RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43

Elsa Zacco[1,2,†], Ricardo Graña-Montes[3,†], Stephen R. Martin[4], Natalia Sanchez de Groot[3], Caterina Alfano[5], Gian Gaetano Tartaglia[3,6,7,8] and Annalisa Pastore[1,2,9]

*1 - **UK Dementia Research Institute at King's College London,** London, SE5 9RT, United Kingdom*
*2 - **The Wohl Institute at King's College London,** London, SE5 9RT, United Kingdom*
*3 - **Centre for Genomic Regulation (CRG),** The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain*
*4 - **The Francis Crick Institute,** London, NW1 1AT, United Kingdom*
*5 - **Fondazione Ri.MED,** Palermo, 90133, Italy*
*6 - **Universitat Pompeu Fabra (UPF),** Barcelona 08003, Spain*
*7 - **Institutio Catalan de Recerca I Estudis Avancats (ICREA),** 23 Passeig Lluıs Companys, 08010 Barcelona, Spain*
*8 - **Department of Biology 'Charles Darwin',** Sapienza University of Rome, P.le A. Moro 5, Rome 00185, Italy*
*9 - **Scuola Normale Superiore,** Piazza dei Cavalieri, Pisa, 56126, Italy*

***Correspondence to Gian Gaetano Tartaglia and Annalisa Pastore:** G.G. Tartaglia is to be contacted at: Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain. A. Pastore is to be contacted at: UK Dementia Research Institute at King's College London, London, SE5 9RT, United Kingdom. annalisa.pastore@crick.ac.uk*
https://doi.org/10.1016/j.jmb.2019.01.028
Edited by Sheena Radford

Abstract

Amyotrophic lateral sclerosis and frontotemporal lobar degeneration are incurable motor neuron diseases associated with muscle weakness, paralysis and respiratory failure. Accumulation of TAR DNA-binding protein 43 (TDP-43) as toxic cytoplasmic inclusions is one of the hallmarks of these pathologies. TDP-43 is an RNA-binding protein responsible for regulating RNA transcription, splicing, transport and translation. Aggregated TDP-43 does not retain its physiological function.

Here, we exploit the ability of TDP-43 to bind specific RNA sequences to validate our hypothesis that the native partners of a protein can be used to interfere with its ability to self-assemble into aggregates. We propose that binding of TDP-43 to specific RNA can compete with protein aggregation. This study provides a solid proof of concept to the hypothesis that natural interactions can be exploited to increase protein solubility and could be adopted as a more general rational therapeutic strategy.

Introduction

The involvement of protein misfolding and aggregation in several neurodegenerative disorders, including Alzheimer and Parkinson diseases, Huntington's chorea and several amyloidosis, has been clearly established over the past 20–30 years [1]. Only recently, RNA and its interaction with proteins have been recognized important players in neurodegeneration [2,3]. Primary transcripts containing expanded nucleotide repeats form intranuclear RNA *foci* where RNA-binding proteins are sequestered and inactivated, as observed in Huntington's, several spinocerebellar ataxias, myotonic dystrophy and C9orf72 [4,5]. RNA–protein interactions facilitate the conversion of the α-rich and prion protein (PrPC) into its infectious β-rich insoluble form (PrPSc) in Creutzfeldt–Jakob's disease [6]. In Alzheimer's disease, the translation of

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

aggregation-prone proteins is regulated by iron-dependent ribonucleo-protein interactions [7]. In many instances, the aggregation of RNA-binding proteins is triggered by a defective transcription and affects the processing of RNA molecules, which leads to progressive cell death [8]. Many RNA-binding proteins involved in neurodegenerative diseases bind to their cognate mRNAs to inhibit translation through a negative feedback loop of auto-regulation: when a protein concentration is high, binding to its transcript may inhibit further translation [2,9–11]. The ability of nucleic acids to exert a chaperone activity has only recently been described [12,13].

Here, we focused on understanding the role of RNA binding on the behavior of TDP-43, a protein that is able of binding to thousands of mRNA transcripts [14,15]. Among other essential physiological functions, TDP-43 participates in mRNA transcription, splicing, stabilization, translation, transportation and degradation [16]. The cytoplasmic aggregation of TDP-43 is considered one of the important causative factors for two incurable neurodegenerative diseases, the amyotrophic lateral sclerosis (ALS) and the frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-U) [17]. Around 40 different TDP-43 mutations are associated with ALS and FTLD-U [18] (https://www.molgen.ua.ac.be/FTDMutations) that are associated with increased TDP-43 insolubility, cytoplasmic localisation and neurotoxicity [19], but the precise mechanism is unknown. Interestingly, TDP-43 plays an important physiological role in the transport of mRNAs from the nucleus to the cytosol. In ALS and FTLD-U, the protein accumulates in ubiquitinated inclusions in the neuronal cytoplasm and its levels are often reduced in the nucleus [20].

TDP-43 is a modular protein that contains an N-terminal region with a nuclear localization signal, two tandem highly conserved RNA-recognition motifs (RRM), and a disordered C-terminus containing a glycine-rich low-complexity region [21]. The RRM domains recognize preferentially single-stranded UG or TG-rich nucleotide sequences [22]. We recently demonstrated that the isolated RRM tandem repeats are able to aggregate and misfold also in the absence of the remaining protein context [23]. This evidence is at variance with previous believes that the TDP-43 aggregation is solely mediated by the C-terminus. Conversely, consistent with its role in nucleotide binding and processing, increasing evidence suggests that aggregation of TDP-43 is strongly influenced by the interaction with DNA/RNA, the natural partner of this protein. Dissecting the impact of RNA binding on aberrant self-assembly is thus of the utmost interest, since it could open a completely unexplored avenue toward the design of novel therapies.

The precise effect of nucleic acids binding on the aggregation of TDP-43 remains controversial. Some authors have reported that cognate DNA and RNA can prevent the formation of TDP-43 oligomers and larger aggregates [24]. The enhancement of TDP-43 solubility upon binding to the 3′UTR of its cognate mRNA was also reported [25]. Conversely, other researchers demonstrated that RNA can induce TDP-43 fragments to adopt a misfolded conformation that appears to be highly toxic [6]. However, clarification of this issue could bring further support to the hypothesis that natural interactions can be exploited as a powerful basis for designing molecules able to compete specifically with protein aggregation [26,27], a concept that is quickly catching on. The identification of natural interaction partners that promote the acquisition or maintenance of protein functional native states, acting as small molecule probes or pharmacological chaperones, is an emerging field with potentialities as a strategy to modulate or disrupt detrimental aggregation [28,29].

To clarify the issue, we developed a hybrid approach that makes use of several spectroscopic and microscopy techniques to accurately define the effects of a high-affinity UG-rich aptamer derived from analysis of iCLIP sites [30], its reverse complementary sequence, and variants with different lengths, on TDP-43 aggregation. We show how RNA aptamers are able to interfere with the TDP-43 aggregation kinetics as a function of their nucleotide composition, binding affinity and length. We also demonstrate that specific RNAs that bind tightly to the sequence can effectively abrogate the deleterious aggregation of the protein. Our work lays the ground toward understanding the mechanism that determines ALS and FTLD-U. It also provides a solid proof of concept to the hypothesis that natural interactions can be exploited to increase protein solubility. A detailed characterization of already known natural interaction partners of aggregation-prone proteins could thus inform the design and betterment of molecular mimics with properties of specific anti-aggregation agents which could, in the future, be used to develop anti-aggregation drugs.

Results

Choice of the sequences

A detailed biophysical characterization of the interactions between TDP-43 and its RNA partners is needed to gain a comprehensive understanding on the interplay between the functional interactions of the protein with nucleic acids and its deleterious deposition. Addressing such characterization is challenging because it requires the protein to be stable in its soluble state for several hours, whereas it is known that full-length TDP-43 readily aggregates within minutes *in vitro* [31]. In the search for a TDP-43 construct that allowed its biophysical analysis without compromising its aggregation potential, we tested a number of constructs all containing one or both RRM repeats. Their expression showed different behaviors and

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

Table 1. RNA sequences employed in this study

Entry	Name	Sequence (5′ ➔ 3′)	Description
A	RNA_{12}	GUGUGAAUGAAU	UG-rich sequence naturally recognized by TDP-43
B	$NegRNA_{12}$	AUUCAUUCACAC	Reverse and complementary of RNA_{12}
C	$RNA_{12}S1$	UAGAGUAUGAUG	Scrambled version of RNA_{12} 1
D	$RNA_{12}S2$	AAGUGUGGUAUA	Scrambled version of RNA_{12} 2
E	$NegRNA_{12}S1$	UUAUACCUACCA	Scrambled version of $NegRNA_{12}$ 1
F	$NegRNA_{12}S2$	CUCACAUUAUCA	Scrambled version of $NegRNA_{12}$ 2
G	RNA_{36}	$(GUGUGAAUGAAU)_3$	3-fold repetition of RNA_{12} sequence
H	$NegRNA_{36}$	$(AUUCAUUCACAC)_3$	3-fold repetition of $NegRNA_{12}$ sequence

The table reports the name attributed to each molecule, the sequence defined from 5′ ➔ 3′ and a short description. We report in the first column an identifier which will be used throughout the text to identify a specific panel of a figure with an aptamer.

solubility properties (Fig. S1). After this initial testing, we focused on a construct containing the N-terminus and both RRM domains, henceforth N-RRM1-2, which has the advantage to allow us to test the aggregation behavior of the less characterized N-terminus of TDP-43 and compare it with that of the isolated tandem RRM domains [23]. The RNA aptamers were designed as variants of a 12-nucleotide single-stranded RNA (RNA_{12}) already known to tightly and specifically bind the TDP-43 RRM tandem repeats [30] (Table 1A). As a control, we used the RNA_{12} reverse and complementary sequence, $NegRNA_{12}$ (Table 1B).

Heat induces a transition toward amyloid-like β-rich conformations in N-RRM1-2

We used circular dichroism (CD) to investigate the secondary structure and thermal stability of N-RRM1-2 measuring spectra at 20 °C before and after heat treatment (Fig. 1A). The CD spectrum of N-RRM1-2 at 20 °C is typical of a well-folded protein, although with a relatively low ellipticity (-4000 deg cm^2 $dmol^{-1}$, which represents ca. 12% of α-helical content). This secondary structure is in excellent agreement with what it would be expected from the secondary structure of the N-terminal domain and the two RRM domains [30,32]. When heat was increasingly applied up to 90 °C, we observed a conformational transition with an apparent melting temperature (T_m) of 54.5 ± 1.4 °C (Fig. 1B). A spectrum recorded lowering again the temperature to 20 °C showed that the transition is irreversible with a significant increase of β-sheet content, indicating that the sample undergoes a folded-to-aggregated transition rather than a reversible folded-to-unfolded one. This is in agreement with previous studies [33].

To establish whether the β-rich structure formed after thermal denaturation is amyloid-like, we exploited the ability of the fluorescent dye thioflavin T (ThT) to preferentially bind amyloid structures proportionally to the amount of aggregates present in the sample (Fig. S2). The fluorescence intensity associated with N-RRM1-2 in the presence of ThT before heat treatment resulted negligible but increased almost 100 times after heat treatment.

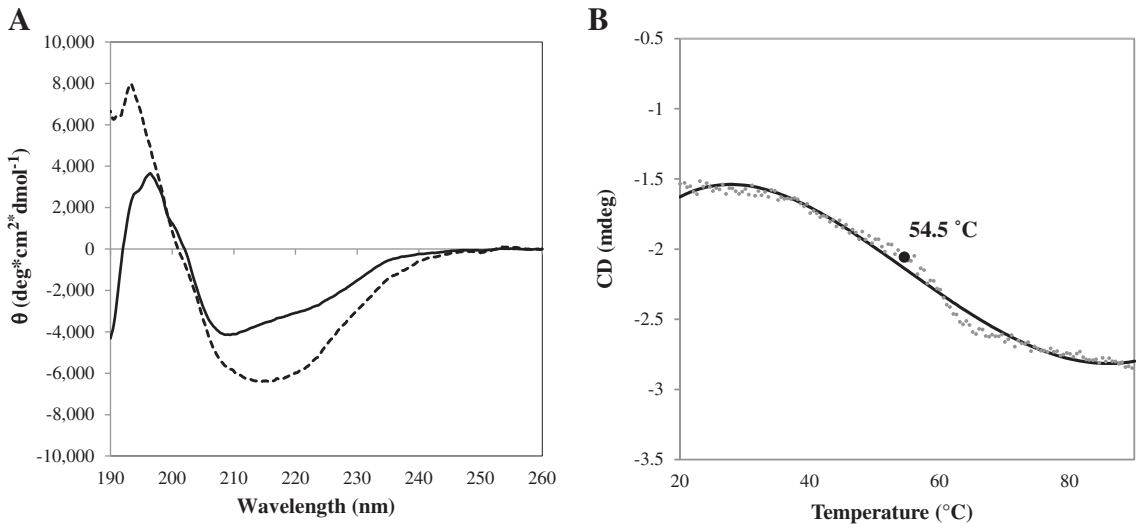

Fig. 1. CD analysis of N-RRM1-2. (A) CD spectrum acquired at 20 °C before heat treatment (continuous line) and after heat treatment (dashed line). (B) N-RRM1-2 melting curve determined according to the variation of the intensity of the minimum at 222 nm as a function of temperature. The melting temperature (T_m) found for N-RRM1-2 is 50.6 °C. An average of three measurements is reported.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

Binding to RNA_{12} or its reverse complementary has opposite effects on the aggregation of N-RRM1-2

To test the affinity of N-RRM1-2 for UG-rich sequences, we measured the apparent dissociation constant K_d of the interaction of N-RRM1-2 with the UG-rich RNA_{12} by biolayer interferometry (BLI). The interaction of N-RRM1-2 with RNA_{12} resulted in a K_d of 6.3 nM (Fig. 2A, Table 2A). Binding of N-RRM1-2 with $NegRNA_{12}$ occurred with a K_d of 1.9 μM (Fig. 2B, Table 2B). These data indicate a tight recognition of the construct with RNA_{12}. The decrease in strength of binding of almost 500-fold with $NegRNA_{12}$ compared to RNA_{12} confirms that the interaction is sequence-specific but also shows that TDP-43 may interact, albeit more weakly, with non-UG-rich RNA sequences.

We next sought to explore the possible effects of these interactions on the aggregation kinetics (Fig. 3). We reasoned that, although ThT is the most commonly used dye to report on amyloid-like aggregation, its strong binding to nucleic acids may interfere with the assay [34,35]. To prevent this possibility, we used the dye ProteoStat. This dye binds to RNA only weakly and did not interfere with the aggregation kinetics, offering a more specific response (Fig. S3). At 37 °C and under a constant agitation, N-RRM1-2 displayed aggregation kinetics characterized by an exponential increase that reaches the fluorescence emission intensity plateau within 24 h, without any noticeable lag phase. Of the initial 15 μM N-RRM1-2 incubated, around 12.5 μM was found aggregated by the end of the assay. When repeating the assay in the presence of RNA_{12}, the kinetics strongly slowed down in a concentration-dependent manner (Fig. 3A): a 1:1 protein/RNA molar ratio induced a noticeable decrease of the slope with a maximum fluorescence associated with ProteoStat of ca. 50% that observed for the isolated N-RRM1-2. This reduction of aggregation was more evident at N-RRM1-2/RNA_{12} molar ratios 1:2 and 1:4, when the apparent degree of aggregate formation was reduced four times, reaching a maximal fluorescence intensity of ca. 25% that of the isolated N-RRM1-2. At these ratios, no plateau was reached within the considered timeframe. Only 7, 3 and 2.5 μM of the protein formed aggregates when incubated with 1:1, 1:2 and 1:4 N-RRM1-2/RNA_{12} ratios, respectively, to be compared to the 12.5 μM N-RRM1-2 found aggregated in the absence of RNA.

The presence of $NegRNA_{12}$ had a strikingly different effect (Fig. 3B): at 1:1 and 1:2 protein/RNA ratios, the aggregation kinetics was not altered. At a 1:4 ratio, the N-RRM1-2 aggregation signal reached 80% of its maximum after only 3 h, compared to 9 h required in the absence of RNA. The slope of the curve increased, although the maximum fluorescence intensity associated with the amount of aggregated N-RRM1-2 did not vary. The amount N-RRM1-2 depleted from the solution in aggregates at the end of the assay was not significantly influenced by the presence of $NegRNA_{12}$ (varying from 12.8 to 13.3 μM at increasing $NegRNA_{12}$ concentration).

To check whether the effects observed on N-RRM1-2 aggregation were simply electrostatically mediated or more specifically controlled by the nucleic acids, we repeated the assay in the presence of increasing concentrations of heparin. It might be arguable that a certain minimal effect of heparin on the total amount of aggregate formed (possible reduction of ca. 25%), which anyway is not concentration dependent, but we did not observe any appreciable effect on the aggregation kinetics (Fig. S4).

These results demonstrate that tight binding of N-RRM1-2 to short UG-rich sequences greatly reduces its aggregation. By contrast, the increased aggregation kinetics promoted by high $NegRNA_{12}$/protein molar ratios indicates that the presence of low-affinity interactions may conversely promote aggregation.

Both recognition specificity and binding strength influence N-RRM1-2 solubility

To investigate whether the contrasting effects of RNA_{12} and $NegRNA_{12}$ on the aggregation of N-RRM1-2 are associated with the RNA sequence or only with the nucleic acid composition, we used randomly scrambled RNA sequences from RNA_{12} ($RNA_{12}S1$ and $RNA_{12}S2$) and from $NegRNA_{12}$ ($NegRNA_{12}S1$ and $NegRNA_{12}S2$). The K_d values of the interactions with $RNA_{12}S1$ and $RNA_{12}S2$ as determined by BLI were still in the low nanomolar range (60 and 75 nM, respectively), but not as tight as for the N-RRM1-2/RNA_{12} pair (Fig. 2C–D, Table 2C–D). Likewise, the K_d values of N-RRM1-2 with $NegRNA_{12}S1$ and $NegRNA_{12}S2$ were comparable to that obtained for the parental $NegRNA_{12}$ (Fig. 2E–F, Table 2E–F). These data suggested that the binding affinity of N-RRM1-2 with these RNAs is mainly determined by the type of bases available but also partially influenced by the nucleic acid sequences. It is also interesting to observe that all sequences had the same percentage of uracil content (30%) and that the main difference between the RNA_{12} and $NegRNA_{12}$ families was the absence of guanine in the latter, where this nucleotide is replaced by a cytosine. This could suggest a role of guanine in binding and an insufficient capacity of uracil to stabilize this protein–RNA interaction.

We also evaluated the effect of the scrambled sequences on the aggregation kinetics of N-RRM1-2. As compared to RNA_{12}, both $RNA_{12}S1$ and $RNA_{12}S2$ displayed an inhibitory effect on aggregate formation that was ca. 20% weaker compared to the parental RNA_{12} (Fig. 4A–B). The slopes became less steep in a concentration-dependent fashion. In contrast, the scrambled versions of $NegRNA_{12}$ exerted no effect on N-RRM1-2 aggregation (Fig. 5A–B). We can thus conclude that both composition and sequence are relevant for RNA binding and aggregation inhibition.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

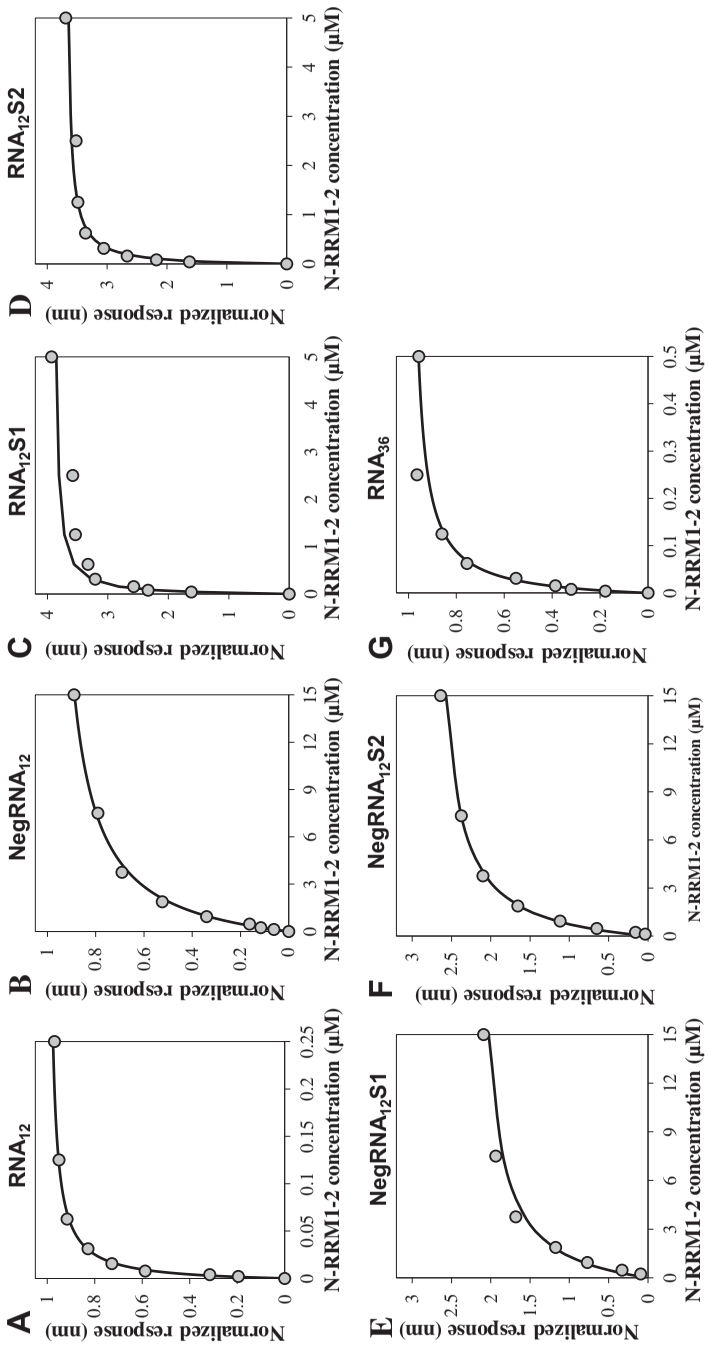

Fig. 2. BLI curves of N-RRM1-2 binding to sensors saturated with different RNA molecules. (A) N-RRM1-2–RNA_{12}, (B) N-RRM1-2–$NegRNA_{12}$, (C) N-RRM1-2–$RNA_{12}S1$, (D) N-RRM1-2–$RNA_{12}S2$, (E) N-RRM1-2–$NegRNA_{12}S1$, (F) N-RRM1-2–$NegRNA_{12}S2$, (G) N-RRM1-2–RNA_{36}. No binding curve was defined by N-RRM1-2–$NegRNA_{36}$ association. Ideal N-RRM1-2 concentration for each association was adjusted in accordance to the strength of each binding.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

Table 2. Summary of K_d for each association pair

Entry	Association pair	K_d (μM)
A	N-RRM1-2–RNA_{12}	6.3×10^{-3}
B	N-RRM1-2–$NegRNA_{12}$	1.9
C	N-RRM1-2–RNA_{12}S1	60×10^{-3}
D	N-RRM1-2–RNA_{12}S2	75×10^{-3}
E	N-RRM1-2–$NegRNA_{12}$S1	1.65
F	N-RRM1-2–$NegRNA_{12}$S2	1.34
G	N-RRM1-2–RNA_{36}	22×10^{-3}
H	N-RRM1-2–$NegRNA_{36}$	ND

The values are all expressed in μM for conformity. The values are to be considered with a ± 10% error due to technique limitations. ND, not determined.

RNA interferes with N-RRM1-2 aggregation in a length-dependent manner

To evaluate whether RNA length has any influence on the aggregation kinetics of N-RRM1-2, we designed longer versions of RNA_{12} and $NegRNA_{12}$ consisting in 3-fold repetitions of the sequences (RNA_{36} and $NegRNA_{36}$). The K_d of RNA_{36} for N-RRM1-2 was ca. 22 nM (Fig. 2G, Table 2G), that is, three to four times weaker than that of RNA_{12}. This affinity reduction could be explained formation of secondary structure and/or by steric hindrance. However, the first explanation is rather unlikely since RNA_{12} repeats are predicted by the Vienna software not to introduce any secondary structure (https://goo.gl/fRZVC2), as also seen in the complex of RRM1-2 with the same RNA [30]. Binding to $NegRNA_{36}$ was so weak that a K_d could not be determined under the set experimental conditions (Table 2H). Both for RNA_{36} and $NegRNA_{36}$, increase of the RNA length seemed to diminish the likelihood of recognition/interaction.

The influence of the RNA length on N-RRM1-2 aggregation did not directly correlate with the binding affinity: despite RNA_{36} exhibited a weaker binding as compared to RNA_{12}, it efficiently prevented N-RRM1-2 aggregation even at 1:1 N-RRM/RNA_{36} ratio (Fig. 6A)

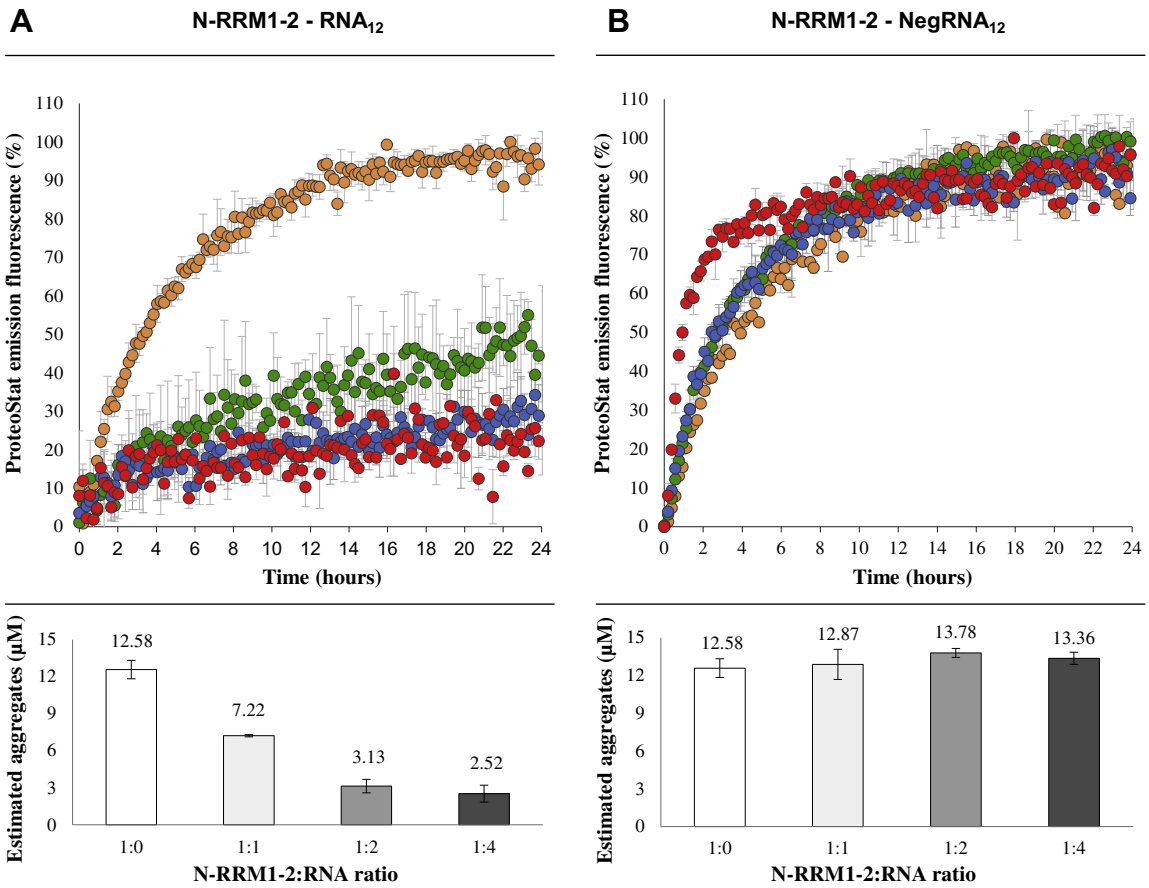

Fig. 3. Kinetics of aggregation and amount of aggregated N-RRM1-2 after 24-h incubation by itself (yellow circles), with protein/RNA ratio = 1:1 (green circles), with protein/RNA ratio = 1:2 (blue circles) and with protein/RNA ratio = 1:4 (red circles). (A) Top panel: N-RRM1-2 with increasing amount of RNA_{12}; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with RNA_{12}. (B) Top panel: N-RRM1-2 with increasing amount of $NegRNA_{12}$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $NegRNA_{12}$. The kinetics are represented proportionately to the variation of the emission fluorescence of ProteoStat™ as a function of time. The fluorescence values are expressed in percentage.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

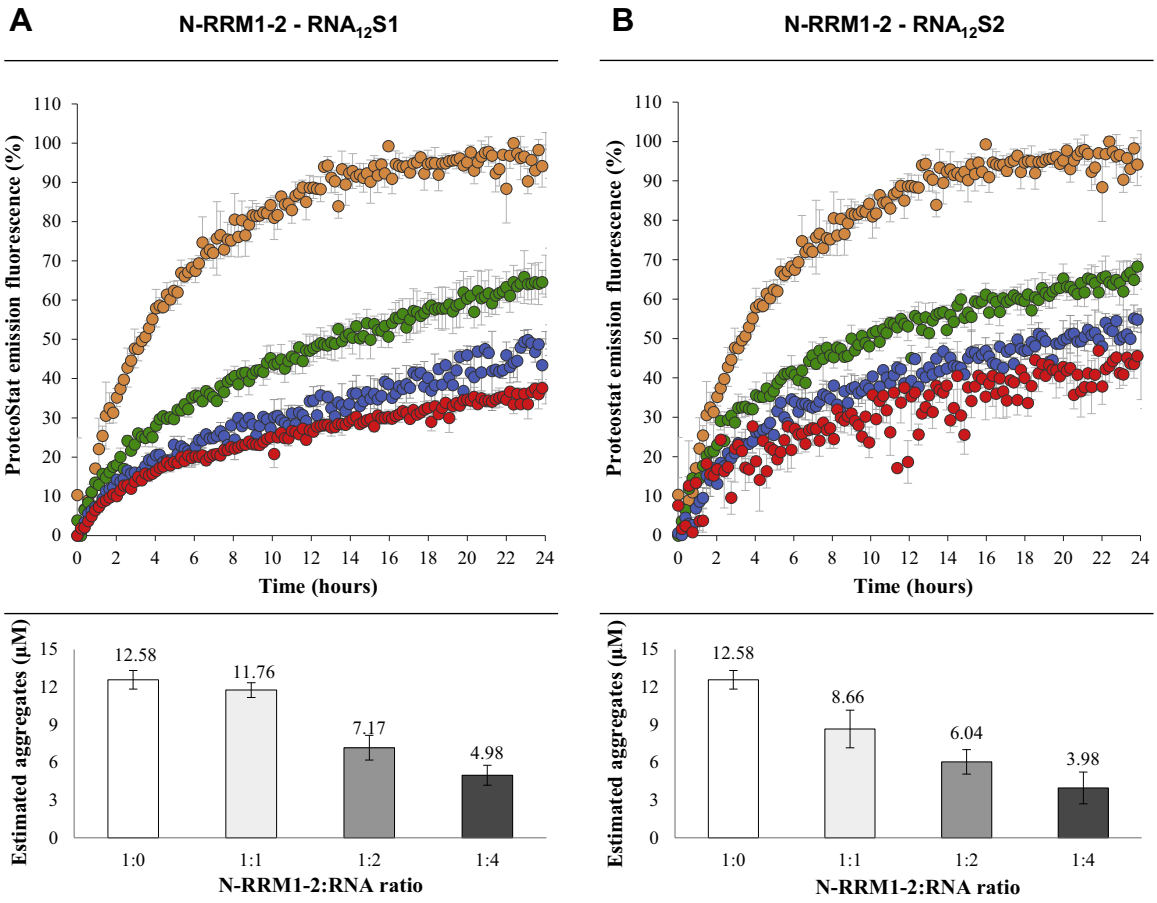

Fig. 4. Kinetics of aggregation and amount of aggregated N-RRM1-2 after 24-h incubation by itself (yellow circles), with protein/RNA ratio = 1:1 (green circles), with protein/RNA ratio = 1:2 (blue circles) and with protein/RNA ratio = 1:4 (red circles). (A) Top panel: N-RRM1-2 with increasing amount of $RNA_{12}S1$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $RNA_{12}S1$. (B) Top panel: N-RRM1-2 with increasing amount of $RNA_{12}S2$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $RNA_{12}S2$. The kinetics are represented proportionately to the variation of the emission fluorescence of ProteoStat™ as a function of time. The fluorescence values are expressed in percentage.

with no increment in the reporter dye fluorescence. Virtually all protein was recovered in its soluble form at assay completion. A drastically opposite effect was observed upon titration with $NegRNA_{36}$, which promoted aggregation already at a 1:2 molar ratio (Fig. 6B), driving 80% of N-RRM1-2 aggregation within 3 h, effect comparable to the 1:4 N-RRM/$NegRNA_{12}$ ratio. The higher the concentration of $NegRNA_{36}$, the higher was also the rate of precipitation observable also by the naked eye, suggesting not only a faster aggregate formation but also the formation of larger, heavier floccules. This effect is not in discrepancy with the weak affinity observed by BLI, since also weak and unspecific interactions could alter protein aggregation [36].

The results obtained with the longer RNAs clearly demonstrated that the length of the binding partner plays an important role in determining the solubility of N-RRM1-2 in solution.

Different RNAs prefer different N-RRM1-2 conformation for their binding

To elucidate how RNA_{12} and $NegRNA_{12}$ affect the aggregation of N-RRM1-2, we looked at possible changes of the secondary structure by CD by incubating the N-RRM1-2 with the aptamers at 37 °C for 24 h and recorded the complete spectra individually (Fig. 7). At t_0 (Fig. 7A), N-RRM1-2 had a typical secondary structure signature of an α + β protein both in the presence and the absence of RNA. In the absence of RNA, the minimum at ~208 nm broadened and lost intensity after 30 min (Fig. 7B), while a single minimum at 218 nm typical of a β-sheet conformation started to appear; the shift from an α-to-β conformation continued progressively and the spectra became dominated by a β-signal after 24 h (Fig. 7H). In contrast, addition of RNA_{12} induced a longer persistency of the minimum at 208 nm, suggesting a

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

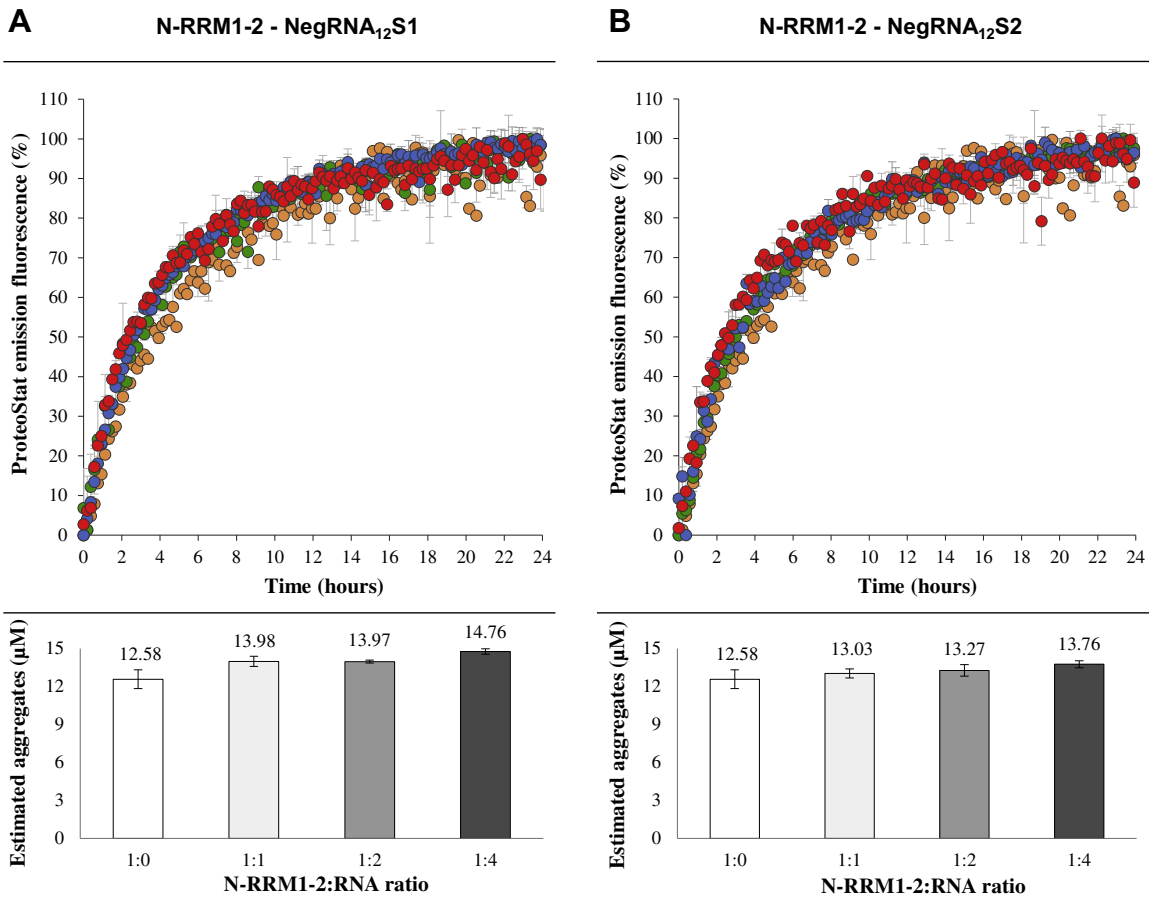

Fig. 5. Kinetics of aggregation and amount of aggregated N-RRM1-2 after 24-h incubation by itself (yellow circles), with protein/RNA ratio = 1:1 (green circles), with protein/RNA ratio = 1:2 (blue circles) and with protein/RNA ratio = 1:4 (red circles). (A) Top panel: N-RRM1-2 with increasing amount of $NegRNA_{12}S1$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $NegRNA_{12}S1$. (B) Top panel: N-RRM1-2 with increasing amount of $NegRNA_{12}S2$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $NegRNA_{12}S2$. The kinetics are represented proportionately to the variation of the emission fluorescence of ProteoStat™ as a function of time. The fluorescence values are expressed in percentage.

stabilization of α-helical content. This signal dominated the protein spectrum for the first 8 h (Fig. 7G), with a slow but progressive loss of the minimum intensity. It was only after 24 h that the α-helical minimum became broad, suggesting the beginning of a conformational transition toward a β conformation.

The N-RRM1-2 spectra in the presence of the $NegRNA_{12}$ reflected an α-to-β transition similar to that observed for the isolated protein, but the transition appeared faster and the β-signal was more evident. A gradual but constant increment of a new maximum at 245 nm became observable between 4 and 24 h. This signal is commonly associated with the stacking of aromatic amino acids, and its gradual increase could indicate a time-dependent change of the environment caused by aggregation and structural conversion. From these results, a stabilizing effect of RNA_{12} toward a more α-helical rich native-like conformation of N-RRM1-2 is evident. $NegRNA_{12}$ induced instead N-RRM1-2 to adopt a β-rich structure.

RNA binding does not alter the aggregate morphology

Furthermore, we assessed the impact of the different RNAs on the morphology of the aggregates formed at the end point of the reaction in the absence and in the presence of a 1:4 protein/RNA ratio. We exploited the presence of the fluorescent ProteoStat by confocal fluorescence microscopy to visualize the overall appearance of the aggregates and transmission electron microscopy (TEM) to study the aggregate morphology. The confocal images showed a clear reduction of the aggregates in the presence of RNA_{12} and RNA_{36} (Fig. 8A, B and D, top panels). Conversely, $NegRNA_{12}$ did not have an appreciable effect (Fig. 8C, top panels). $NegRNA_{36}$ induced dramatic effects leading to the formation of bigger and more localized aggregates (Fig. 8E, top panels).

TEM micrographs revealed the presence of floccular species with a cotton wool appearance (Fig. 8A,

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

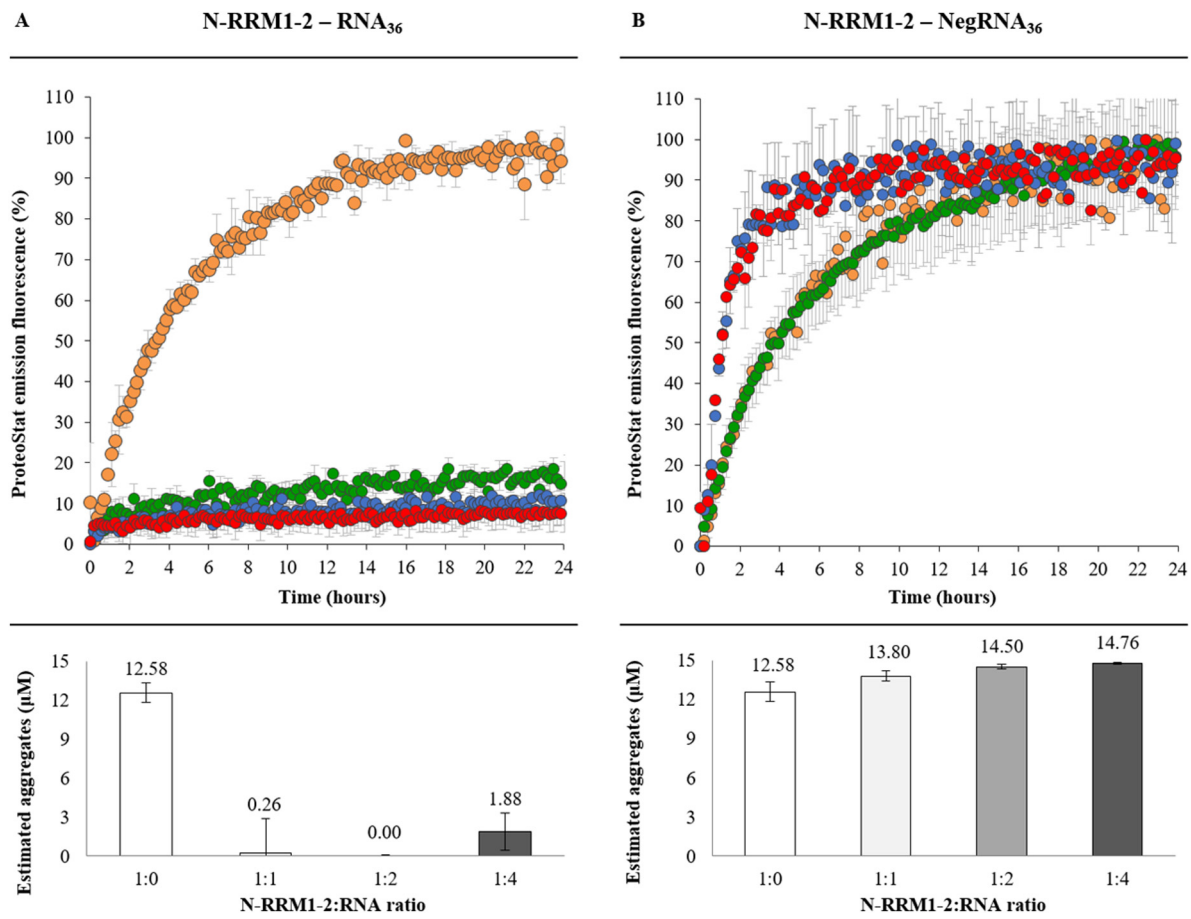

Fig. 6. Kinetics of aggregation and amount of aggregated N-RRM1-2 after 24-h incubation by itself (yellow circles), with protein/RNA ratio = 1:1 (green circles), with protein/RNA ratio = 1:2 (blue circles) and with protein/RNA ratio = 1:4 (red circles). (A) Top panel: N-RRM1-2 with increasing amount of RNA_{36}; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with RNA_{36}. (B) Top panel: N-RRM1-2 with increasing amount of $NegRNA_{36}$; bottom panel: amount of N-RRM1-2 found aggregated at the end of the assay when incubated with $NegRNA_{36}$. The kinetics are represented proportionately to the variation of the emission fluorescence of ProteoStat™ as a function of time. The fluorescence values are expressed in percentage.

bottom panels), somewhat reminiscent of the aggregates reported for SOD1, another protein involved in ALS [37]. The aggregates retained the same morphology in the presence of RNA but became slightly more dense with $NegRNA_{12}$ (Fig. 8B and C, bottom panels). The effects became more evident with the longer RNA aptamers. The aggregates observed in the presence of RNA_{36} were morphologically quite different being distributed as a thin layer without identifiable fibrillar features (Fig. 8D, bottom panels). These structures could be representative of an early stage of the aggregation process. As in the confocal micrographs, $NegRNA_{36}$ drove N-RRM1-2 to form large and compact aggregates that readily disrupted the carbon coating of the grids by TEM (Fig. 8E, bottom panels).

These results suggest that TDP-43 binding to the chosen RNAs does not significantly alter the morphology of the aggregates formed by the N-RRM1-2, but their formation is either enhanced or inhibited depending on the specific nucleic acid sequence.

Predicting the effect of clinically important mutations on RNA binding

Most disease-associated mutations occur in the C-terminus of the protein that is absent in N-RRM1-2. However, two interesting clinically relevant mutations (P112H and D169G) are in RRM1 [18]. We used different approaches to predict the effects of mutations on RNA binding using *cat*RAPID, protein stability through FoldX and protein aggregation with Aggrescan3D. *cat*RAPID uses secondary structure predictions and a potential for the van der Waals and hydrogen bonding contributions. Both D169G (*Z*-score −1) and P112H (*Z*-score −0.5) mutations led to a decrease of the binding propensities (Fig. 9). The P112H mutation was predicted in FoldX to destabilize

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

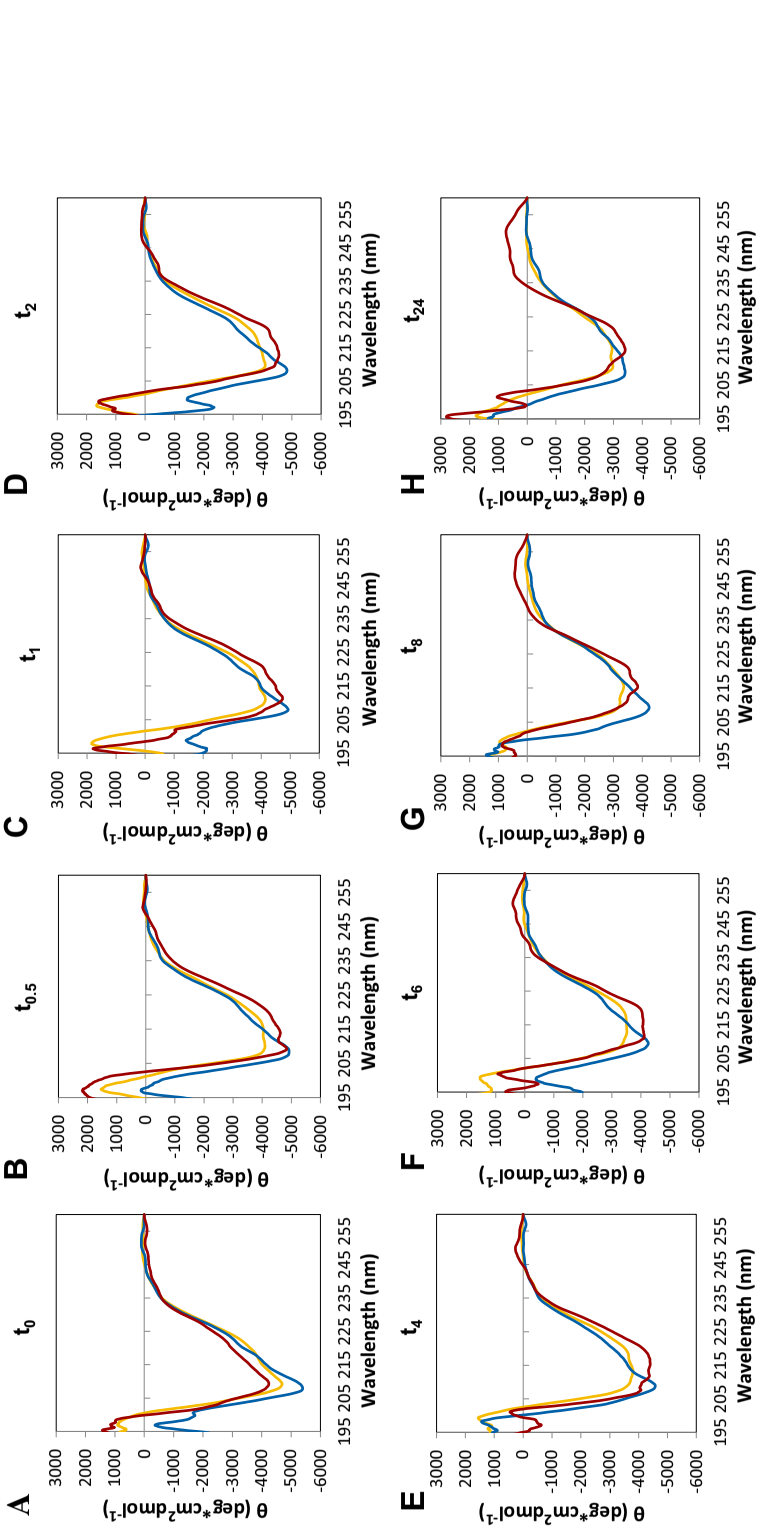

Fig. 7. CD analysis of N-RRM1-2 in the presence of RNA_{12} and $NegRNA_{12}$. Yellow line, N-RRM1-2; blue line, N-RRM1-2/RNA_{12} = 1,4; red line, N-RRM1-2: $NegRNA_{12}$ = 1:4. (A) t_{0h}, (B) $t_{0.5h}$, (C) t_{1h}, (D) t_{2h}, (E) t_{4h}, (F) t_{6h}, (G) t_{8h}, (H) t_{24h}. CD data are acquired at 37 °C and are shown from 260 to 190 nm, blanked and normalized according to protein concentration, light path length and number of peptide bonds. Each spectrum represents an average of 10 acquisitions.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

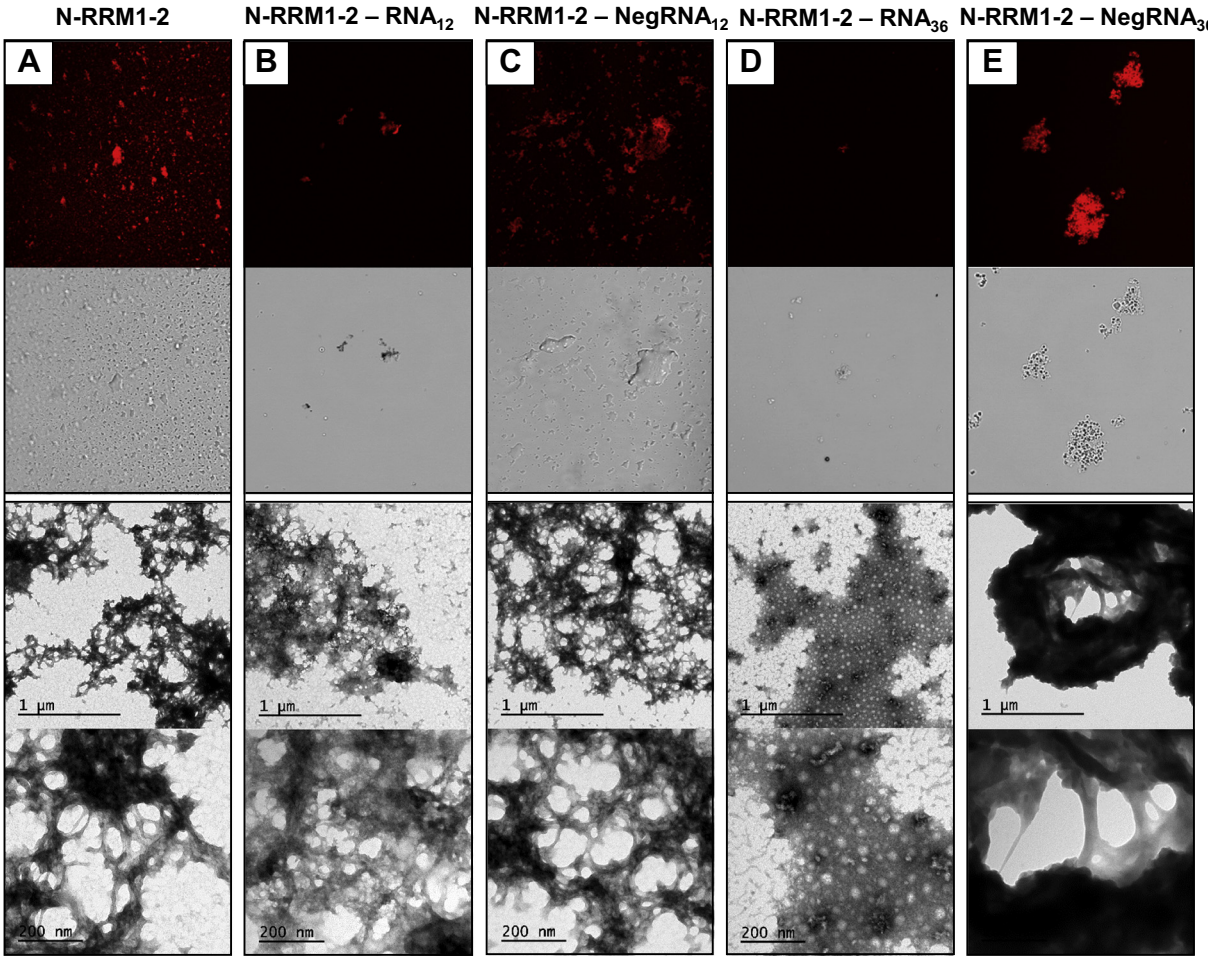

Fig. 8. Study of aggregates' morphology. From top to bottom: confocal microscopy images acquired after aggregation with a 40× magnification, using fluorescence and transmitted light detection, TEM micrographs acquired after aggregation at 5000× and 15,000× magnification. (A) N-RRM1-2, (B) N-RRM1-2–RNA_{12}, (C) N-RRM1-2–$NegRNA_{12}$, (D) N-RRM1-2–RNA_{36}, (E) N-RRM1-2–$NegRNA_{36}$. N-RRM1-2: 15 μM. N-RRM1-2:RNA = 1:4.

the protein ($\Delta\Delta G > 3$ kcal mol^{-1}), while D169G was estimated to slightly increase the stability ($\Delta\Delta G < -1$ kcal mol^{-1}), in good agreement with a mild increase in the melting temperature of isolated RRM1 observed experimentally [23]. In addition, we predicted the structural aggregation propensity corrected by the surface exposure using Aggrescan3D. The calculations suggested that the D169G mutation increases the exposed aggregation propensity, whereas the P112H mutation does not change significantly the aggregation propensity. Thus, the reduced RNA binding propensity of P112H is associated with a decrease in stability and mild increase of aggregation propensity. The mutation D169G stabilizes the structure of the two RRM domains but simultaneously increases the aggregation propensity, which in turn reduces the RNA binding propensity. The mutation seems to result in a persistent exposure of the highly aggregation-prone RNA-binding interface, whose aggregation potential is further increased by the mutation itself, thus promoting the formation of toxic aggregates.

Discussion

While protein aggregation has received attention over the last 25 years, an interest on the effects of nucleic acids in aggregation is relatively new. In this work, we studied the effects of RNA interaction with TDP-43, a protein interesting both for its physiological function and for its involvement in human disease.

To understand the mechanism, we dissected TDP-43 into smaller fragments according to the protein architecture. TDP-43 is a modular protein with an N-terminal domain linked by a presumably unstructured region to two RRM motifs that are followed by a C-terminal domain. As other proteins, TDP-43 seems to have multiple aggregation-prone hotspots: the N-terminus was shown to enhance binding activity to

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

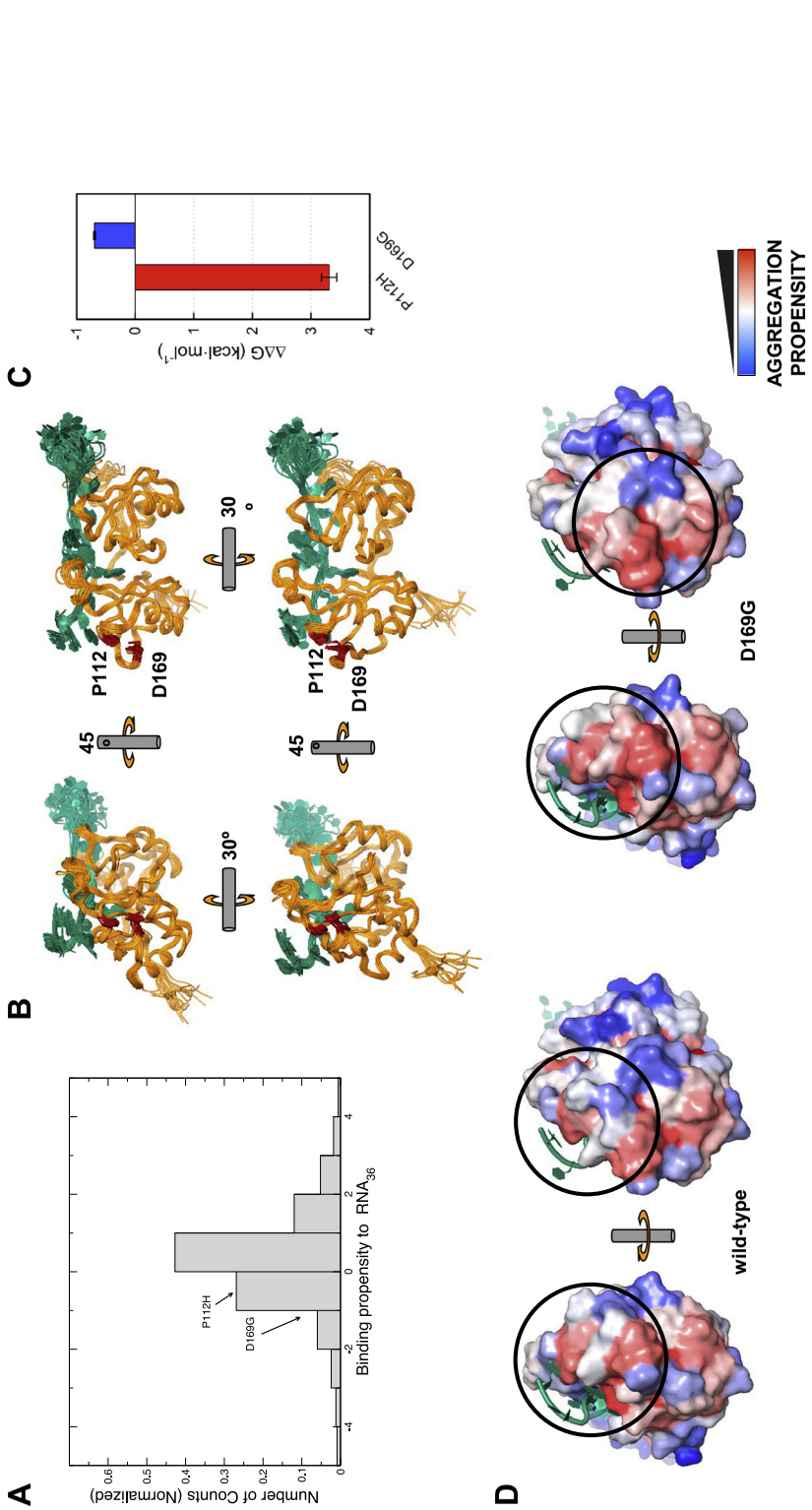

Fig. 9. Prediction of the effects of two clinically relevant mutations. (A) RNA binding affinities computed with *cat*RAPID. The results are presented as the Z-normalized scores with respect to the wild-type. (B) Cartoon representation of the ensemble of 20 conformers of the RRM1-2 resolved in the presence of a high affinity RNA (PDB 4BS2). (C) Prediction of changes in stability upon mutation of the RRM1-2 structure computed with FoldX. (D) Prediction of changes in structural aggregation propensity computed with Aggrescan3d. The black circles highlight the region where D169G in RRM1 induces a shift from low aggregation propensity (blue patch) to medium–high aggregation propensity (light-red patch). (E) The surface of the wild-type RRM1-2 shows overlap between patches with a high aggregation propensity in both RRM1 and RRM2, and the RNA binding site.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

RNA and promote oligomerization and toxicity [38,39]. We recently proved that RRM2 is soluble and aggregation resistant, while RRM1 and the two tandem domains are highly aggregation prone and able to form amyloid-like structures [23]. Other researchers identified sequence stretches in the RRM2 domain with a high potential to form amyloid-like fibrils [40]. The C-terminus of TDP-43 contains a G-rich low-complexity region and a QN-rich motif that confers prion-like properties involved in the physiological phase transition of the protein [41].

We chose to use a construct that encompasses both the N-terminus and the RRM domains taking the start from our previous work on the RRM domains [23]. We decided to include also the N-terminus for various reasons: while the prion-like potential of the C-terminus has been associated to the pathological aggregation of TDP-43 [42], growing evidence indicates the importance of the N-terminally driven oligomerization for physiological self-assembly into stress granules [43,44]. The N-terminal domain is also involved in the native functional dimerization of the protein, which is considered to increase the risk of aberrant self-interactions and aggregation of the protein [45,46]. Literature assessing how TDP-43 dimerization/oligomerization is directed by the N-terminus has shown that this region participates in the physiological role of the protein by enhancing phase transition [47]. Deletions or mutations within the first 10 residues of the N-terminal domain disrupt the deposition of a TDP-43 into cytoplasmic inclusions, resulting in a diffuse cytosolic localization [33]. Oligomerization directed by the N-terminus might thus regulate TDP-43 self-assembly and trigger aggregation, more than the prion-like properties of the C-terminus. A construct similar to that we used was recently shown to foster concentration-dependent oligomerization [25]. All this evidence indicates that N-RRM1-2 constitutes a relevant model to test the impact of RNA binding on the pathological aggregation of full-length TDP-43 and to model the pathophysiological aggregation of TDP-43.

Our data provide precious information on the molecular events occurring as a consequence of RNA binding. To ensure tight and specific binding, we used an RNA aptamer derived from the literature that binds the RRM repeats tightly and specifically [30]. We show that binding to the natural sequence improves TDP-43 solubility. Conversely, the reverse or the scramble sequences have the opposite effect. This is an important result that has a number of important consequences. The effects of RNA or DNA binding on TDP-43 have been described before [24], but none of the previous studies have directly addressed the relationship between nucleic acid binding and protein aggregation in the attempt to understand the molecular mechanisms of disease and used this knowledge to prevent aberrant function. The employment of artificial tags (Flag or vYFP) in the previous studies could also have masked or altered the protein behavior. We have for the first time developed a platform aimed to assess the effects of different RNA molecules on TDP-43 binding and the kinetics of aggregation.

Inclusion of the N-terminus did not significantly affect the binding affinities to RNA, that are comparable but had a noticeable effect on the kinetics of aggregation which are faster and more cooperative for the longer construct: we previously reported that RRM1-2 aggregation curve describes a linear increment and does not reach a plateau within the 48 h of the experiment [23]. Under identical conditions, N-RRM1-2 displays a significantly faster aggregation described by a concave curve, with a plateau at ca. 24 h. This is well in agreement with a role of the N-terminus in oligomerization [43]. In fact, we did not observe a distinguishable lag phase in the binding curves of N-RRM1-2 to the fluorescent dye ProteoStat. This aggregation kinetics suggests the absence of a nucleation step, or at least the concomitant existence of nucleation and aggregate growth [48]. This may indicate the presence of pre-formed soluble oligomers at the very beginning of the assay and is compatible with the reported aggregation curves of other pre-seeded proteins [49,50]. It is also known that both nucleation and aggregate elongation rate determine the slope of the curve [51]. This implies that inhibition of the aggregation will probably translate into a reduction of the speed by which oligomers grow into large aggregates/fibrillar structures. We identified RNA sequences that efficiently slow down TDP-43 aggregation in a sequence-specific manner. The inhibitory effect of specific RNA aptamers on TDP-43 aggregation was evident from aggregation assays and microscopic analysis. The slope of the aggregation curve of N-RRM1-2 in the presence of RNA_{12} and RNA_{36} is dramatically lower, suggesting a severe delay in aggregate growth. Much smaller TDP-43 precipitates were observed by confocal microscopy and TEM. Likewise, Bradford assays confirm that the addition of $NegRNA_{12}$ and $NegRNA_{36}$ to N-RRM1-2 does not increase the amount of total protein aggregates but accelerates the speed of their formation. For this system, in which the total amount of aggregates probably does not vary with time but new aggregates form by incorporation of new monomers into existing aggregates, the effect of RNA is to slow down an inevitable cascade of aggregate growth. The specificity by which certain RNA sequences (and not others) appear to affect N-RRM1-2 aggregation kinetics leads us to think of the existence of a delicately balanced, highly regulated mechanism controlling temporal and spatial interactions between TDP-43 and its RNA binding partners. Any dysregulation of RNA levels in the cell could lead to an increase in TDP-43 aggregation. Likewise, it suggests that any reduction in the RNA binding efficiency of TDP-43, for instance due to mutations, may reduce

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

protein conformational stability and drive it toward a more severe and more toxic aggregation [52].

The K_d values reported in this study as result of our BLI experiments suggested that TDP-43 might interact, albeit less tightly, with non-UG-rich RNAs, and that these interactions could have a destabilizing conformational effect on the protein. Our data also bring new support to the concept of the "conformational selection" model in molecular recognition [53]. According to our results, TDP-43 binding to nucleic acids can be considered a prime example of the "conformational selection" model, in which the protein can arrange itself in a collection of coexisting conformations with different population distribution. Each one of these conformations selectively binds the most suitable partner, shifting the equilibrium toward this state [54,55]. If this theory were proven correct, TDP-43 could display a wide modulation of specificities and affinities in its interactions with nucleic acids: RNA could either drive the protein toward its aggregation, when inducing it to adopt a more disordered conformation, or act as molecular chaperones, by stabilizing a more soluble state [56]. The same concept could in principle be extrapolated to other RNA-binding proteins involved in other human disorders. Several reports describe the effects of nucleic acids on the folding of RBPs. It has, for example, been shown that ribosomal RNA can increase the conformational stability of a protein [57] and that nucleic acid binding can alter the folding of aggregating proteins, leading to their misfolding [58]. The concept of nucleic acids acting as molecular chaperones has been proposed for mRNA in stress granules, when, as a consequence of stress and translation inhibition, RNA is retained with multiple RBPs and chaperones in response to cellular stress [59]. In this light, our results support the thesis that sequence-specific nucleic acids can stabilize RBP structures (e.g., UG repeats and TDP-43), but we propose a new theory that an excess of low-affinity RNA might conversely promote RBP aggregation.

Last, but not least, our results shed further light into the hypothesis that aggregation is the dark side of normal function extending the meaning of this statement. It has been suggested and demonstrated in a number of specific cases that functionally important regions on the surface of a protein that have evolved to recognize specific benign cellular partners may be the same that promote aggregation [26,27]. These observations have highlighted a dualism between normal function and aberrant aggregation and proposed the urgency of studying non-pathologic interactions also in the context of misfolding diseases [60]. Because of its nature of a modular protein with long intrinsically unfolded regions, TDP-43 might provide a complex and interesting paradigm to further develop the concept, which could in turn, in the future, be exploited in the design of RNA-based pharmacological chaperones. The utilization of RNA as a way to increase protein solubility should not be intended as a way to substitute or displace the normal function of the protein since this could be as dangerous as protein aggregation if not more. Future studies aiming at the use of RNA as a solubilizing agent will have to consider instead ways to interfere with aggregation by "chaperoning" the protein to a more stable non-aggregated conformation without interfering with the normal partners. Further studies in cell are thus now needed to assess the effects of these and other RNA sequences on aggregation and physiological function of TDP-43 and other aggregation prone proteins.

Materials and Methods

Protein production

A construct encompassing residues M1-Q269 of the TDP-43 sequence (Uniprot entry: Q13148) which includes both the N-terminus and the two RRM repeats (henceforth N-RRM1-2) was cloned into a pET-SUMO plasmid and expressed in Rosetta2(DE3) *Escherichia coli* cells as fusion protein with a N-terminal SUMO solubilization domain and 6xHis tag. Cells transformed with this plasmid were first grown overnight at 37 °C in Luria–Bertani medium containing 50 μg/ml kanamycin. Cell cultures were then diluted 1:100 in fresh Luria–Bertani medium with 50 μg/ml kanamycin and grown to a 0.7 optical density recorded at 600 nm, before adding 0.5 mM IPTG to induce protein expression overnight at 18 °C. Next, cells were collected by centrifugation at 4000 rcf for 20 min at 4 °C, resuspended in lysis buffer [10 mM potassium phosphate buffer (pH 7.2), 150 mM KCl, 5 mM imidazol, 5% v/v glycerol, 1 mg/ml lysozyme, cOmplete™ EDTA-free Protease Inhibitor tablet by Roche, 1 μg/ml DNase I and 1 μg/ml RNaseA], and disrupted by sonication. The soluble fusion protein was recovered in the supernatant by centrifugation at 70,000 rcf for 45 min at 4 °C, and isolated by nickel affinity chromatography (Super Ni-NTA agarose resin, Generon), eluting 6xHis-SUMO construct with high-salt phosphate buffer [10 mM potassium phosphate buffer (pH 7.2), 150 mM KCl] with the addition of 250 mM imidazole. The 6xHis-SUMO domain was cleaved by incubating the construct with the tobacco etch virus protease (1:20 protein construct/tobacco etch virus molar ratio) overnight at 4 °C, while dialyzing the mixture against low-salt phosphate buffer [10 mM potassium phosphate (pH 7.2), 15 mM KCl]. N-RRM1-2 was isolated with a second nickel affinity chromatography and nucleic acids bound non-specifically to the protein construct were removed by heparin affinity chromatography (HiTrap Heparin HP; GE Healthcare) and eluted with potassium phosphate buffer with 1.5 M KCl. Pure N-RMMs was finally obtained with a size-exclusion chromatography performed on a Äkta pure system (HiLoad 16/60 Superdex 75 prep grade column; GE Healthcare). The protein was eluted in

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

low-salt phosphate buffer (10 mM potassium phosphate buffer at pH 7.2 and 15 mM KCl), aliquoted, flash-frozen and stored at −20 °C. This protocol allowed to obtain a pure protein construct as based on SDS-PAGE and size-exclusion chromatography.

Circular dichroism

CD spectra were recorded on a JASCO-1100 spectropolarimeter and deconvoluted with the online analysis software DichroWeb [61]. One-millimeter path-length quartz cuvettes (type S3/Q/1; Starna Scientific) were employed, and constant N_2 flush at 4.0 L/min was applied. Far-UV CD spectra of 15 μM N-RRM1-2 before and after thermal denaturation were acquired at 20 °C in low-salt phosphate buffer as an average of 10 scans. For measurements in the presence of RNA, the contribution of the isolated aptamers was subtracted from the sample spectra, assuming that the nucleotides do change conformation upon binding to the protein. This assumption was based on secondary predictions which excluded secondary structure formation for all aptamers. N-RRM1-2 thermal denaturation in low-salt phosphate buffer was monitored following changes in CD signal at 222 nm between 20 °C and 90 °C, using a 1 °C/min gradient. The apparent melting temperature (T_m) was derived by fitting the data according to literature [62].

ThT binding assays

N-RRM1-2 was speedily thaw, spinned down and filtered with 0.2-μm syringe filter before each assay, to ensure sample homogeneity. Binding of N-RRM1-2 to the amyloid dye ThT upon thermal denaturation was assessed using spectrofluorometry. Protein samples were diluted to 15 μM in high-salt phosphate buffer with a final ThT concentration of 20 μM. ThT fluorescence emission spectra were recorded from 460 to 600 nm with 440-nm excitation wavelength, in a 1-cm path-length quartz cuvette using a Jasco FP-8200 spectrofluorometer (JASCO). Each spectrum represents the average of at least three scans.

Biolayer interferometry

All experiments were performed in high-salt potassium phosphate buffer with the addition of 0.05% Tween on an Octet Red instrument (ForteBio, Inc., Menlo Park, CA) operating at 25 °C. Biotinylated RNA molecules (1 μg/ml) were loaded onto streptavidin-coated biosensors (FortèBio Dip and Read™) for 100 s, to ensure sensor saturation. Binding to N-RRM1-2, with concentrations ranging from 1 nM to 15 μM, was allowed for 200 s, or until the response signal reached plateau. Apparent K_d values were estimated by fitting the response intensity (nm) as a function of the protein concentration using the Octet Data Analysis Software. Since BLI does not give information about reaction stoichiometry, the data were fitted to a simple 1:1 binding model using non-linear least squares methods (Levenberg Marquardt algorithm).

Aggregation kinetics assays

A previously established protocol for high-throughput monitoring of protein aggregation was specifically customized to allow working in the presence of RNA. The N-RRM1-2 aggregation reaction was monitored on a 96-well plate reader (Corning black with clear flat bottom plates), using a final N-RRM1-2 concentration of 15 μM in high-salt phosphate buffer. The reaction mixture included ProteoStat, a fluorescent aggregation reporter dye, and a broad-spectrum RNAse inhibitor to avoid nucleic acid degradation. A Teflon ball was added to each to ensure sample homogeneity along the kinetics. The plate was sealed with an optic seal to avoid evaporation. The assay plates were incubated at 37 °C under constant ≈100 rpm agitation in an Infinite 200 PRO M Plex plate reader (Tecan), and fluorescence intensity was recorded every 5 min. At the end of each assay, samples were collected from the plate and spinned down in 1.5-ml tubes for 20 min at 4 °C and at 13000 rcf [63]. Concentration of the protein found in the supernatant (soluble protein) was determined by means of Bradford assay, according to a standard curve described by known N-RRM1-2 concentrations in high-salt phosphate buffer [64]. The experiments were all performed at least in triplicate, and the results were referred to the blank, normalized and reported as percentage average.

Confocal fluorescence microscopy

Confocal micrographs were acquired at the end of aggregation assays with a confocal TCS SP8 microscope (Leica) using a 40× oil immersion objective (HC PL APO CS2 40×/1.30 OIL). Protein aggregates were imaged using both transmitted light detection and fluorescent detection upon excitation with an Argon laser line at 496 nm and filtering emission between 575 and 625 nm.

Transmission electron microscopy

The aggregation reaction mixture (10 μl) was loaded onto carbon-coated 400-mesh copper grids, allowing aggregate deposition for 5 min. Excess of aggregated samples was subsequently removed, and the grids were washed with 10 μl of distilled water and dried. The aggregates were negatively stained with 10 μl of 2% uranyl acetate for 1 min. After dye removal, the grids were washed and dried again. Micrographs of adsorbed aggregates were acquired on a JEM-1400 transmission electron microscope (JEOL) operating at a 120-kV accelerating voltage.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

Predictions of RNA binding affinities, protein stability and aggregation

RNA binding was predicted by the *cat*RAPID software [65]. We calculated 1000 random mutations in the TDP-43 structure (PDB 4BS2) and used them to compute the binding propensities to RNA_{36}. Protein stability and aggregation propensity corrected by the surface exposure were estimated by FoldX [66] and Aggrescan3D [67], respectively. We selected model 7 from the NMR bundle of 4BS2 as the representative structure according to Olderado (model: https://goo.gl/zHT31K) [68]. Since FoldX is parametrized to take into account DNA molecules but not RNA, we discarded the RNA molecule present in the PDB file.

Acknowledgments

We greatly acknowledge Chris E. Shaw for continuous encouragement and Laura Masino from the Francis Crick Institute for technical support with BLI. We thank Prof. Salvador Ventura's group and the Servei de Microscòpia of the Universitat Autonoma de Barcelona for their assistance with sample preparation and image acquisition for TEM analysis. This work was supported by a Newton International Fellowship (Royal Society), the Motor Neuron Disease Association, the UK Dementia Research Institute (UKDRI) and European Research Council (ERC, Ribomylome).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.01.028.

Received 17 September 2018;
Received in revised form 22 January 2019;
Accepted 24 January 2019
Available online xxxx

Keywords:
amyotrophic lateral sclerosis;
frontotemporal dementia;
neurodegeneration;
protein aggregation;
RNA binding

†E.Z. and R.G.M. have contributed equally to this work.

Current address: R. Graña-Montes,Department of Biochemistry, Univeristy of Zürich. Winterthurerstrasse 190, 8057, Zürich, Switzerland.

Abbreviations used:
ALS, amyotrophic lateral sclerosis; FTLD-U, frontotemporal lobar degeneration with ubiquitin-positive inclusion; RRM, RNA-recognition motif; RNA_{12}, 12-nucleotide single-stranded RNA; CD, circular dichroism; ThT, thioflavin T; BLI, biolayer interferometry; TEM, transmission electron microscopy.

References

[1] C.A. Ross, M.A. Poirier, Protein aggregation and neurodegenerative disease, Nat. Med. 10 (Suppl) (2004) S10–S17.
[2] D. Cirillo, F. Agostini, P. Klus, D. Marchese, S. Rodriguez, B. Bolognesi, G.G. Tartaglia, Neurodegenerative diseases: quantitative predictions of protein–RNA interactions, RNA 19 (2013) 129–140.
[3] J.M. Gallo, P. Jin, C.A. Thornton, H. Lin, J. Robertson, I. D'Souza, W.W. Schlaepfer, The role of RNA and RNA processing in neurodegeneration, J. Neurosci. 25 (2005) 10372–10375.
[4] A. Kryshtafovych, C. Venclovas, K. Fidelis, J. Moult, Progress over the first decade of CASP experiments, Proteins 61 (2005) 225–236.
[5] Y.B. Lee, H.J. Chen, J.N. Peres, J. Gomez-Deza, J. Attig, M. Stalekar, C. Troakes, A.L. Nishimura, E.L. Scotter, C. Vance, Y. Adachi, V. Sardone, J.W. Miller, B.N. Smith, J.M. Gallo, J. Ule, F. Hirth, B. Rogelj, C. Houart, C.E. Shaw, Expanded G4C2 repeats linked to C9ORF72 ALS and FTD form length-dependent RNA foci, sequester RNA binding proteins and are neurotoxic, Cell Rep. 5 (2013) 1178–1186.
[6] N.R. Deleault, R.W. Lucassen, S. Supattapone, RNA molecules stimulate prion protein conversion, Nature 425 (2003) 717–720.
[7] G. Bartzokis, D. Sultzer, J. Cummings, L.E. Holt, D.B. Hance, V.W. Henderson, J. Mintz, In vivo evaluation of brain iron in Alzheimer disease using magnetic resonance imaging, Arch. Gen. Psychiatry 57 (2000) 47–53.
[8] F.C. Fiesel, P.J. Kahle, TDP-43 and FUS/TLS: cellular functions and implications for neurodegeneration, FEBS J. 278 (2011) 3550–3568.
[9] Y.M. Ayala, L. De Conti, S.E. Avendaño-Vázquez, A. Dhir, M. Romano, A. D'Ambrogio, J. Tollervey, J. Ule, M. Baralle, E. Buratti, et al., TDP-43 regulates its mRNA levels through a negative feedback loop, EMBO J. 30 (2011) 277–288.
[10] H.D. Kim, T.-S. Kim, Y.J. Joo, H.-S. Shin, S.-H. Kim, C.-Y. Jang, C.E. Lee, J. Kim, RpS3 translation is repressed by interaction with its own mRNA, J. Cell. Biochem. 110 (2010) 294–303.
[11] A. Zanzoni, D. Marchese, F. Agostini, B. Bolognesi, D. Cirillo, M. Botta-Orfila, C.M. Livi, S. Rodriguez-Mulero, G.G. Tartaglia, Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein, Nucleic Acids Res. 41 (2013) 9987–9998.
[12] B.E. Docter, S. Horowitz, M.J. Gray, U. Jakob, J.C. Bardwell, Do nucleic acids moonlight as molecular chaperones? Nucleic Acids Res. 44 (2016) 4835–4845.
[13] S.W. Choi, A. Kano, A. Maruyama, Activation of DNA strand exchange by cationic comb-type copolymers: effect of cationic moieties of the copolymers, Nucleic Acids Res. 36 (2008) 342–351.
[14] C.F. Sephton, B. Cenik, B. Kutluk Cenik, J. Herz, G. Yu, TDP-43 in CNS development and function: clues to TDP-43-associated neurodegeneration, Biol. Chem. 393 (2012) 589–594.

[15] J.R. Tollervey, T. Curk, B. Rogelj, M. Briese, M. Cereda, M. Kayikci, J. König, T. Hortobágyi, A.L. Nishimura, V. Zupunski, R. Patani, S. Chandran, G. Rot, B. Zupan, C.E. Shaw, J. Ule, Characterising the RNA targets and position-dependent splicing regulation by TDP-43, Nat. Neurosci. 14 (2011) 452–458.

[16] P.H. Kuo, L.G. Doudeva, Y.T. Wang, C.K. Shen, H.S. Yuan, Structural insights into TDP-43 in nucleic-acid binding and domain interactions, Nucleic Acids Res. 37 (2009) 1799–1808.

[17] M. Neumann, D.M. Sampathu, L.K. Kwong, A.C. Truax, M.C. Micsenyi, T.T. Chou, J. Bruce, T. Schuck, M. Grossman, C. M. Clark, L.F. McCluskey, B.L. Miller, et al., Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, Science 314 (2016) 130–133.

[18] M. Zhao, J.R. Kim, R. van Bruggen, J. Park, RNA-binding proteins in amyotrophic lateral sclerosis, Mol Cells 41 (2018) 818–829.

[19] J. Sreedharan, I.P. Blair, V.B. Tripathi, X. Hu, C. Vance, B. Rogelj, S. Ackerley, J.C. Durnall, K.L. Williams, E. Buratti, F. Baralle, J. de Belleroche, J.D. Mitchell, P.N. Leigh, A. Al-Chalabi, C.C. Miller, G. Nicholson, C.E. Shaw, TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis, Science 319 (2008) 1668–1672.

[20] A.S. Chen-Plotkin, V.M. Lee, J.Q. Trojanowski, TAR DNA-binding protein 43 in neurodegenerative disease, Nat. Rev. Neurol. 6 (2010) 211–220.

[21] E. Buratti, F.E. Baralle, Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9, J. Biol. Chem. 276 (2001) 36337–36343.

[22] E. Buratti, A. Brindisi, F. Pagani, F.E. Baralle, Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: a functional link with disease penetrance, Am. J. Hum. Genet. 74 (2004) 1322–1325.

[23] E. Zacco, S.R. Martin, R. Thorogate, A. Pastore, The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein, Front. Mol. Neurosci. 11 (2018) 32.

[24] Y.C. Huang, K.F. Lin, R.Y. He, P.H. Tu, J. Koubek, Y.C. Hsu, J.J. Huang, Inhibition of TDP-43 aggregation by nucleic acid binding, PLoS One 8 (2013), e64002. .

[25] Y. Sun, P.E. Arslan, A. Won, C.M. Yip, A. Chakrabartty, Binding of TDP-43 to the 3′UTR of its cognate mRNA enhances its solubility, Biochemistry 53 (2014) 5885–5894.

[26] L. Masino, G. Nicastro, L. Calder, M. Vendruscolo, A. Pastore, Functional Interactions as a survival strategy against abnormal aggregation, FASEB J. 25 (2010) 45–54.

[27] A. Pastore, P.A. Temussi, The two faces of Janus: functional interactions and protein aggregation, Curr. Opin. Struct. Biol. 22 (2012) 30–37.

[28] M. Convertino, J. Das, N.V. Dokholyan, Pharmacological chaperones: design and development of new therapeutic strategies for the treatment of conformational diseases, ACS Chem. Biol. 11 (2016) 1471–1489.

[29] L.M. Young, A.E. Ashcroft, S.E. Radford, Small molecule probes of protein aggregation, Curr. Opin. Chem. Biol. 39 (2017) 90–99.

[30] P.J. Lukavsky, D. Daujotyte, J.R. Tollervey, J. Ule, C. Stuani, E. Buratti, F.E. Baralle, F.F. Damberger, F.H. Allain, Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43, Nat. Struct. Mol. Biol. 20 (2013) 1443–1449.

[31] B.S. Johnson, D. Snead, J.J. Lee, J.M. McCaffery, J. Shorter, A.D. Gitler, TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity, J. Biol. Chem. 284 (2009) 20329–20339.

[32] M. Mompeán, V. Romano, D. Pantoja-Uceda, C. Stuani, F.E. Baralle, E. Buratti, D.V. Laurents, The TDP-43 N-terminal domain structure at high resolution, *FEBS J.* 283 (2016) 1242–1260.

[33] Y.-J. Zhang, T. Caulfield, Y.-F. Xu, T.F. Gendron, J. Hubbard, C. Stetler, H. Sasaguri, E.C. Whitelaw, S. Cai, W.C. Lee, et al., The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation, Hum. Mol. Genet. 22 (2013) 3112–3122.

[34] S. Sugimoto, K.-I. Arita-Morioka, Y. Mizunoe, K. Yamanaka, T. Ogura, Thioflavin T as a fluorescence probe for monitoring RNA metabolism at molecular and cellular levels, *Nucleic Acids Res.* 43 (2015) e92.

[35] S. Liu, P. Peng, H. Wang, L. Shi, T. Li, Thioflavin T binds dimeric parallel-stranded GA-containing non-G-quadruplex DNAs: a general approach to lighting up double-stranded scaffolds, *Nucleic Acids Res.* 45 (2017) 12080–12089.

[36] F. Prischi, P.V. Konarev, C. Iannuzzi, C. Pastore, S. Adinolfi, S.R. Martin, D.I. Svergun, A. Pastore, Structural bases for the interaction of frataxin with the central components of iron–sulphur cluster assembly, Nat. Commun. 1 (2010) 95.

[37] S.S. Leal, J.S. Cristovao, A. Biesemeir, I. Cardoso, C.M. Gomes, Aberrant zinc binding to immature conformers of metal-free copper–zinc superoxide dismutase triggers amorphous aggregation, Metalomics 7 (2015) 333–346.

[38] C.-K. Changa, T.-H. Wub, C.-Y. Wube, M.-H. Chianga, E. Khai-Woon Toh, Y.-C. Hsu, K.-F. Lin, Y.-H. Liao, T.-H. Huang, J.J.-Z. Huang, The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity, Biochem. Biophys. Res. Commun. 425 (2012) 219–224.

[39] H. Sasaguri, J. Chew, Y.-F. Xu, T.F. Gendron, A. Garrett, C. W. Lee, K. Jansen-West, P.O. Bauer, E.A. Perkerson, J. Tong, C. Stetler, Y.-J. Zhang, The extreme N-terminus of TDP-43 mediates the cytoplasmic aggregation of TDP-43 and associated toxicity in vivo, Brain Res. 1647 (2016) 57–64.

[40] E.L. Guenther, P. Ge, H. Trinh, M.R. Sawaya, D. Cascio, D. R. Boyer, T. Gonen, Z.H. Zhou, D.S. Eisenberg, Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2, Nat. Struct. Mol. Biol. 25 (2018) 311–319.

[41] Y. Lin, E. Mori, M. Kato, S. Xiang, L. Wu, I. Kwon, S.L. McKnight, Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers, *Cell.* 167 (2016) 789–802.e12.

[42] T. Nonaka, M. Masuda-Suzukake, T. Arai, Y. Hasegawa, H. Akatsu, T. Obi, M. Yoshida, S. Murayama, D.M.A. Mann, H. Akiyama, M. Hasegawa, Prion-like properties of pathological TDP-43 aggregates from diseased brains, *Cell Rep.* 4 (2013) 124–134.

[43] T. Afroz, E.-A. Hock, P. Ernst, C. Foglieni, M. Jambeau, L.A. B. Gilhespy, F. Laferriere, Z. Maniecka, A. Plückthun, P. Mittl, P. Paganetti, F.H.T. Allain, M. Polymenidou, Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation, Nat. Commun. 8 (2017) 45.

[44] P.S. Tsoi, K.-J. Choi, P.G. Leonard, A. Sizovs, M.M. Moosa, K.R. MacKenzie, J.C. Ferreon, A.C.M. Ferreon, The N-

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028

terminal domain of ALS-linked TDP-43 assembles without misfolding, Angew. Chem. Int. Ed. 56 (2017) 12590–12593.
[45] M.B. Borgia, A. Borgia, R.B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, J. Clarke, Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins, Nature 474 (2011) 662–665.
[46] S. Pechmanna, E.D. Levyb, G.G. Tartaglia, M. Vendruscolo, Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins, PNAS 106 (2009) 10159–10164.
[47] A. Wang, A.E. Conicella, H. Broder Schmidt, E.W. Martin, S. N. Rhoads, A.N. Reeb, A. Nourse, D. Ramirez Montero, V.H. Ryan, R. Rohatgi, F. Shewmaker, M.T. Naik, T. Mittag, Y.M. Ayala, N.L. Fawzi, A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing, *EMBO J.* 37 (2018), e97452.
[48] F. Ferrone, Analysis of protein aggregation kinetics, *Methods Enzymol.* 309 (1999) 256–274.
[49] A.K. Buell, C. Galvagnion, R. Gaspar, E. Sparr, M. Vendruscolo, T.P.J. Knowles, S. Linse, C.M. Dobson, Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation, PNAS 111 (2014) 7671–7676.
[50] S.-J. Lee, H.-S. Lim, E. Masliah, H.-J. Lee, Protein aggregate spreading in neurodegenerative diseases: problems and perspectives, Neurosci. Res. 70 (2011) 339–348.
[51] S.I.A. Cohen, M. Vendruscolo, C.M. Dobson, T.P.J. Knowles, From macroscopic measurements to microscopic mechanisms of protein aggregation, J. Mol. Biol. 421 (2012) 160–171.
[52] C. Maurel, B. Madji-Hounoum, R.A. Thepault, S. Marouillat, C. Brulard, V. Danel-Brunaud, J.P. Camdessanche, H. Blasco, P. Corcia, C.R. Andres, P. Vourc'h, Mutation in the RRM2 domain of TDP-43 in amyotrophic lateral sclerosis with rapid progression associated with ubiquitin positive aggregates in cultured motor neurons, Amyotroph Lateral Scler Frontotemporal Degener 19 (2018) 149–151.
[53] J.P. Changeux, S. Edelstein, Conformational selection or induced fit? 50 years of debate resolved, *Biol Rep.* 3 (2011) 19.
[54] K.A. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature 450 (2007) 964–972.
[55] D.W. Miller, K.A. Dill, Ligand binding to proteins: the binding landscape model, *Protein Sci.* 6 (1997) 2166–2179.
[56] S. Horowitz, J.C.A. Bardwell, RNAs as chaperones, *RNA Biol.* 13 (2016) 1228–1231.
[57] B.K. Pathak, S. Mondal, A.N. Ghosh, C. Barat, The ribosome can prevent aggregation of partially folded protein intermediates: studies using the *Escherichia coli* ribosome, PLoS One 9 (2014), e96425.
[58] Y. Cordeiro, B. Macedo, J.L. Silva, M.P.B. Gomes, Pathological implications of nucleic acid interactions with proteins associated with neurodegenerative diseases, Biophys. Rev. 6 (2014) 97–110.
[59] P. Anderson, N. Kedersha, Stress granules, Curr. Biol. 19 (2009) R397–R398.
[60] H. Yang, H.-Y. Hu, Sequestration of cellular interacting partners by protein aggregates: implication in a loss-of-function pathology, FEBS J. 283 (2016) 3705–3717.
[61] L. Whitmore, B.A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, Biopolymers 89 (2008) 392–400.
[62] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nat. Protoc. 1 (2006) 2527–2535.
[63] J. Pujols, S. Peña-Díaz, M. Conde-Giménez, F. Pinheiro, S. Navarro, J. Sancho, S. Ventura, High-throughput screening methodology to identify alpha-synuclein aggregation inhibitors, Int. J. Mol. Sci. 18 (2017), e478. .
[64] M.M. Bradfod, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.
[65] M. Bellucci, F. Agostini, M. Masin, G.G. Tartaglia, Predicting protein associations with long noncoding RNAs, Nat. Methods 8 (2011) 444–445.
[66] R. Guerois, J.E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations, J. Mol. Biol. 320 (2002) 369–387.
[67] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik, S. Ventura, AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures, *Nucleic Acids Res.* 43 (W1) (2015) W306–W313.
[68] L.A. Kelley, S.P. Gardner, M.J. Sutcliffe, An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies, Protein Eng. 9 (1996) 1063–1065.

Please cite this article as: E. Zacco, R. Graña-Montes, S. R. Martin, et al., RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.01.028