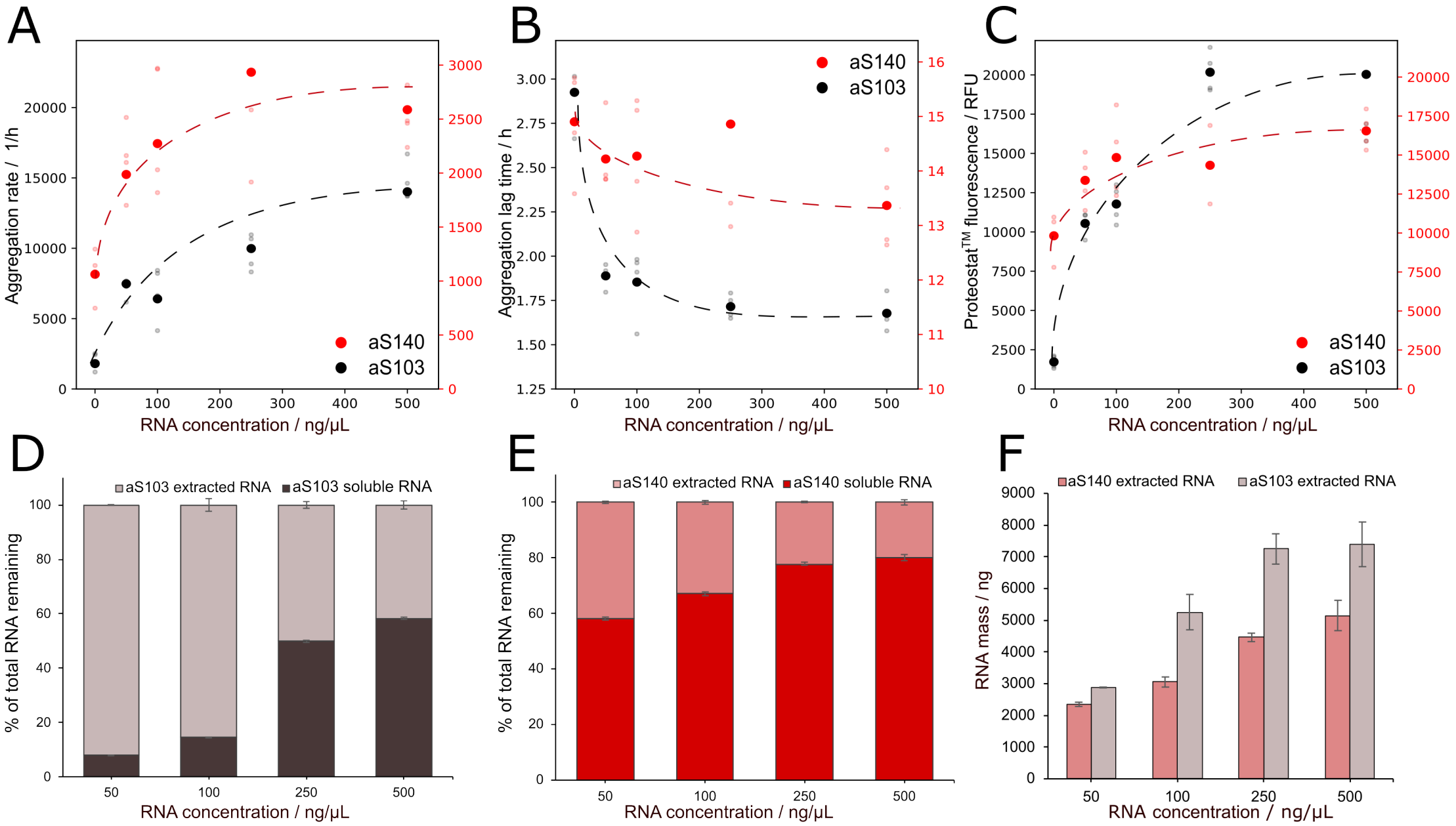
Figure 4. *In vitro aggregation assays of aS140 and aS103 in the presence of RNA.* **A**) Calculated aggregation rates from aggregation data (**Supplementary Fig. 3**) of time resolved Proteostat™ fluorescence. Data shown are calculated from parameters, derived from the fitted curve to 6 technical replicates of 3 separate experiments (see **Materials and Methods**). The presence of RNA increases the rate of aggregation for both proteins, with the aS140 rate reaching a plateau with concentrations higher than 250 ng/μL. aS103 has a significantly shorter aggregation lag time than aS140. **B**) The addition of RNA further decreases the lag phase, however increasing concentrations do not seem to drastically affect if further while the opposite is true for aS140. **C)** Increasing concentration of RNA also elevates the plateau of fluorescence for both proteins, especially significantly for aS103. **D and E)** Quantification of nucleic acids from aS aggregation assays indicates different behavior for aS140 and aS103. Comparing RNA in solution and the fraction extracted from aggregates of aS103 and aS140, we found that RNA is preferentially sequestered by aS103 aggregates. **F)** The average amount

of RNA extracted from aS103 and aS140 aggregates at different initial starting concentrations show again that aS103 aggregates have a higher propensity of sequestering soluble RNA upon aggregation.


A
Tau 1-441
Aggregation propensity
Aminoacid Position
B
Experimental External Region in Residues 391-426
Hydrophobicity
RNA Binding
Disorder
In
Out
C
Hydrophobicity
RNA binding
Disorder
In
Out
D
In/Out Regions in Disordered Proteins
AUC = 0 84 .
True Positive Rate
False Positive Rate


A
Hydrophobicity
Density
B
RNA Binding Propensity
Density
C
Disorder
Density
External Residues
Internal Residues
D
Hydrophobicity
Low
High
E
RNA Binding Propensity
Low
High
F
Disorder
Low
High
G
Experimnetal annotated regions:
RNA Binding Domain
DNA binding domain
Probability Density Function
Fraction of domain in the external part of the amyloid
H
HNRDL
Aggregation propensity
Aminoacid position
I
Poor Aggregation Propensity
Probability Density Function
percentage of Exposure on the predicted RBD
J
Medium Aggregation Propensity
Probability Density Function
percentage of Exposure on the predicted RBD
K
High Aggregation Propensity
Probability Density Function
percentage of Exposure on the predicted RBD
Prion domain
RNA binding domain


A
aS140 Internal
aS103 Internal
aS140 External
aS103 External
Net charge WT= -8
Net Charge = 4
Net Charge = -12
Aggregation Propensity
D
Aggregation Susceptibility
Aminoacid Position
B
Hydrophobicity
RNA binding Score
Disorder
In
Out
C
Δ RNA-binding-propensity %
Δ Hydrophobicity %
Δ Disorder %
E
RNA Binding Propensity
Cterm
NAC
aS140
aS103
Nterm
TDP43
TRA2A
SRSF3


A
Aggregation rate / 1/h
RNA concentration / ng/µL
aS140
aS103
B
Aggregation lag time / h
RNA concentration / ng/µL
aS140
aS103
C
Proteostat™ fluorescence / RFU
RNA concentration / ng/µL
aS140
aS103
D
aS103 extracted RNA
aS103 soluble RNA
% of total RNA remaining
RNA concentration / ng/µL
E
aS140 extracted RNA
aS140 soluble RNA
% of total RNA remaining
RNA concentration / ng/µL
F
aS140 extracted RNA
aS103 extracted RNA
RNA mass / ng
RNA concentration / ng/µL