A Theoretical Model reveals RNA sequestration in Alpha Synuclein Aggregates

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

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
*correspondence to gian@tartaglialab.com

ABSTRACT

The modulating effect of nucleic acids on protein aggregation has recently come into the spotlight, with RNA shown to either prevent or promote protein assembly depending on the molecular context. Here, we computed the biophysical properties of amyloids and observed a trend indicating that regions outside the aggregates core are highly prone to interact with nucleic acids. In the case of alpha synuclein (aS), an intrinsically disordered protein abundantly expressed in the brain, found in the nucleus and involved in Parkinson's disease, our predictions indicate that regions outside the aggregate are able to contact RNA, but the acidic C-terminal prevents the formation of stable interactions. We performed aggregation assays with both the wild type α-synuclein (aS140) as well as a C-terminally truncated isoform (aS103) and found that while RNA increases the aggregation rate of aS103, it decreases that of aS140, although at higher RNA concentrations the trend is inverted. To further elucidate the effects driving this behavior, we built a general dynamic model that describes the aggregation process of monomers in the presence of RNA. Our predictions indicate that RNA affects aS103 and aS140 aggregation in a non-linear manner and prioritize the interaction between RNA and aggregate as the most relevant. To confirm this, we extracted RNA from aS103 and aS140 aggregates and observed that they indeed acquire distinct RNA-sequestering abilities. Our research demonstrates that binding to transcripts can drastically alter the aggregation of a protein and represents an important gain-of-function mechanism to be further investigated.

INTRODUCTION

In recent years, the view on protein aggregation and its correlation to pathology has evolved from the mechanistic concepts of misfolding and seeding[1,2] to the more biology-centered phenomena of protein phase separation, aberrant condensation and aging of the condensates[3–5]. The prion model of exponential growth of long, beta sheet-enriched fibrils is giving way to the phase separation of low complexity domain-containing proteins and the importance of cellular factors contributing to liquid-to-solid phase transition[6]. Among these, nucleic acids are particularly relevant in modulating phase transitions of aggregation-prone proteins[7,8] and have been isolated from both physiological condensates and pathological inclusions[9,10]. The fact that numerous proteins found aggregated in neurological disorders are RNA-binding and form ribonucleoprotein assemblies brought RNA into the spotlight as one of the crucial modulators of phase transition[7]. For instance, RNA has been shown to be able to both actively promote[11,12] and abrogate aggregation[13] of RNA-binding proteins.

Not much is known about the ability of proteins to interact with nucleic acids after aggregation has taken place, or whether aggregation might confer novel nucleic acid binding ability to proteins that are not classified as RNA/DNA-binding. With this work, we aim to investigate the possibility of this instance to occur by developing a theoretical model that takes into account how protein aggregate exteriors might acquire RNA binding ability upon phase transition. Among these proteins, alpha synuclein (aS) represents a particularly suited case. aS is an intrinsically disordered protein involved in synaptic vesicle trafficking and machinery assembly, mitochondrial homeostasis and DNA repair[14–16]. It has been identified as the primary component of Lewy bodies[17], pathological aggregates found in certain neurodegenerative conditions such as Parkinson's Disease (PD) and Multiple System Atrophy (MSA)[1,17,18]. aS has on occasion been termed a "chameleon" protein[19] due to its highly flexible structural conformation ranging from alpha helical, beta sheet to completely unfolded as a result of changes in the cellular environment[20]. The structural flexibility also reflects in a wide variety of multimeric and aggregated forms, from diverse oligomers to protofibrils and amyloid fibrils[21–23]. The modular organization of the protein is also varied. The N-terminus (residues 1-95) contains 12 basic residues, has been shown to be acetylated and contains the amyloid central region (NAC, residues 61-95)[24]. The N-terminal and NAC regions have been shown to fold into an alpha helical structure upon binding the membrane of synaptic vesicles, acting as an inter-vesicular tether[20,25]. The C-terminus (residues 96-140) contains 15 acidic residues, is completely disordered and has been shown to negatively affect fibrillation and oligomerization[26–28]. The C-terminal domain can undergo several post-translational modifications such as phosphorylation of Ser129[29] and truncation at various residues[30]. These shorter forms occur naturally in the cell, presumably as incomplete protein degradation products. It has been shown however that PD-related mutations determine an increase in protein abundance, compared to the wild type aS[30–32]. C-terminal truncations also boost the protein fibrillation rates and result in diverse aggregate morphologies, leading to an increase in cellular toxicity[27,33–35]. Long-range electrostatic interactions between the N-terminus and C-terminus of the protein influence the partially folded intermediary states of the protein in solution and are highly affected by changes in pH and ionic strength[28,36–38]. 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[39–42]. In the absence of the acidic C-terminus, it is possible that polyanions such as nucleic-acids could have an impact on aS aggregation.

aS has also been shown to undergo liquid-liquid phase separation and liquid-to-solid phase transition both *in vitro*[43] and *in vivo*[44]. It has been identified to be part of functional molecular condensates of the synaptic vesicle protein machinery[45], processing bodies[46] and assemblies with protein tau[47]. Tau and aS have long been linked as "partners in pathology", with several studies showing the cross-seeding of aggregation[48–50]. Tau has also been shown to form RNA-dependent condensates[12] that are able to recruit aS within[47]. Yet, proteomic analyses of pathological inclusions found in the brain of PD and MSA patients revealed little cross-seeding[51], implying that RNA-dependent coacervate condensation with aS as the client protein should not be the prevalent mechanism of inclusion formation.

Here, we use a computational approach to investigate whether RNA molecules play a role in protein aggregation. Based on our calculations, we hypothesize and experimentally demonstrate that RNAs modulates aS self-assembly by affecting the aggregation rates in a non-linear way. The analysis of regions prone to interact with RNA indicate that C-terminus deprivation can promote aS aggregates sequestration of transcripts. 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. Our results confirm that RNA molecules differentially influence the aggregation rates of the full length aS (aS140) and of a truncated version deprived of the C-terminal region (aS103), and this is reflected on the different amount of RNA that can be recovered from their aggregates. More in general, our work indicates that aggregation and misfolding can vary the RNA-binding ability of any given protein.

MATERIALS AND METHODS

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 ug/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, afterwards 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 nucleic acids 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 protein was purified using heparin affinity chromatography on a HiTrap™ Heparin column to remove the remaining nucleic acids. The protein was eluted with the 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 protein was pooled and checked for concentration with BCA assay and purity with spectropolarimetry and SDS PAGE, afterwards it was stored aliquoted at - 80 °C.

Protein aggregation assays

A frozen protein aliquote of aS140 or aS103 was quickly thawed and filtered with a 0.2 um 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 MgCl2). 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 15 min 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.

Soluble nucleic acid quantification and extraction

After aggregation, the 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, soluble RNA was quantified using Qubit RNA BR and DNA with Qubit DNA BR kit (Thermo Fisher Scientific, USA).

For RNA extraction, the insoluble part 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 part 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.

Aggregation data analysis

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

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

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) :

$$\begin{aligned} \alpha &\equiv \tilde{h}(x = K) = h(K) + \frac{dh(x)}{dx}|_{x=K} \\ \alpha &= A\frac{h}{2K} \end{aligned} \tag{2}$$

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

$$\begin{aligned} y(x) &= \alpha x + q \\ y(K/2) &= \alpha K/2 + q = A/2 \rightarrow q = \frac{A - \alpha K}{2} \\ \alpha\tau + \frac{A - \alpha K}{2} &= 0 \rightarrow \tau = \frac{\alpha K - A}{2\alpha} \\ \tau &= K\frac{(h-1)}{h} \end{aligned} \tag{3}$$

Analysis of the inner and outer part of aggregates

Using the Zyggregator algorithm it is possible to compute the amyloid propensity of proteins[52]. For each residue in a given sequence, the method predicts the individual contribution to aggregation[2]. 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[2]. Thus, Zyggregator can be exploited to identify the inner and outer part of the fibrils.

Indeed, given an aggregation profile of a protein, a threshold $\hat{t}$ can be set so that the residues with a Zyggregator score higher than $\hat{t}$ are considered the inner part of the amyloid and the ones that are lower the outer part. Using this approach, the recognition of the inside and outside of the amyloid is quite precise with the threshold $\hat{t} = 1.3$ as in the original research work of Tartaglia et al[2].

Using this procedure we assessed the physical properties of the residues of the inner and the outer part of amyloids. Running this analysis in an amyloid dataset recently collected[6] we analyzed the distribution of the surface hydrophobicity, net charge and nucleic acid binding propensity of different parts of the sequence. We computed the chemical physical properties of the sequence considering Roseman's hydrophobicity[53], nucleic acid binding propensity[54] and net charge of protein sequences. These properties have been selected using the CleverMachine[55].

Random Mutation analysis

In order to define the role of each residue of aS sequence in the aggregation propensity of the protein, we performed an *in silico* random mutation analysis. We selected random positions in the sequence and replaced them, one by one, with random amino acids from an uniform distribution. We used the Zyggregator algorithm[2] 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[2]. 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.

***cat*RAPID predictions of protein-RNA interactions**

In *cat*RAPID, the interaction propensity between a protein-RNA pair is computed using secondary structure properties as well as van der Waals and hydrogen bonding potentials[56]. The algorithm is able to separate interacting vs. 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-terminal of aS (MDVFMKGLSKAKEGVVAAAEKTKQG) is predicted to interact with RNA, while the C-terminal (MPVDPDNEAYEMPSEEGYQDYEPEA) shows poor binding ability. In our calculations[58], we used the yeast transcriptome and computed the overall interaction propensity (>0.4) of the different aS regions. TDP43 is used as a positive control. aS103 and aS140 are not predicted to interact.

RESULTS

Internal and external regions of amyloids have different biophysical properties

The recently published cryo-EM structure of tau fibrils, grown in the presence of total RNA, shows RNA trapped in a pocket on the surface of the fibril[59]. The critical interactions have been shown to be cation-π interactions between an arginine residue and the RNA bases, as well as hydrophobic interactions. Here we decided to assess the overall propensity of aggregates to bind RNA with the following three parameters: hydrophobicity, RNA-binding propensity and overall charge.

Using a dataset available from a recent publication[6] (**Supplementary Table 1**) we defined the interior and exterior of amyloid fibrils using the Zyggregator algorithm as described in **Materials and Methods**. In essence, the Zyggregator algorithm allows the accurate distinction between inner and outer portions of an aggregate, which can be exploited to define regions to analyze. In **Fig. 1,** we report the computed overall hydrophobicity[55] and RNA–binding propensity[56] for all amyloids present in the dataset. Electrostatic charge distribution is reported in **Supplementary Fig. 1A,B.**

Previous works indicate that amino acids with high aggregation propensities are buried in the fibril, while those with low aggregation propensity are exposed to the solvent[2,60]. As expected, our calculations show that the internal part of an amyloid is more hydrophobic than the external one **(Fig. 1A**). There is a significant shift for the RNA-binding propensity towards the external part (**Fig. 1B)** while the electrostatic charge distribution does not seem to differ substantially (**Supplementary Fig. 1A,B**). Given the overall negative charge of the phosphate backbone of RNA and no significant difference in the aggregate charge distribution, it is likely the RNA-binding propensity of the aggregates in the absence of any canonical RNA-binding motifs is not solely electrostatically driven. To better visualize this trend we chose the structure of the amyloid formed by *Podospora anserina* protein HET-s (PDB 2RMN) as an example (**Fig. 1C,D**). 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. 1C**). The RNA-binding propensity is considerably higher in the unstructured external portion, while the charge is non differentially distributed (**Fig. 1D, Supplementary Fig. 1A,B**). Overall, the analysis of the HET-s amyloid structure shows excellent agreement with the predicted results from the computational analysis.

Random mutation analysis of alpha synuclein identifies the NAC region as critical for aggregation propensity

We used the Zyggregator approach to identify the aS amino acid clusters critical for its aggregation and their associated properties. For this, we computed both the overall score of the wild type sequence (**Fig. 2A**) and performed a mutational analysis with 10 000 random single mutations (Fig. 2C)[61]. The two approaches show consistent results. In accordance with the data shown in the literature[62], the N-terminal 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. 2A**).

Mutations in the NAC region display a stronger impact on aS aggregation. In particular, 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[63]. 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 significantly differently compared to the wild type protein in terms of aggregation propensity as well as the properties of their aggregated species. This truncated form is in fact found in the brain of patients with Alzheimer's disease. Produced by cleavage with asparagine protease, aS103 has a significantly higher aggregation rate and results in a worse prognosis for the patients[64].

Computing the scores of the three parameters for both wild-type aS140 and aS103 shows exactly this stark difference (**Fig. 2B**). The removal of the 14 negatively charged residues in the C-terminal portion results in a net charge increment, as well as higher RNA-binding propensity (**Fig. 2B**). The external regions of aS103 and aS140 show a lower degree of hydrophobicity compared to the average of the dataset (**Fig. 2B**), likely due to an abundance of charged and polar residues outside the NAC region.

This observation would suggest that the truncated isoform aS103 has a higher ability of sequestering RNA molecules. As we will show in the following sections, this effect could non-trivially shape the dynamic of the aggregation propensity, tuning the concentration of the RNA in solution.

Modeling the effect of RNA on aS aggregation kinetics

To better characterize and systematically assess the effect of RNA on protein aggregation, we developed a general dynamic model that describes the aggregation process of monomers in the presence of RNA. It has been shown that RNA can both induce[8,13] or reduce[65,66] protein aggregation in a context-dependent manner. We postulated that the key difference between the aS140 and aS103 is the electrostatic contribution of the RNA in the formation of the aggregate (**Fig. 2A**). While not specifically binding the monomers, the presence of RNA in solution during the aggregation process could modulate the protein behavior through weak, transient electrostatic interactions. It has been shown that weak, long-range interactions between the N- and C-terminal domains of aS have a key role in its aggregation propensity[38]. In the case of aS140, the presence of RNA in solution slows down the aggregation by buffering the solvent exposure of the N-terminus and stabilizing the intra- and intermolecular interactions of aS140. In the case of aS103, however, the protective C-terminus is missing and we propose that the transient interaction of the N-terminus with the RNA further exposes the NAC region. We therefore assume that RNA would exert opposite effects on aS140 and on aS103: slow down the aggregation rate of the former while speeding up the rate of the latter.

Our model to describe RNA-dependent aggregation of aS140 is reported in (4):

$$
\begin{aligned}
\frac{dp}{dt} &= -\alpha p \frac{p}{p+Kp} \frac{\beta^{h_1}}{(\beta^{h_1}+r^{h_1})} - \delta r \cdot p \\
\frac{dA}{dt} &= \alpha p \frac{p}{(p+Kp)} \frac{\beta^{h_1}}{(\beta^{h_1}+r^{h_1})} - \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{dAr}{dt} &= \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{dr}{dt} &= -\delta r \cdot p - \gamma r \frac{r^{h_3}}{(r^{h_3}+K_r^{h_3})} - \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{drp}{dt} &= \delta r \cdot p \\
\frac{drr}{dt} &= \gamma r \frac{r^{h_3}}{(r^{h_3}+K_r^{h_3})}
\end{aligned}
\tag{4}
$$

The model adapted for aS103 is described in (5):

$$
\begin{aligned}
\frac{dp}{dt} &= -\alpha p \frac{p}{p+Kp} \left(\frac{r}{(r+\beta)} + \eta \right) - \delta r \cdot p \\
\frac{dA}{dt} &= \alpha p \frac{p}{(p+Kp)} \left(\frac{r}{(r+\beta)} + \eta \right) - \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{dAr}{dt} &= \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{dr}{dt} &= -\delta r \cdot p - \gamma r \frac{r^{h_3}}{(r^{h_3}+K_r^{h_3})} - \xi r \frac{A^{h_2}}{(A^{h_2}+K_a^{h_2})} \\
\frac{drp}{dt} &= \delta r \cdot p \\
\frac{drr}{dt} &= \gamma r \frac{r^{h_3}}{(r^{h_3}+K_r^{h_3})}
\end{aligned}
\tag{5}
$$

where:

- *p* is the monomer in solution;
- *A* represents the aggregates;
- *r* representes RNA in solution;
- *rp* are the RNA-protein interactions;
- Ar are the RNA-aggregate interactions;

The model contains five essential parameters describing the contribution of each force in the equations, **α**, **γ**, **β**, **δ**, **ξ**:

- **α** is the rate of protein aggregate formation;

- **β** represents the rate associated with the addition of a monomer to the already formed aggregate in presence of RNA;
- **ξ** is the RNA-aggregate binding rate;
- **δ** is the RNA-monomer binding rate;
- **γ** is the RNA-RNA binding rate.

In this phenomenological description of the system, we considered that monomeric proteins interact with each other promoting aggregation. We also assumed that, once the aggregation starts, it can no longer be stopped. No specific binding between any particular class of RNAs with proteins or aggregates was taken into account. Since we assume RNA to have an overall, non-specific electrostatic effect on the system, our system does not take copy numbers explicitly into account. Indeed, low copy numbers could result in a noise effect that can render the mean field system described in the equations not suitable for a correct description.

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 first confirmed that neither variant binds to RNA in their monomeric forms (**Supplementary Fig. 2A**). We then 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. 3; Supplementary Fig. 2B**).

As per literature, aS103 aggregates significantly faster than aS140, irrespective of the concentration of RNA, in agreement with our predictions (see **Modeling the effect of RNA on aS aggregation kinetics**)[64]. Our data analysis shows that the estimated aggregation rate of aS103 is approximately 10-fold higher in the absence of RNA and up to 25-fold higher in the presence of 250 ng/μL total RNA extract. Importantly, RNA does not seem to affect the aggregation rate in a linear manner and its effect on the aggregation rate varies as the concentration of RNA increases. For aS140, the aggregation rate in the presence of RNA initially decreases by ca. 30 % up to a concentration of 100 ng/μL and then starts increasing by a maximum of 1.5-fold at 500 ng/μL. The opposite is true for aS103, for which the addition of up to 250 ng/μL RNA increases the aggregation rate by approximately 2.5-folds. 500 ng/μL RNA decreases the rate to reach a level similar to the one defined for 100 ng/μL. 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. 2B**).

Interpretation of the experimental data through the theoretical model reveals aggregate-RNA interaction as the key driving force for aggregation

Next we fitted our theoretical model to the experimental aggregation rates to determine the contribution of protein and RNA components. Five key parameters were fitted: **α**, **β**, **γ**, **δ** and **ξ**. They represent the different contributions of the biochemical effects in action. The model does indeed manage to recapitulate the trend inversion and shows excellent correlation (>0.95 Pearson's correlation) with the experimental data (**Fig. 4 A,B**).

Given the equation system is degenerate, we imposed constraints using *in silico* predictions and backed them up with experimental evidence. Zyggregator was used to estimate protein

aggregation rate **α** (Zyggregator scores are 0.46 for aS140 and 0.62 for aS103, which is in agreement with experiments reported in **Figure 3**; **Materials and Methods**) and *cat*RAPID was employed to estimate the parameter **δ** for monomeric protein-RNA binding rate (we predict that aS103 and aS140 do not interact with transcripts, which is in agreement with experiments reported in **Supplementary Fig. 2A**). Also the prediction of RNA-aggregate binding rate **ξ** was estimated with *cat*RAPID [56,58] using the external portion of the aggregate as input sequence (the outer part of the fibril, comprising the N- and C-terminus, has an overall RNA-interaction propensity of 0.12, while the N-terminus alone has RNA-interaction propensity of 0.22, **Supplementary Figure 4A**).

For the RNA-RNA interaction (**γ**) parameter, we did not observe any RNA-only condensation in the conditions used for the aggregation assays (**Supplementary Fig. 2B**) and therefore postulated the interactions to be constant in the system and not contributing to the sequestration of available RNA. This is consistent with the observation that RNA-RNA interaction scores reported are significantly high for phase-separating transcripts such as human NEAT and ribosomal RNAs (0.30+/-0.05)[67], while total yeast RNA is associated with low values (0.19+/-0.03)[67], which suggest that experiments with greater concentrations are needed for RNA condensation.

Finally, the RNA effect term **β** aims to describe the addition of a monomer to the aggregate in presence of RNA. This is a phenomenological term that mimics the effect of an increasing electrostatic effect on the intra- and intermolecular interactions that promote aggregations[16,37,68]. We imposed no costrains on this term.

Overall, thee weight of the parameters (**Figure 4 C,D**) is in good agreement with our theoretical estimates. Critically, it shows RNA-aggregate interactions to be the key contributors in recapitulating the trend observed in the experimental data (**Supplementary Fig. 4C,D**).

aS aggregated species preferentially sequester RNA in a sequence-dependent manner

The results of the computational analysis clearly indicate that RNA-aggregate interactions are the key drivers of the trend inversion observed in the kinetic curves. To validate our predictions, we seeked to validate it experimentally by quantifying the RNA present both in solution and in the aggregated fraction after aggregation (**Fig. 5**). The results confirm that both aS140 and aS103 are able to sequester RNA through aggregation.

There is a notable difference between the amount of RNA left in solution after aggregation of the two protein constructs. The amount quantified for aS140 remains higher than for aS103 in all conditions , which is in agreement with the computational predictions. The difference is especially noticeable at lower RNA concentrations, with less than 5% of RNA remaining soluble for aS103 at the concentration of 50 ng/μL compared to almost 60% for aS140 (**Fig. 5A,B**). Irrespective of the initial RNA concentration, there is no large variability in the percentage of RNA recovered in solution, when incubated with aggregating aS140 and the value reaches the maximum of 80% with 500 ng/μL initial RNA (see also **Supplementary Fig. 5A,B**). 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, and total RNA alone

to account for RNA degradation by itself. The quantification of RNA in both conditions showed comparable values while no protein aggregation was observed (**Supplementary Fig. 5C**).

To determine whether what we observed is an RNA-specific effect, we repeated the aggregation of aS103 and aS140 in the presence of total cell DNA extract. Since DNA and RNA degrade at different rates, we adjusted the normalization to account for it (**Supplementary Fig. 5D**). aS103 still shows the ability to sequester DNA at lower initial DNA concentrations, however higher concentrations apparently limits the binding ability (**Supplementary Fig. 6**). aS140 on the other hand barely sequesters any DNA in the aggregates, with less than 10% quantified from the insoluble part (**Supplementary Fig. 6)**.

These results suggest that while both aS140 and aS103 have the ability to sequester nucleic acids upon aggregation, RNA is preferentially incorporated in their aggregated species. The grade of incorporation is 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.

CONCLUSIONS

In a previous study we predicted that amyloids are a class of protein format with strong propensity to interact with RNA[69]. Our hypothesis was supported by experimental evidence on amyloidogenic proteins such as TAR DNA-binding protein 43 and fragile X mental retardation protein[70], whose aggregation in presence of RNA we studied in subsequent works[13,65]. aS was also amongst the amyloids identified in our screening and one for which our previous calculations suggested an RNA-mediated interaction network controlling protein abundance[69]. While our calculations revealed a hypothetical functional circuit, which would modulate aS production and aggregation, we did not compute how RNA interactions would affect aggregation.

In this work, we reasoned that aggregation requires the hydrophobic residues that form the core of aS fibrils (i.e., belonging to the NAC region), and leaves the charged amino acids (i.e., N-terminal and C-terminal) available for interactions with other molecules such as RNA (**Fig. 6**). This is corroborated by a number of experiments such as site-directed spin labeling coupled with EPR, hydrogen/deuterium exchange, and limited proteolysis[37,71–73], 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. Importantly, in our calculations we found that removal of the negative charge of the amino acids 103-140, which are not included in the core of the aggregate upon misfolding, sharply increases the RNA-binding propensity of the aggregated species. Thus, we proceed to test the hypothesis of RNA-binding gain-of-function upon aggregation.

Our experimental results confirm the hypothesis of the 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 (**Fig. 6**). 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 part of the protein, which shield the NAC region and prevent its misfolding and intermolecular self-assembly[37,38]. RNA likely enters in contact with the N-terminal, thus interfering with these interactions. In the case of aS140, the aggregation rate in the presence of lower RNA concentrations decreases (**Fig. 6**). This is likely due to a competition mechanism with the C-terminus, which at these concentrations favors RNA-protein interactions, lowering the overall aggregation propensity. This also reflects in the increased amount of RNA sequestered within or on the protein aggregates at lower concentrations. aS103, however, lacks the C-terminal domain and *a priori* cannot engage in the shielding effect of the NAC region. It is likely 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. This would increase the aggregation rate at lower concentrations. However, at higher RNA concentrations the elevated RNA-protein ratio would act as a partial electrostatic shield and lower the aggregation rate. It is not yet clear how the RNA sequestration itself contributes to this process. Molecules bound to the surface of the aggregates could inhibit both the elongation and the secondary nucleation rate of aggregate growth, even if the primary nucleation phase is sped up. It has been shown that tau fibrils grown in the presence of RNA exhibit higher structural rigidity and disintegrate upon RNase addition[59]. RNA could thus also impede fragmentation and secondary nucleation, however further studies are necessary to better characterize the process.

In our experiments RNA concentrations were near-to-physiological. Indeed, a mammalian cell has a cell volume between 1,000 and 10,000 μm^3 (i.e., 10^{-6} µL -10^{-5} µL [74]) and an RNA amount of 10 - 30 pg (i.e., 0.01 - 0.03 ng [75]), 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). In fact, we expect the local nucleic acid concentration to be higher in the nucleus, suggesting that the effects measured should be even stronger. Also, aS was shown to localize in the nucleus of the cells[76,77], associated with nucleic acids[16,78], nucleosomes[79] and processing bodies[46], but no proof of direct and functional RNA binding effects on aggregation have been so far carried out. While Siegert and colleagues have shown RNA has little effect and can actually abrogate aS phase separation[47], the molecular aspects of the effect of RNA on aS aggregation was not characterized. It should be mentioned that Cohlberg and colleagues have previously shown that negatively charged molecules such as heparin have a significant effect than RNA of 1000 ng/µL on aS aggregation[41]. Yet, Cohlberg and colleagues used thioflavin T which, by interacting with RNA[80], creates interference in the aggregation assay while we used another dye, PROTEOSTAT™, that binds to RNA only weakly[13], thus obtaining significantly different results.

Most importantly, our experiments revealed that the RNA effect on aS aggregation is concentration-dependent and does not affect the aggregation rate of aS in a linear manner. An intriguing aspect is the pathological implication of RNA sequestration in synucleinopathies. Our results show that the C-terminally truncated aS variants found in disease might have a higher RNA-binding propensity compared to the wild type protein. 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, which is afflicted early in disease[30].

Overall our study shows that aggregation and the misfolding associated with it can lead to varying biophysical and functional properties of aggregated species. Our theoretical model can successfully replicate experimental data and reveal the critical contributions to the process of protein aggregation in the presence of RNA. aS, a non-RNA-binding protein, can acquire the ability to sequester RNA in a sequence-dependent manner, thus bringing further emphasis on the fluent 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.

Figures

Figure 1. Distribution score of the amyloid dataset[6] (available in **Supplementary table 1**). (**A**) Distribution of hydrophobicity score from [53] computed using Zyggregator method for internal (black) and external (red) portions of the amyloids. (**B**) Distribution of RNA-binding propensity[55,81] for the amyloid dataset. Upon aggregation, the propensity of nucleic acid binding for the external part is higher than for the internal part. **C,D**) Structure of the HET-s (218-289) prion protein, determined by solid-state NMR (PDB 2RNM)[82], depicted by different color scales for (**C**) hydrophobicity and (**D**) RNA-binding propensity.

Figure 2. Analysis of aS aggregation propensity and aggregate physicochemical properties. (**A**). The analysis of wild type aS140 with Zyggregator shows the region comprising residues 1-100, roughly corresponding to the N-terminus of the protein (residues 1-95), as the driver of protein aggregation. The C-terminal region between residues 101 and 140 shows a significant reduction of aggregation propensity. The red line indicates experimentally determined regions of alpha synuclein fibril[2] (**B**) The region comprising residues 1-105 is the most susceptible to single random mutations in terms of effect on the aggregation propensity, while mutations in the C-terminal region are not predicted to alter it. (**C**) Physicochemical properties of aS140 and aS103 aggregate. The removal of the negatively charged C-terminal part results in a net positive electrostatic charge for aS103. The overall hydrophobicity is very similar to the dataset in the amyloid core, however is much lower in external regions.

Figure 3. *In vitro* aggregation assays of both aS constructs. (**A**) Experimental data of time resolved PROTEOSTAT™ fluorescence in the absence and presence of total RNA. Data shown are the fitted curve to 6 experimental replicates, with reads every 15 min (see Materials and Methods). As per literature[64], aS103 shows significantly faster aggregation kinetics compared to aS140, reaching the fluorescence plateau already after 4 h of incubation compared to aS140. The presence of RNA decreases the aggregation lag time of aS103 and increases the one of aS140 (**Supplementary figure 3**). (**B**) Calculated aggregation rates (α) of fitted experimental data plotted against starting RNA concentration. There is a clear inversion of trend, showing the non-linear effect of RNA on aggregation kinetics.

Figure 4. Theoretical model to predict RNA-aggregate interaction. (**A,B**) Correlation plots between the experimental aggregation rates (α) and the theoretical fitted values predicted by our computational model. There is excellent correlation between both values, showing the model manages to recapitulate the aggregation process well. The weights calculated from these fits show aggregate-RNA as the critical parameter for trend inversion, both for aS140 (**C**) and aS103 (**D**).

Figure 5. 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 (**A**) aS103 and (**B**) aS140, we found that RNA is preferentially sequestered by aS103 aggregates.

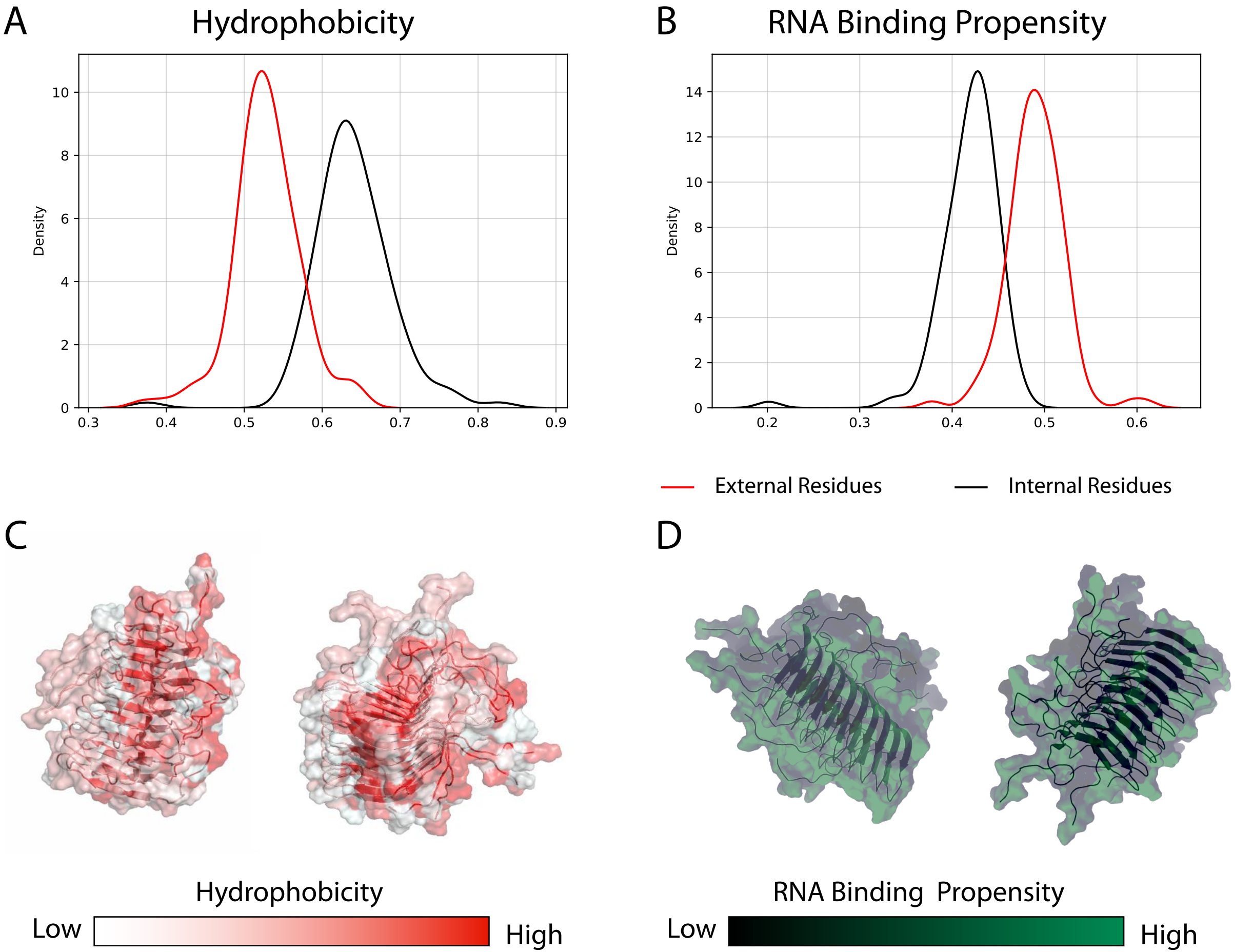
Figure 6. Proposed mechanism of aS nucleic acid/RNA sequestration. RNA differentially modulates aS140 and aS103 aggregation rates and this is reflected on the different amount of RNA that the two proteins can entrap in their aggregates. While N- and C-terms can interact in aS140, in aS103 the N-term is free to interact with RNA species.

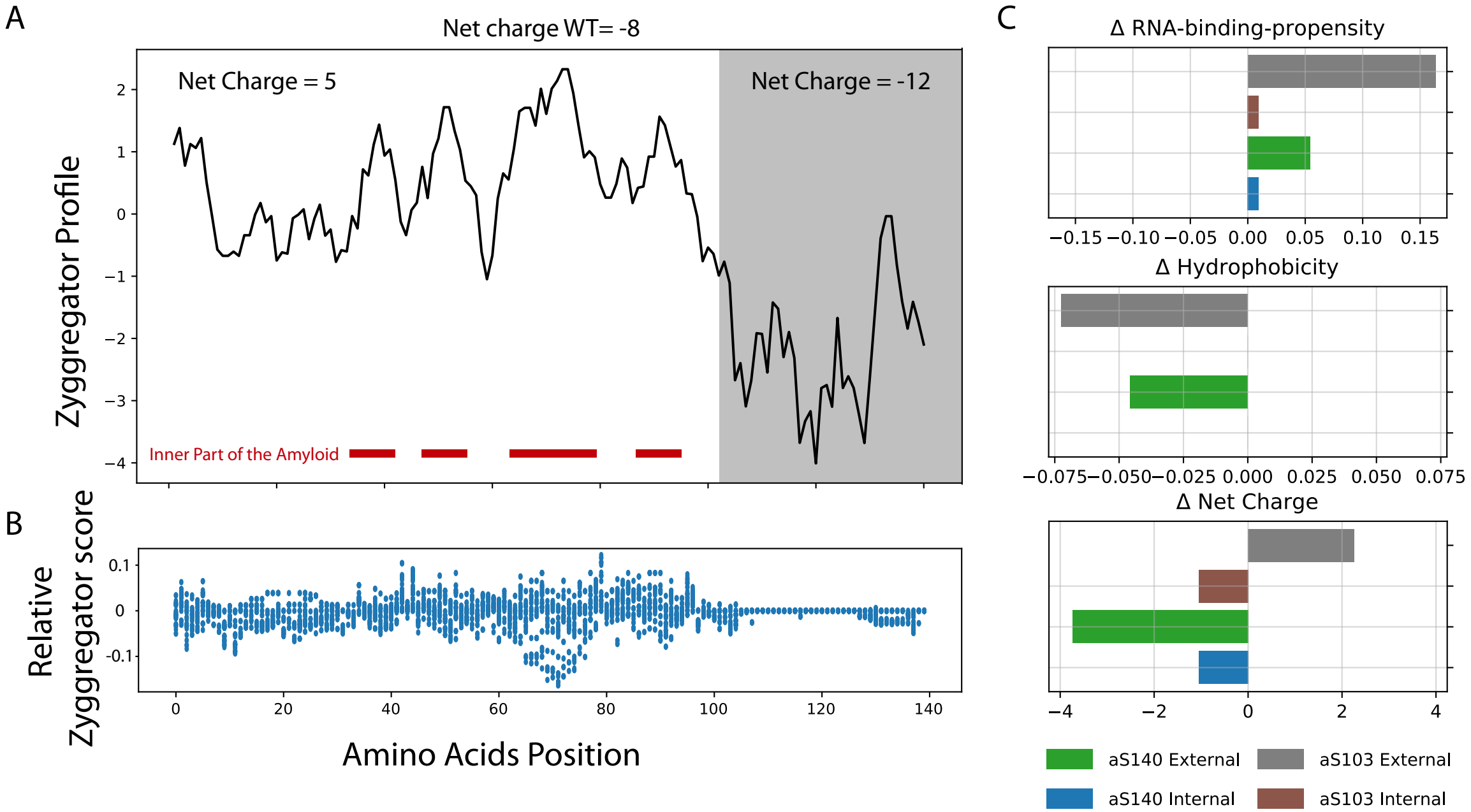
BIBLIOGRAPHY

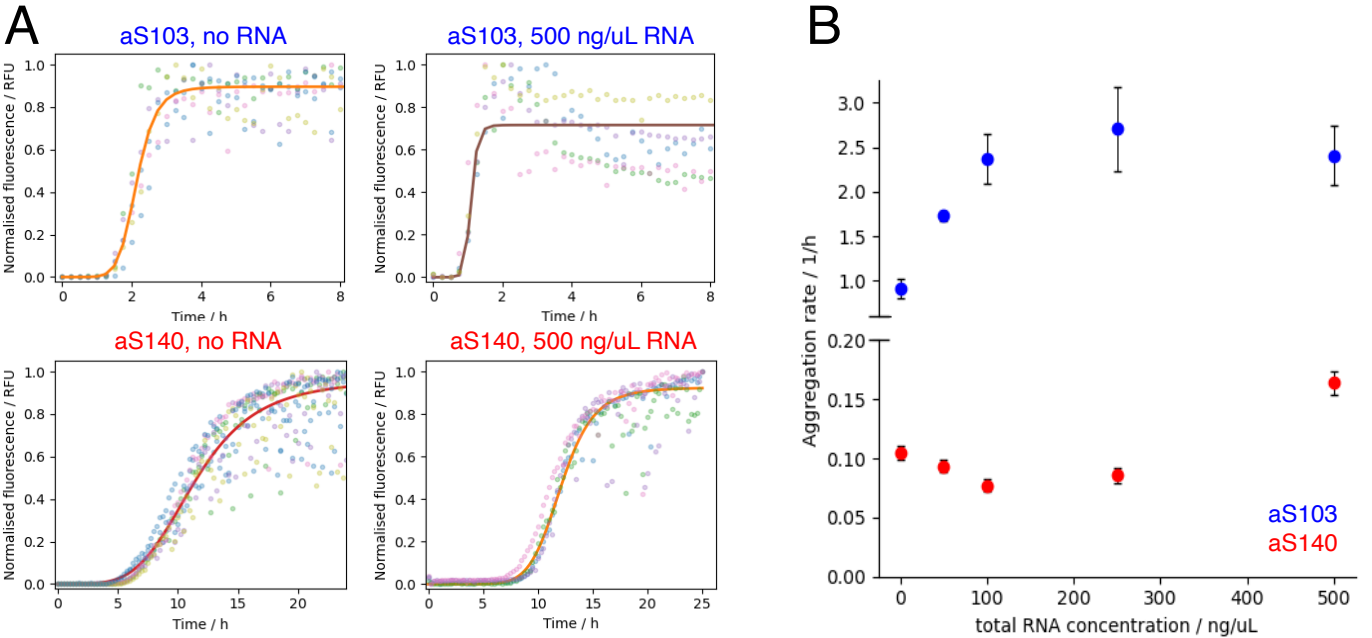
1. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
2. Tartaglia, G. G. *et al.* Prediction of Aggregation-Prone Regions in Structured Proteins. *J. Mol. Biol.* **380**, 425–436 (2008).
3. Polymenidou, M. The RNA face of phase separation. *Science* **360**, 859–860 (2018).
4. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).
5. Louka, A., Zacco, E., Temussi, P. A., Tartaglia, G. G. & Pastore, A. RNA as the stone guest of protein aggregation. *Nucleic Acids Res.* **48**, 11880–11889 (2020).
6. 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).
7. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
8. Zacco, E. *et al.* Probing TDP-43 condensation using an in silico designed aptamer. *Nat. Commun.* **13**, 3306 (2022).
9. 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).
10. Lester, E. *et al.* Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron* **109**, 1675-1691.e9 (2021).
11. 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).
12. Zhang, X. *et al.* RNA stores tau reversibly in complex coacervates. *PLOS Biol.* **15**, e2002183 (2017).
13. 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).
14. 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).
15. 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).
16. 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).
17. 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).
18. 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).
19. Uversky, V. N. A Protein-Chameleon: Conformational Plasticity of α-Synuclein, a Disordered Protein Involved in Neurodegenerative Disorders. *J. Biomol. Struct. Dyn.* **21**, 211–234 (2003).
20. Fusco, G. *et al.* Structural basis of synaptic vesicle assembly promoted by α-synuclein. *Nat. Commun.* **7**, 12563 (2016).
21. Hasecke, F. *et al.* Origin of metastable oligomers and their effects on amyloid fibril self-assembly. *Chem Sci* **9**, 5937–5948 (2018).
22. Chen, S. W. *et al.* Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation. *Proc. Natl. Acad. Sci.* **112**, E1994–E2003 (2015).
23. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9**, 3609 (2018).
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. Fusco, G. *et al.* Direct observation of the three regions in α-synuclein that determine its membrane-bound behaviour. *Nat. Commun.* **5**, 3827 (2014).
26. 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).
27. 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).
28. 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).
29. Anderson, J. P. *et al.* Phosphorylation of Ser-129 Is the Dominant Pathological Modification of α-Synuclein in Familial and Sporadic Lewy Body Disease*. *J. Biol. Chem.* **281**, 29739–29752 (2006).
30. Sorrentino, Z. A. & Giasson, B. I. The emerging role of α-synuclein truncation in aggregation and disease. *J. Biol. Chem.* **295**, 10224–10244 (2020).
31. Li, W. *et al.* Aggregation promoting C-terminal truncation of α-synuclein is a normal cellular process and is enhanced by the familial Parkinson's disease-linked mutations. *Proc. Natl. Acad. Sci.* **102**, 2162–2167 (2005).
32. Liu, C.-W. *et al.* A Precipitating Role for Truncated α-Synuclein and the Proteasome in α-Synuclein Aggregation: IMPLICATIONS FOR PATHOGENESIS OF PARKINSON DISEASE*. *J. Biol. Chem.* **280**, 22670–22678 (2005).

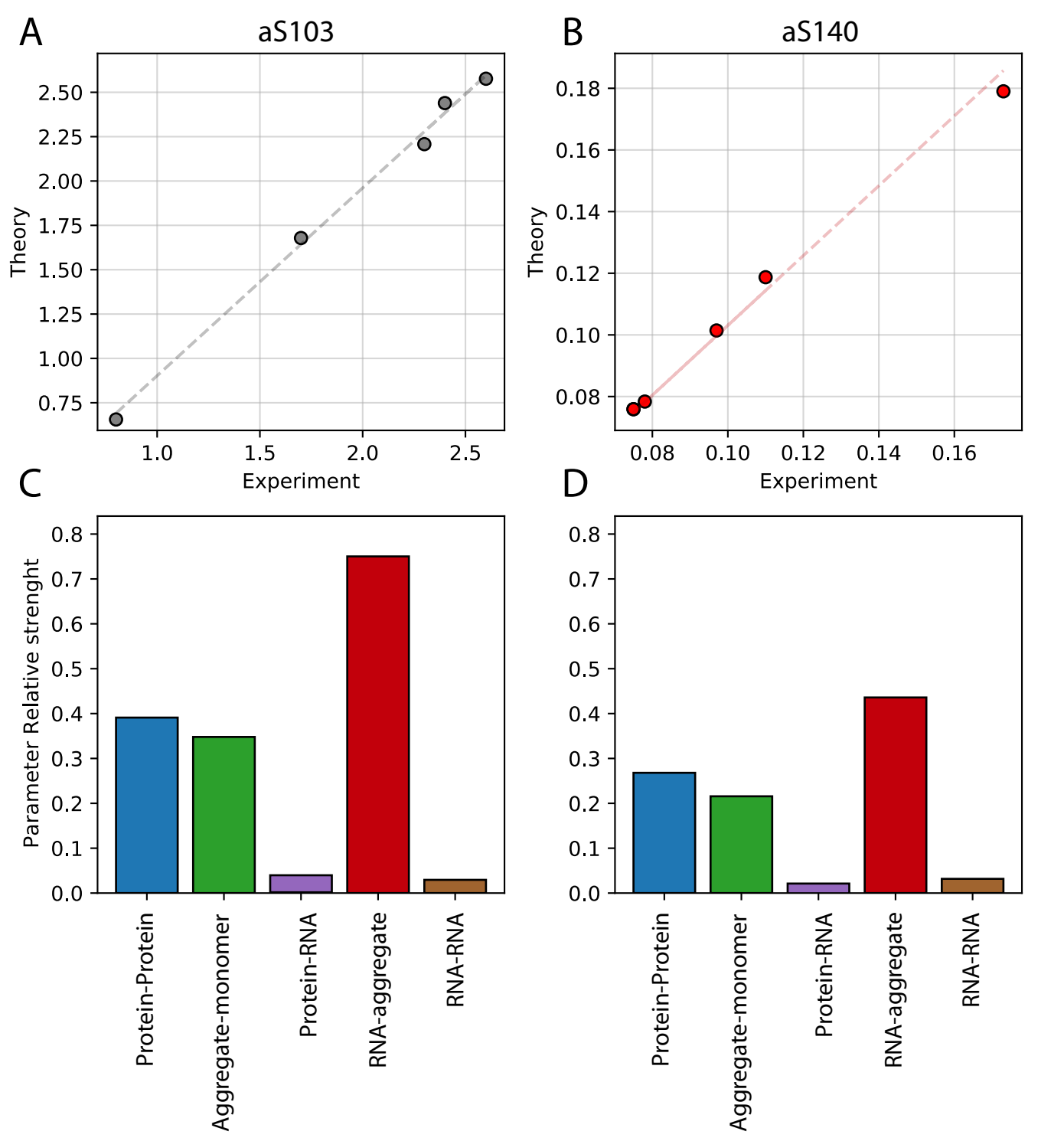
33. Iyer, A. *et al.* C-Terminal Truncated α-Synuclein Fibrils Contain Strongly Twisted β-Sheets. *J. Am. Chem. Soc.* **139**, 15392–15400 (2017).
34. van der Wateren, I. M., Knowles, T. P. J., Buell, A. K., Dobson, C. M. & Galvagnion, C. C-terminal truncation of α-synuclein promotes amyloid fibril amplification at physiological pH. *Chem Sci* **9**, 5506–5516 (2018).
35. Delenclos, M. *et al.* Cellular models of alpha-synuclein toxicity and aggregation. *J. Neurochem.* **150**, 566–576 (2019).
36. 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).
37. 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).
38. Stephens, A. D. *et al.* Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nat. Commun.* **11**, 2820 (2020).
39. 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).
40. Antony, T. *et al.* Cellular Polyamines Promote the Aggregation of α-Synuclein*. *J. Biol. Chem.* **278**, 3235–3240 (2003).
41. 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).
42. Kam, T.-I. *et al.* Poly(ADP-ribose) drives pathologic α-synuclein neurodegeneration in Parkinson's disease. *Science* **362**, eaat8407 (2018).
43. Ray, S. *et al.* α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nat. Chem.* **12**, 705–716 (2020).
44. Hardenberg, M. C. *et al.* Observation of an α-synuclein liquid droplet state and its maturation into Lewy body-like assemblies. *J. Mol. Cell Biol.* **13**, 282–294 (2021).
45. Hoffmann, C. *et al.* Synapsin Condensates Recruit alpha-Synuclein. *Phase Sep. Biol. Dis. Chapter* **433**, 166961 (2021).
46. 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).
47. Siegert, A. *et al.* Interplay between tau and α-synuclein liquid–liquid phase separation. *Protein Sci.* **30**, 1326–1336 (2021).
48. Giasson, B. I. *et al.* Initiation and Synergistic Fibrillization of Tau and Alpha-Synuclein. *Science* **300**, 636–640 (2003).
49. Guo, J. L. *et al.* Distinct α-Synuclein Strains Differentially Promote Tau Inclusions in Neurons. *Cell* **154**, 103–117 (2013).
50. Moussaud, S. *et al.* Alpha-synuclein and tau: teammates in neurodegeneration? *Mol. Neurodegener.* **9**, 43 (2014).
51. Laferrière, F., Claverol, S., Bezard, E., De Giorgi, F. & Ichas, F. Similar neuronal imprint and no cross-seeded fibrils in α-synuclein aggregates from MSA and Parkinson's disease. *Npj Park. Dis.* **8**, 10 (2022).
52. Tartaglia, G. G. & Vendruscolo, M. The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* **37**, 1395–1401 (2008).
53. Roseman, M. A. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.* **200**, 513–522 (1988).
54. Terribilini, M. *et al.* Prediction of RNA binding sites in proteins from amino acid sequence. *RNA* **12**, 1450–1462 (2006).
55. Klus, P. *et al.* The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. *Bioinformatics* **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. 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).
60. Zambrano, R. *et al.* AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **43**, W306–W313 (2015).
61. 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).
62. 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).
63. Waxman, E. A., Mazzulli, J. R. & Giasson, B. I. Characterization of Hydrophobic Residue Requirements for α-Synuclein Fibrillization. *Biochemistry* **48**, 9427–9436 (2009).
64. 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).
65. Cid-Samper, F. *et al.* An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.* **25**, 3422-3434.e7 (2018).

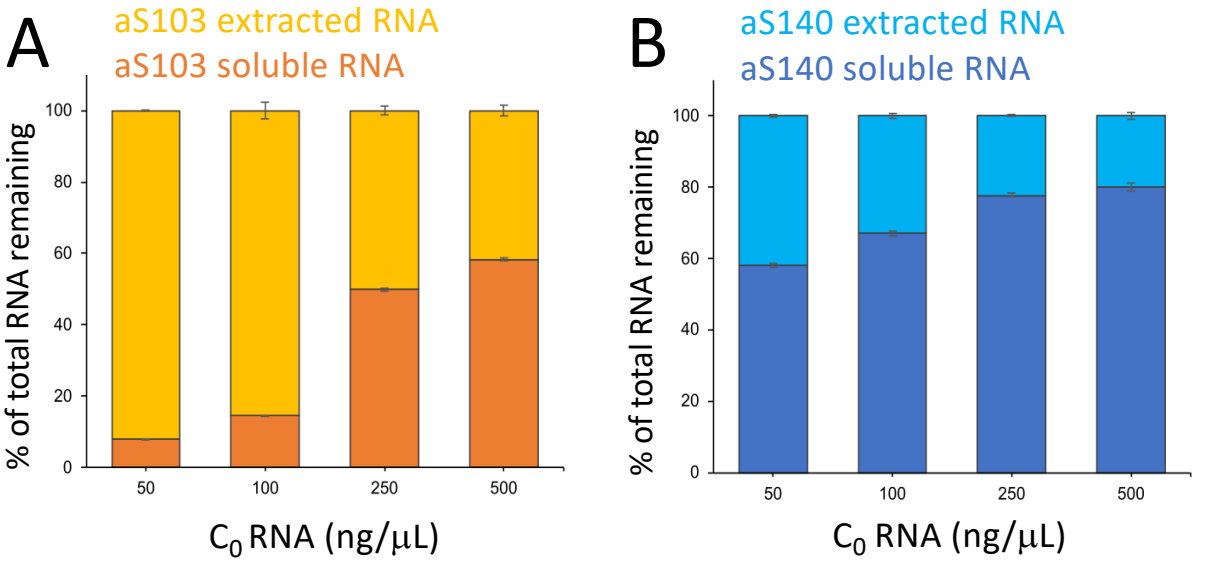
66. Cerase, A., Calabrese, J. M. & Tartaglia, G. G. Phase separation drives X-chromosome inactivation. *Nat. Struct. Mol. Biol.* **29**, 183–185 (2022).
67. Kang, J. *et al.* RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.* **50**, D326–D332 (2022).
68. Galvagnion, C. *et al.* Chemical properties of lipids strongly affect the kinetics of the membrane-induced aggregation of α-synuclein. *Proc. Natl. Acad. Sci.* **113**, 7065–7070 (2016).
69. Zanzoni, A. *et al.* Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. *Nucleic Acids Res.* **41**, 9987–9998 (2013).
70. Cirillo, D. *et al.* Neurodegenerative diseases: Quantitative predictions of protein–RNA interactions. *RNA* **19**, 129–140 (2013).
71. 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).
72. Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. Biochemical Characterization of the Core Structure of α-Synuclein Filaments*. *J. Biol. Chem.* **277**, 19213–19219 (2002).
73. 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).
74. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* **14**, e1002533 (2016).
75. Russo, J. & Russo, I. H. *Techniques and Methodological Approaches in Breast Cancer Research*. (Springer, 2014). doi:10.1007/978-1-4939-0718-2.
76. Maroteaux, L., Campanelli, J. & Scheller, R. Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *J. Neurosci.* **8**, 2804–2815 (1988).
77. Pinho, R. *et al.* Nuclear localization and phosphorylation modulate pathological effects of alpha-synuclein. *Hum. Mol. Genet.* **28**, 31–50 (2019).
78. 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).
79. Jos, S. *et al.* Molecular insights into α-synuclein interaction with individual human core histones, linker histone, and dsDNA. *Protein Sci.* **30**, 2121–2131 (2021).
80. 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).
81. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).
82. Wasmer, C. *et al.* Amyloid Fibrils of the HET-s(218–289) Prion Form a β Solenoid with a Triangular Hydrophobic Core. *Science* **319**, 1523–1526 (2008).

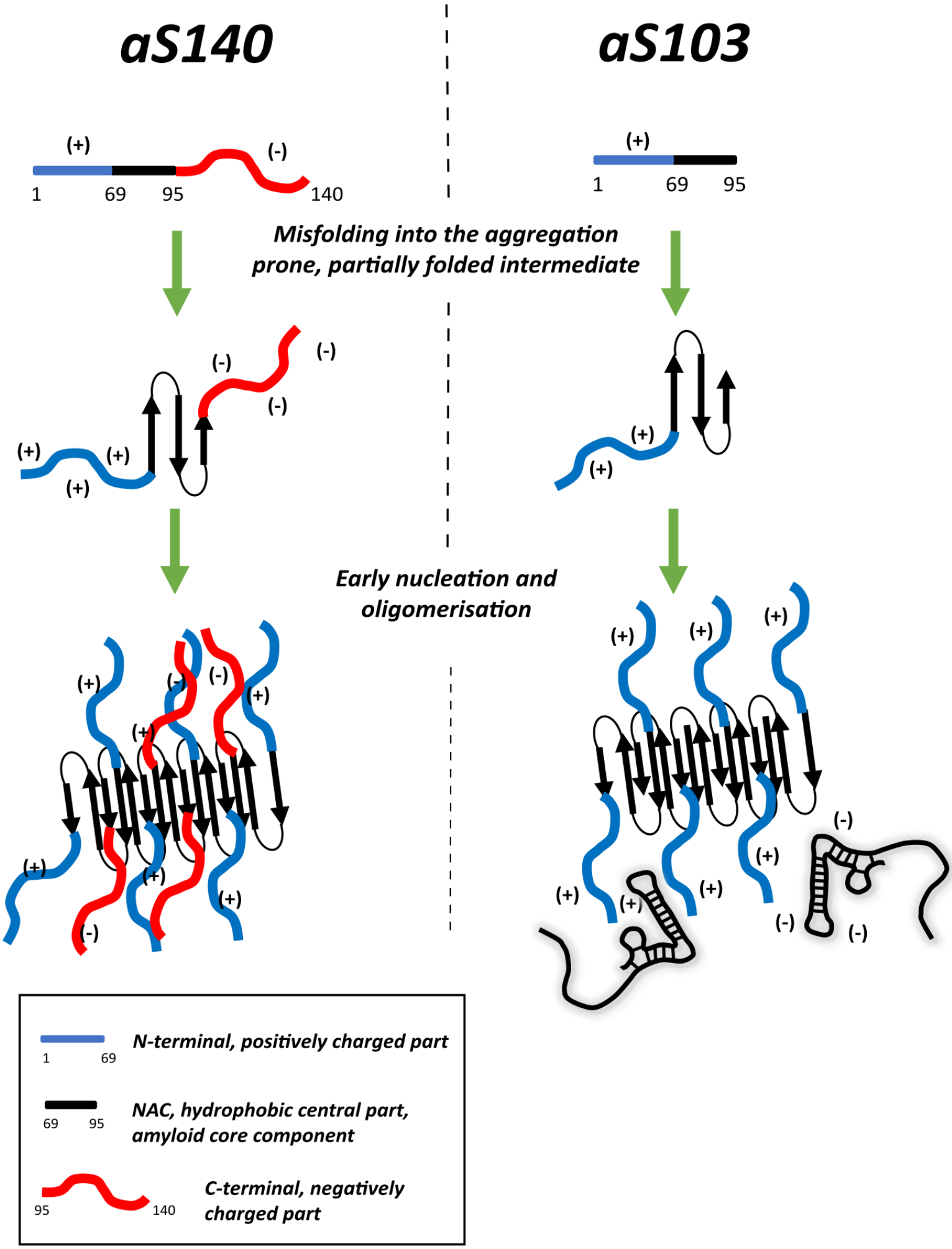

A
Hydrophobicity
Density
0
2
4
6
8
10
0.3
0.4
0.5
0.6
0.7
0.8
0.9
B
RNA Binding Propensity
Density
0
2
4
6
8
10
12
14
0.2
0.3
0.4
0.5
0.6
External Residues
Internal Residues
C
Hydrophobicity
Low
High
D
RNA Binding Propensity
Low
High


A
Net charge WT= -8
Net Charge = 5
Net Charge = -12
Zyggregator Profile
Inner Part of the Amyloid
B
Relative Zyggregator score
Amino Acids Position
C
Δ RNA-binding-propensity
Δ Hydrophobicity
Δ Net Charge
aS140 External
aS103 External
aS140 Internal
aS103 Internal


A
aS103, no RNA
aS103, 500 ng/uL RNA
aS140, no RNA
aS140, 500 ng/uL RNA
Normalised fluorescence / RFU
Time / h
B
Aggregation rate / 1/h
total RNA concentration / ng/uL
aS103
aS140


A
aS103
Theory
Experiment
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
1.0
1.5
2.0
2.5
B
aS140
Theory
Experiment
0.18
0.16
0.14
0.12
0.10
0.08
0.08
0.10
0.12
0.14
0.16
C
Parameter Relative strenght
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Protein-Protein
Aggregate-monomer
Protein-RNA
RNA-aggregate
RNA-RNA
D
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Protein-Protein
Aggregate-monomer
Protein-RNA
RNA-aggregate
RNA-RNA


A
aS103 extracted RNA
aS103 soluble RNA
% of total RNA remaining
100
80
60
40
20
0
50
100
250
500
C0 RNA (ng/µL)
B
aS140 extracted RNA
aS140 soluble RNA
% of total RNA remaining
100
80
60
40
20
0
50
100
250
500
C0 RNA (ng/µL)


aS140
aS103
(+)
(-)
1
69
95
140
Misfolding into the aggregation prone, partially folded intermediate
Early nucleation and oligomerisation
N-terminal, positively charged part
NAC, hydrophobic central part, amyloid core component
C-terminal, negatively charged part