



Aromatic and aliphatic residues of the disordered region of TDP-43 are on a fast track for self-assembly



Douglas V. Laurents^a, Cristiana Stuani^b, David Pantoja-Uceda^a, Emanuele Buratti^b, Miguel Mompeán^{a,*}

^a "Rocasolano" Institute for Physical Chemistry, Spanish National Research Council, Serrano 119, 28006, Madrid, Spain

^b International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano 99, I-34149, Trieste, Italy

ARTICLE INFO

Article history:

Received 10 August 2021

Accepted 17 September 2021

Available online 20 September 2021

Keywords:

TDP-43

Biomolecular condensate

NMR

¹H–¹⁵N HSQC

Kinetics

Aggregation

ABSTRACT

The C-terminal, intrinsically disordered, prion-like domain (PrLD) of TDP-43 promotes liquid condensate and solid amyloid formation. These phase changes are crucial to the normal biological functions of the protein but also for its abnormal aggregation, which is implicated in amyotrophic lateral sclerosis (ALS) and certain dementias. We and other previously found that certain amyloid forms emerge from an intermediate condensed state that acts as a nucleus for fibrillization. To quantitatively ascertain the role of individual residues within TDP-43's PrLD in its early self-assembly we have followed the kinetics of NMR ¹H–¹⁵N HSQC signal loss to obtain values for the lag time, elongation rate and extent of condensate formation at equilibrium. The results of this analysis represent a robust corroboration that aliphatic and aromatic residues are key drivers of condensate formation.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

TDP-43 is an essential protein that regulates RNA splicing and translation [1]. To carry out these functions, TDP-43 forms part of biomolecular condensates, such as nuclear speckles and stress granules [2]. TDP-43 also assembles functional amyloids to function in muscle regeneration [3]. In addition, this protein occasionally forms other assembled, aberrant species that are implicated in ALS, FTD and other dementia [4,5]. A rather large protein, TDP-43 contains a N-terminal domain with a singular structure [6], two RNA recognition motifs (RRM) and a long, disordered C-terminal tail. Research over the last decade has shown that this C-terminal, intrinsically disordered region or prion-like domain (PrLD, residues 267–414) plays key roles in both physiological and pathological phase transitions and oligomeric assembly. In particular, mutational studies [7], mass spectrometry [8] and solution NMR spectroscopy have shown that the PrLD contains a hydrophobic segment which tends to form a sticky α -helix (residues 321–340) [9–11], several aromatic residues that are scattered throughout the sequence and promote condensate formation [12], and finally a Q/N-rich segment (residues 341–366) with strong amyloid-forming

propensity [13,14]. More recently, consensus has started to emerge that condensate formation precedes the assembly of certain fibril forms [15–17], and a combination of solution and solid-state NMR spectroscopies has revealed that Phe residues in the PrLD become fixed when the condensate converts into the fibrillar solid state [17]. A recent cryo-electron microscopy (cryo-EM) structure provided direct structural evidence for the presence of such residues in the amyloid core [18], thus highlighting them as potentially valuable pharmaceutical targets. Despite these advances, many residue-level details of the initial assembly events are still unclear since both solid-state NMR and cryo-EM study the final, end-stage fibril forms. The main objective of this study, therefore, has been to characterize TDP-43 condensate formation quantitatively by following the kinetics of this process on a per-residue level by solution NMR spectroscopy.

2. Materials and methods

A ¹³C, ¹⁵N-labeled TDP-43 PrLD construct, consisting of residues 267–414 was produced and handled as previously described to ensure conditions where fibrils emerge from a droplet, condensed state (Fig. 1A) [17]. Under these conditions, the monomer-to-droplet conversion can be tracked by solution NMR [17]. To conduct NMR experiments, the protein is initially stored in 8 M urea and the denaturant is removed using a PD-10 (GE-Healthsystems)

* Corresponding author.

E-mail address: mmompean@iqfr.csic.es (M. Mompeán).

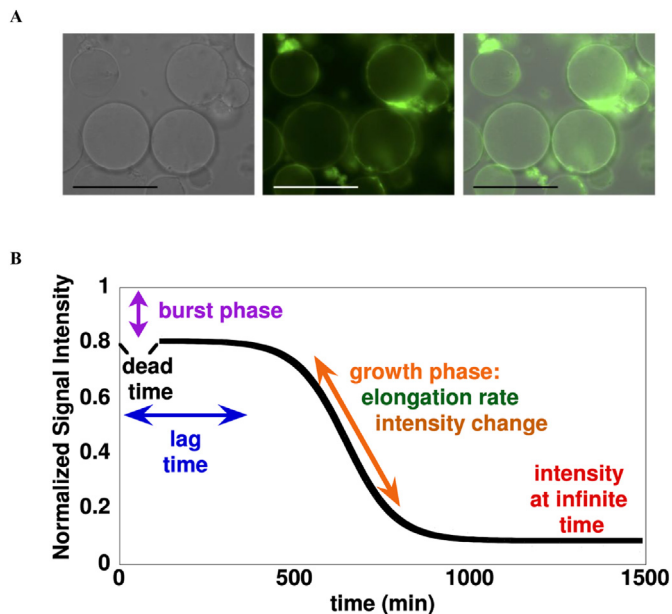


Fig. 1. A. Thioflavin T (ThT) staining of condensates formed by the PrLD of TDP-43. ThT shines on the condensate/solvent interface, but not the condensate interior, indicating that fibrils emerge from the condensate surface. Left: light transmitted image; middle: fluorescent image; right: overlay. Bar scale is 100 μ m. B. Schematic illustrating different stages and events in TDP-43 phase transitions, as monitored by the intensity of liquid state NMR ^1H - ^{15}N HSQC spectra. The **dead time** is the period prior to the recording of the first NMR spectrum. During the **dead time**, part of the intensity that is lost (presumably due to condensate formation) is the **burst phase intensity**. The **lag time** is the period prior to the formation of amyloid nuclei. Once formed, there is a rapid decrease in intensity as monomers quickly added on to the growing fibril. From this period, it is possible to measure an **intensity change** as well as a growth or **elongation rate**. Finally, at long times, there is a small amount of signal remaining, the **intensity at infinite time**, which is due to the population of monomer in solution that is in equilibrium with condensate and amyloid phase.

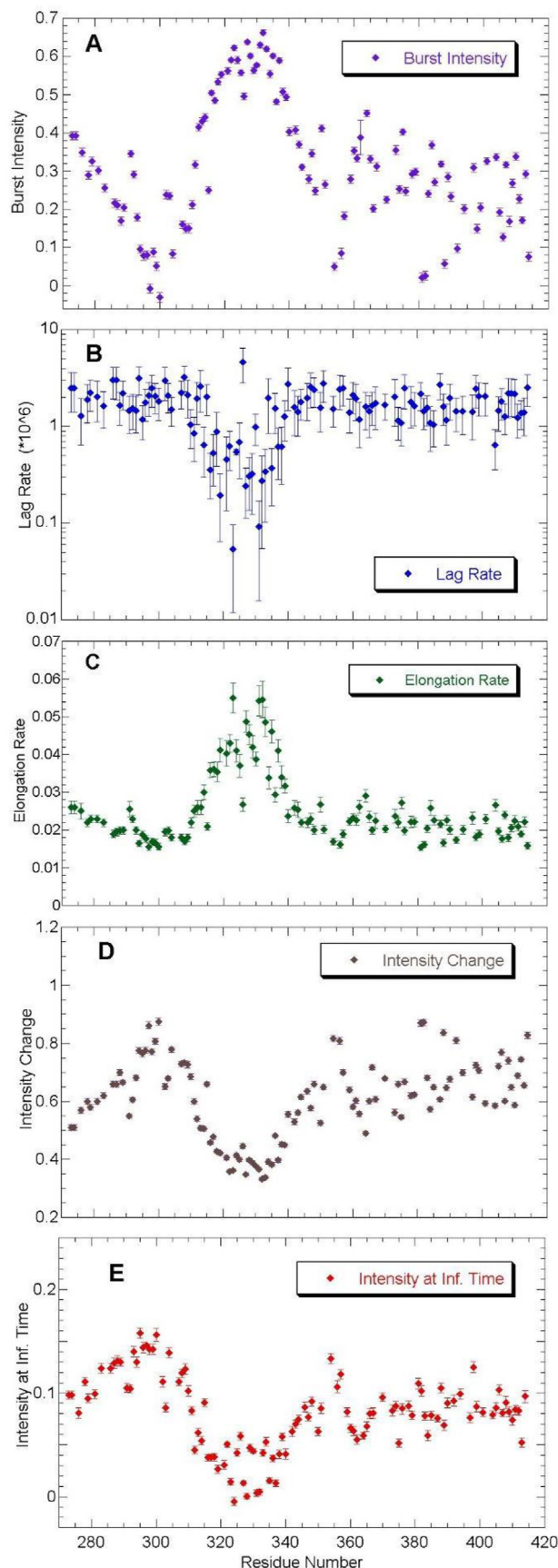


Fig. 3. TDP-43 PrLD show distinct residue-level phase transition kinetics.

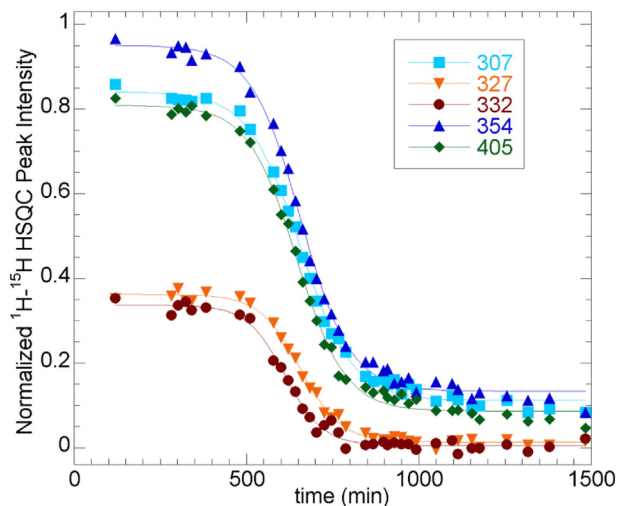
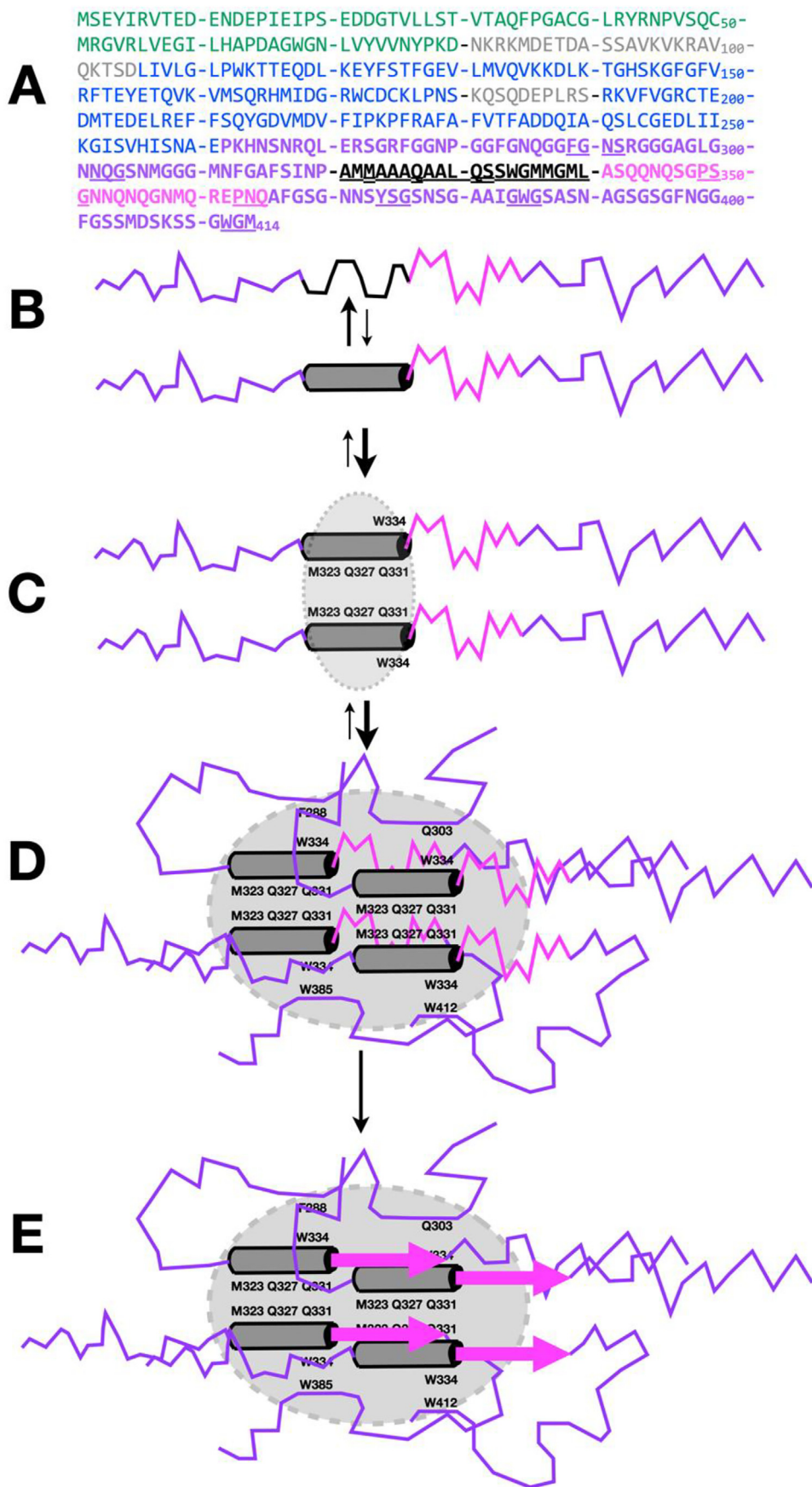


Fig. 2. Signal versus time data for representative residues: 307 from the first glycine/aromatic-rich region (**light blue squares**), 327 and 332 from the hydrophobic segment (**inverted orange triangles** and **red circles**, respectively), 354 from the Q/N-rich segment (**blue triangles**) and 405 from the second glycine/aromatic-rich region (**green diamonds**). The solid curves show the fit of Equation (1) to the data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

desalting column to buffer exchange the protein to 1 mM DAC in 90% $\text{H}_2\text{O}/10\%$ D_2O . The final urea concentration was <10 mM as



determined using refractive index [17,19]. Then, a series of 2D ^1H – ^{15}N HSQC NMR spectra were recorded in an 800 MHz spectrometer equipped with a cryogenic probe, at 25 °C over a period of 25 h. The ^1H – ^{15}N HSQC peak assignments used are deposited in the BMRB under accession code 50154, and all the experimental procedures used to obtain them are reported in Ref. [19].

The intensities of each resolved peak in the ^1H – ^{15}N HSQC spectra was measured using the program NMRFAM-SPARKY [20]. The kinetic model for amyloid formation described by Morris et al. [21]:

$$I_t = (k_1 / k_2 + I_0) / [1 + (k_1 / k_2 \cdot I_0) \cdot \exp((k_1 + k_2 \cdot I_0) \cdot t)] \quad (1)$$

Where I_t is the signal intensity at time t , k_1 is the nucleation rate (inversely related to the lag time to form a minimal amyloid nucleus that here is a condensed form), k_2 is the elongation rate (which senses the amyloid fibril growth rate after nucleation from the condensed state), and I_0 is the initial intensity at time 0.

Parameters describing the lag time, growth rate and extent of amyloid formation at equilibrium were obtained by fitting this model by a least-squares algorithm to the per-residue intensity versus time data sets (Fig. 1B). The burst amplitude is the intensity that is lost during the dead time prior to the acquisition of the first experiment (Fig. 1B). Here, it was estimated as the intensity of the strongest ^1H – ^{15}N peak in the first spectrum minus the sum of the intensity changes during amyloid growth plus that at infinite time (Fig. 1B).

3. Results

Following denaturant removal, TDP-43 self-assembly manifests a residue-specific loss of peak intensities in the ^1H – ^{15}N HSQC NMR spectra, as we and others have shown previously [10–12,17]. The intensity decays are rapid for residues within the sticky, hydrophobic α -helix of TDP-43's PrLD and have been interpreted as line broadening resulting from intermolecular interactions, establishing the unique role of this region in driving TDP-43 self-assembly under very different conditions [10–12,17]. We note, however, that a per-residue peak analysis has not been accomplished yet. To address this gap, in this work we have analyzed in detail such intensity changes during self-assembly by collecting a series of ^1H – ^{15}N HSQC spectra over the course of 25 h following removal of denaturant. Data fitting of the peak intensities to a model for amyloid growth affords a quantitative interpretation of the dynamic picture for the monomer-to droplet-to fibril conversion that we reported for TDP-43's PrLD (Fig. 1). Noticeable intensity decreases are observed for most residues over time. Certain residues, such as 327 and 332 show signal decrease sooner and more completely than others. These distinctions are seen more clearly when the signal intensities are plotted versus time (Fig. 2). In particular, the intensity changes occurring in the burst phase (Fig. 3A) is greater for residues in the hydrophobic helix.

The values for the burst phase intensities, lag rate, elongation rate, intensity change during elongation and the intensity at infinite time, $I_t(\infty)$, obtained from fitting equation (1) to the data are plotted in Fig. 3A–E. From these data, it can be observed that residues of the hydrophobic helix have significantly elevated values during the burst phase intensity (Fig. 3A). This strongly suggests

that these residues coalesce to form a condensate more quickly and to a greater extent than other residues. This signal loss occurs as tumbling of the local polypeptide segment slows upon condensate formation leading to high transverse relaxation rates and very broad signals [22]. In addition, short segments containing a few residues centered on residues 291, 303, 350, 364, 375 and 384 also show burst intensities which are significantly higher than average. Interestingly, the lag rate is lowest for the hydrophobic segment; in fact, it is an order of magnitude lower than the average of the remaining residues (Fig. 3B). The residues of the hydrophobic segment also show the fastest elongation rates (Fig. 3C). In addition, other short segments centered on residues 291, 350, 364, 375 and 384 also manifest faster than average elongation rates. Regarding the intensity change during elongation, residues of the hydrophobic segment and the other short segments aforementioned have the smallest values (Fig. 3D), as a rather substantial fraction of their signal was lost in the burst phase. In contrast, neighboring residues like 355, 381, or 398 show high intensity changes during elongation as their burst phase amplitudes are small. Finally, the intensity remaining at infinite time, $I_t(\infty)$, also shows quantitative differences (Fig. 3E). The hydrophobic segment shows the smallest values, approaching zero. For this segment the conformational evolution from monomer to high-molecular weight species appears to be almost complete. Other short segments near residues 364, 375, 384 and 413 also have lower than average $I_t(\infty)$ values, whereas short segments near residues 294, 300, 354, 357, 381 and 398 have higher than average $I_t(\infty)$ values (Fig. 3E). Within the hydrophobic helix itself, certain residues like 323, 327, 331 and 332 have larger burst phase amplitude changes, short lag times, faster growth rates and lower $I_t(\infty)$ values (Fig. 3). Keeping in mind that the hydrophobic segment forms a α -helix, residues 323, 327, 331 and 332 are positioned on the same side of the α -helix. This would explain the recently reported helix-driven dimerization, critical for condensate formation [23], illustrating the key participation of these four residues in association processes as gauged from the analysis presented in this manuscript.

4. Discussion

The residue sequence of TDP-43 is shown in Fig. 4A. Residues of the hydrophobic segment, which show fast phase change kinetics are underlined, as are the residues of other short segments which also show faster-than-average kinetics and larger burst phase amplitudes. It is interesting that most of these short segments contain an aromatic residue, and have already been reported to be mutated in ALS patients and to promote aggregation in several *in vitro* and *in cellulo* systems [24]. Within the hydrophobic segment, residues with especially quick and complete condensate kinetics, namely 323, 327 & 331 + 332, are underlined in Fig. 4A. These residues would be on the same side of the α -helix and we propose that they interact with the same residues in other TDP-43 molecules to form initial dimers and then coiled-coils (Fig. 4B). Such coiled-coils would form quickly and explain why the NMR ^1H – ^{15}N HSQC signal drops faster for the hydrophobic segment, as well as the initial increase in light scattering and fluorescence anisotropy that we reported elsewhere during this process [17]. In such an α -helix environment, Trp334 would be positioned on the exterior face (Fig. 4C), and would be available to interact with aromatic, aliphatic and/or sp^2 groups found in residues like Asn, Gln, Pro,

Fig. 4. Model for residue contributions to TDP-43 PrLD condensate and amyloid formation

A. Residue sequence of TDP-43, coloring the N-terminal domain **green**, the RRM domains **blue** and the NLS (residues 80–100) and the linker connecting the RRM domains (residues 181–190) **gray**. Regarding the C-terminal disordered region, the glycine/aromatic-rich segments are colored **purple**, the hydrophobic segment shown in **black** and the Q/N-rich segment is colored **magenta**. **B.** The C-terminal region is mostly disordered with a minor α -helical population. **C.** Association with another α -helix would increase stability and position W334 to favor long-range contacts (**D**) as suggested by these residues' faster and more complete aggregation kinetics. **E.** These events would configure the Q/N-rich segments (**magenta**) to form β -strands (arrows) as a first step towards amyloid formation.

Phe, Tyr, Trp to drive the formation of larger oligomers/condensates (Fig. 4D). As these interactions are formed, more light would be scattered, the Trp side chains would become more rigid and attain a somewhat more nonpolar milieu, concurring with reported fluorescence data [17] and in line with previous results that uncovered the important contributions of Trp residues for the self-association of this region [12]. During this conformational evolution, other segments would retain more liberty and still show relatively strong signals in the ^1H – ^{15}N HSQC spectra.

According to the model we present in this manuscript, amyloid would be expected to form in the last stage of the conformational evolution (Fig. 4E) that follows condensate formation, in agreement with previous findings [15–17] (Fig. 1A). Although the results reported here cannot identify which residues would be comprising the amyloid, it is interesting to keep in mind that our recent confocal fluorescence results [17] showed that amyloid formation occurs mostly on the surface of the condensates, not in its core (Fig. 1A). This suggests that some more flexible/exterior/polar residues, such as the Q/N-rich segment, might actually play a pivotal role in amyloid formation, which is in accord with some previous [7,8,13,14] and recent findings [12,17,18]. This and other questions remains to be addressed in future studies. For example, aromatic residues that seemed not critical for the assembly of condensates [12], but comprise the fibril core in amyloids emerging from the condensate interface [17,18], could be mutated to e.g. alanines to test whether TDP-43 remains active in cells. Finally, we propose that this approach of following self-assembly on the residue-level instead of just globally by ThT kinetics can give valuable insight as shown here, and it would be useful to apply this tool to other assembling proteins to interrogate what the critical interactions are *in vitro* and their relevance (e.g. through mutational studies) in cells.

Declaration of interests

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

Acknowledgments

This work has been funded by Grants CTQ2017-84371-P to D.P.-U. and BTC-PID2019-109306RB-I00 to D.V.L. from the Spanish Ministry of Science and Innovation; AriSLA (PATHENSTDP project) to E.B.; and Grant LCF/BQ/PR19/11700003 from La Caixa Foundation (ID 100010434) to M.M. M.M. is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2019-026574-I). NMR experiments were performed in the “Manuel Rico” NMR Laboratory (LMR) of the Spanish National Research Council (CSIC), a node of the Spanish Large-Scale National Facility (ICTS R-LRB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- [1] E. Buratti, Trends in understanding the pathological roles of TDP-43 and FUS proteins, *Adv. Exp. Med. Biol.* 1281 (2021) 243–267, https://doi.org/10.1007/978-3-030-51140-1_15.
- [2] C. Colombrita, E. Zennaro, C. Fallini, M. Weber, A. Sommacal, E. Buratti, V. Silani, A. Ratti, TDP-43 is recruited to stress granules in conditions of oxidative insult, *J. Neurochem.* 111 (2009) 1051–1061, <https://doi.org/10.1111/j.1471-4159.2009.06383.x>.
- [3] T.O. Vogler, J.R. Wheeler, E.D. Nguyen, M.P. Hughes, K.A. Britson, E. Lester, B. Rao, et al., TDP-43 and RNA Form Amyloid-like Myo-Granules in Regenerating Muscle, vol. 563, 2018, pp. 508–513, <https://doi.org/10.1038/s41586-018-0665-2>.
- [4] M. Neumann, D.M. Sampathu, L.K. Kwong, A.C. Truax, M.C. Micsenyi, T.T. Chou, et al., Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, *Science* 314 (2006) 130–133, <https://doi.org/10.1126/science.1134108>.
- [5] P.T. Nelson, D.W. Dickson, J.Q. Trojanowski, C.R. Jack, P.A. Boyle, K. Arfanakis, R. Rademakers, I. Alafuzoff, J. Attems, C. Brayne, I. Coyle-Gilchrist, H.C. Chui, D.W. Fardo, M.E. Flanagan, G. Halliday, S. Hokkanen, S. Hunter, G.A. Jicha, Y. Katsumata, C.H. Kawas, et al., Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report, *Brain* 142 (2019) 1503–1527, <https://doi.org/10.1093/brain/awz099>.
- [6] 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, <https://doi.org/10.1111/febs.13651>.
- [7] M. Budini, E. Buratti, C. Stuani, C. Guarnaccia, V. Romano, L. De Conti, F.E. Baralle, Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region, *J. Biol. Chem.* 287 (2012) 7512–7525, <https://doi.org/10.1074/jbc.M111.288720>.
- [8] F. Kametani, T. Obi, T. Shishido, H. Akatsu, S. Murayama, Y. Saito, M. Yoshida, M. Hasegawa, Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains, *Sci. Rep.* 6 (2016) 23281, <https://doi.org/10.1038/srep23281>.
- [9] L.L. Jiang, M.X. Che, J. Zhao, C.J. Zhou, M.Y. Xie, H.Y. Li, J.H. He, H.Y. Hu, Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion, *J. Biol. Chem.* 288 (2013) 19614–19624, <https://doi.org/10.1074/jbc.M113.463828>.
- [10] A.E. Conicella, G.H. Zerze, J. Mittal, N.L. Fawzi, ALS mutations disrupt phase separation mediated by α -helical structure in the TDP-43 low-complexity C-terminal domain, *Structure* 24 (2016) 1537–1549, <https://doi.org/10.1016/j.str.2016.07.007>.
- [11] L. Lim, Y. Wei, Y. Lu, J. Song, ALS-causing mutations significantly perturb the self-assembly and interaction with nucleic acid of the intrinsically disordered prion-like domain of TDP-43, *PLoS Biol.* 14 (2016), e1002338, <https://doi.org/10.1371/journal.pbio.1002338>.
- [12] H.R. Li, W. C. Chiang, P.C. Chou, W.J. Wang, J.R. Huang, TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues, *J. Biol. Chem.* 293 (2018) 6090–6098, <https://doi.org/10.1074/jbc.AC117.001037>.
- [13] M. Mompeán, E. Buratti, C. Guarnaccia, R.M. Brito, A. Chakrabarty, F.E. Baralle, D.V. Laurents, Structural characterization of the minimal segment of TDP-43 competent for aggregation, *Arch. Biochem. Biophys.* 545 (2014) 53–62, <https://doi.org/10.1016/j.abb.2014.01.007>.
- [14] M. Mompeán, R. Hervás, Y. Xu, T.H. Tran, C. Guarnaccia, E. Buratti, F.E. Baralle, L. Tong, M. Carrión-Vázquez, A.E. McDermott, D.V. Laurents, Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43, *J. Phys. Chem. Lett.* 6 (2015) 2608–2615, <https://doi.org/10.1021/acs.jpcclett.5b00918>.
- [15] W.M. Babinchak, R. Haider, B. K. Dumm, P. Sarkar, K. Surewicz, J.K. Choi, W.K. Surewicz, The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain, *J. Biol. Chem.* 294 (2019) 6306–6317, <https://doi.org/10.1074/jbc.RA118.007222>.
- [16] B.D. Fonda, K.M. Jami, N.R. Boulos, D.T. Murray, Identification of the rigid core for aged liquid droplets of an RNA-binding protein low complexity domain, *J. Am. Chem. Soc.* 143 (2021) 6657–6668, <https://doi.org/10.1021/jacs.1c02424>.
- [17] D. Pantoja-Uceda, C. Stuani, D.V. Laurents, A.E. McDermott, E. Buratti, M. Mompeán, Phe-Gly motifs drive fibrillization of TDP-43's prion-like domain condensates, *PLoS Biol.* 19 (2021), e3001198, <https://doi.org/10.1371/journal.pbio.3001198>.
- [18] Q. Li, W.M. Babinchak, W.K. Surewicz, Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43, *Nat. Commun.* 12 (2021) 1620, <https://doi.org/10.1038/s41467-021-21912-y>.
- [19] D. Pantoja-Uceda, C. Stuani, D.V. Laurents, A.E. McDermott, E. Buratti, M. Mompeán, NMR assignments for the C-terminal domain of human TDP-43, *Biomol. NMR Assign* 15 (2021) 177–181, <https://doi.org/10.1007/s12104-020-10002-7>.
- [20] W. Lee, M. Tonelli, J.L. Markley, Nmrfam-Sparky, Enhanced software for biomolecular NMR spectroscopy, *Bioinformatics* 31 (2015) 1325–1327, <https://doi.org/10.1093/bioinformatics/btu830>.
- [21] A.M. Morris, M.A. Watzky, J.N. Agar, R.G. Finke, Fitting neurological protein aggregation kinetic data via a 2-step, minimal/Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth, *Biochemistry* 47 (2008) 2413–2427, <https://doi.org/10.1021/bi701899y>.
- [22] K.A. Burke, A.M. Janke, C.L. Rhine, N.L. Fawzi, Residue-by-Residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II, *Mol. Cell.* 60 (2015) 231–241, <https://doi.org/10.1016/j.molcel.2015.09.006>.
- [23] A.E. Conicella, G.L. Dignon, G.H. Zerze, H.B. Schmidt, A.M. D'Ordine, Y.C. Kim, R. Rohatgi, Y.M. Ayala, J. Mittal, N.L. Fawzi, TDP-43 α -helical structure tunes liquid-liquid phase separation and function, *Proc. Natl. Acad. Sci. U.S.A.* 117 (2020) 5883–5894, <https://doi.org/10.1073/pnas.1912055117>.
- [24] 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, <https://doi.org/10.1074/jbc.M109.010264>.