

UCLA

UCLA Electronic Theses and Dissertations

Title

Structural studies of functional and pathogenic amyloid fibrils

Permalink

<https://escholarship.org/uc/item/0c24p4j1>

Author

Lu, Jiahui

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structural studies of functional and pathogenic amyloid fibrils

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Biochemistry, Molecular and Structural Biology

by

Jiahui Lu

2023

© Copyright by

Jiahui Lu

2023

ABSTRACT OF THE DISSERTATION

Structural studies of functional and pathogenic amyloid fibrils

by

Jiahui Lu

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2023

Professor David S. Eisenberg, Chair

In the last decade, there has been increasing research into the function and the mechanism of pathogenicity of membraneless organelles (MLOs). MLOs are essential for RNA metabolism including RNA transportation, transcription, translation, and degradation. Ribonucleoproteins (RNPs) participate in these processes and are key components of MLOs. For example, the RNPs hnRNPA2 and hnRNPF bind myelin basic protein mRNA, and this complex associates with a type of MLO called an RNA transport granule, which facilitates transportation to oligodendrocytes for translation. Similarly, the RNPs hnRNPU and LSM14 bind and localize RNA, and this complex associates with a type of MLO called a processing body, which facilitates RNA destabilization and decay. Another example is that the RNPs FUS and TDP-43 bind mRNA, and this complex associates with a type of MLO called a stress granule which forms and disappears in response to stress stimuli. Increasing studies are performed to span our knowledge of the relationship between MLOs and neurodegenerative diseases.

RNPs often contain two types of functional domains: an RNA binding domain (RBD) for recognizing RNA and a low complexity domain (LCD) to facilitate MLO formation; most attention in neurodegeneration research has focused on the LCD. LCDs are composed of a limited repertoire of amino acid types such as Glycine, Tyrosine, and polar residues. LCDs of many RNPs, including FUS and LSM4, form a separate, dense phase in vitro similar to cellular MLOs and can further self-assemble into stable amyloid-like fibrils with cross- β structures. Their propensity to phase separate implicates the LCDs of RNPs as being principle drivers of the formation and dynamics of MLOs. Mutations in the LCDs can accelerate their self-assembly rates, leading to hyper-stable MLOs which contribute to diseases. Over 50 mutations are found to be located in the LCDs and convert functional amyloid-like fibrils into pathogenic amyloid, but only a few atomic structures of the LCDs are known due to their lack of tertiary structures. Knowing the structures of the LCDs can help us better understand their functions and pathogenic mechanisms. So far, over 40 amyloid fibril structures are determined to atomic resolutions, but none before the work described here gives a clear structural explanation of its pathogenic mechanism.

To fill this gap, we used cryoEM and determined the atomic structures of both wildtype and the variant hnRNPA2 LCD, an RNP component of RNA transport granules and stress granules. The wildtype structure of hnRNPA2 LCD shares intrinsic structural and energetic properties with other RNPs such as FUS: it has only one fibril morphology with a single protofilament and it is stabilized by hydrogen bonding and polar interactions. Wildtype hnRNPA2 has a poor solvation energy value, caused by the scarcity of residues that adopt the β -sheet conformation in the structure, which thereby weakens fibril stability. The variant structure turns out to be very different from wildtype in three ways: 1) It has at least 6 morphologies with two or more protofilaments. 2) It has a solvation energy value close to other pathogenic amyloids such as Tau. 3) More residues adopt β -

sheet conformations. Compellingly, the variant structures of hnRNPA2 LCD offer evidence for its pathogenic mechanism. A loss of function arises because the variant's PY nuclear localization signal (PY-NLS) is hidden in the fibril core, disrupting its ability to bind to transporters and be localized into the nucleus. Over half of the inward facing residues in the wildtype structure turn outward facing in the variant structure, exposing new surface epitopes that could potentially contribute to the variant's toxicity. Thermostability and phase separation assays also confirm our structural results that the variant hnRNPA2 LCD fibrils are more stable than wildtype fibrils. We believe that the variant hnRNPA2 LCD fibril structure is unusual among other pathogenic amyloid fibrils because it suggests an obvious mechanism for pathogenicity and may shed light on analogous mutational conversions occurring in other neurological diseases.

The dissertation of Jiahui Lu is approved.

Douglas L. Black

Todd O. Yeates

Gerard C. L. Wong

David S. Eisenberg, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

List of Figures.....	vii
List of Tables.....	ix
Acknowledgements.....	x
Vita.....	xiii
Publications.....	xiv
Overview.....	1
Chapter 1 The cryoEM structure of the fibril-forming low-complexity domain of hnRNPA2 reveals distinct differences from pathogenic amyloid and shows how mutation converts it to the pathogenic form.....	5
Chapter 2 Cryo-EM structures of the D290V mutant of the hnRNPA2 low-complexity domain mask the nuclear import signal required for function.....	42
Chapter 3 Small molecules disaggregate alpha-synuclein and prevent seeding from patient brain-derived fibrils.....	83
Chapter 4 Atomic structures of TDP-43 LCD segments and Insights into reversible or pathogenic aggregation.....	121
Chapter 5 Cryo-EM structures of hIAPP fibrils seeded by patient-extracted fibrils reveal new polymorphs and conserved fibril cores.....	165
Chapter 6 Low complexity domains of the nucleocapsid protein of SARS-CoV-2 form amyloid fibrils.....	202

LIST OF FIGURES

Chapter 1

Figure.1.....	29
Figure.2.....	30
Figure.3.....	31
Figure.4.....	31

Chapter 2

Figure.1.....	66
Figure.2.....	68
Figure.3.....	69
Figure.4.....	71
Figure.5.....	72

Chapter 3

Figure.1.....	108
Figure.2.....	109
Figure.3.....	111
Figure.4.....	113

Chapter 4

Figure.1.....	150
Figure.2.....	151
Figure.3.....	152
Figure.4.....	153

Chapter 5

Figure.1.....	192
Figure.2.....	193
Figure.3.....	194
<u>Chapter 6</u>	
Figure.1.....	247
Figure.2.....	249
Figure.3.....	250
Figure.4.....	252
Figure.5.....	254

LIST OF TABLES

Chapter 1

Table 1.....	32
Table 2.....	33
Table 3.....	33
Table 4.....	34
Table 5.....	34
Table 6.....	34

Chapter 2

Table 1.....	74
--------------	----

Chapter 4

Table 1.....	154
Table 2.....	155

Chapter 5

Table 1.....	195
Table 2.....	196

Chapter 6

Table 1.....	255
--------------	-----

ACKNOWLEDGEMENTS

I would like to dedicate my deepest appreciation to my PI, Professor David Eisenberg. David started me on the project based on my research interests, and he gave me invaluable and immeasurable guidance and advice during my PhD years. David gives me freedom to explore the scientific world of amyloid diseases; he tolerates my research failures; he tirelessly edits and improves my verbal and written English; he gives loving support through mine and our lab's toughest times. David molded me into a good scientist and a better person. I love and respect David as my mentor and as my friend.

I want to thank my committee Professor David Eisenberg, Professor Douglas Black, Professor Todd Yeates, and Professor Gerard Wong, who were not afraid or hesitant to give me critical suggestions when I needed them. Their help and support navigated me through my scientific discovery during my PhD years.

At the beginning of my PhD, I worked with a hard-working and brilliant graduate student Michael Hughes for a year and a half, and he gave me immense support and guidance for my project, and led me into the world of LCDs and amyloid fibrils. Many thank you to Hughes for making me the scientist I am right now.

I want to express my most sincere gratitude to Michael Sawaya and Duilio Cascio. When I joined the Eisenberg lab, I had minimum knowledge about the principles of crystallography and structure determination. M.S. and D.C. led me into the beautiful world of protein structures, took me on synchrotron trips, and taught me generously, tirelessly, and patiently about X-ray crystallography. Thank you both for making me a good structural biologist.

I would like to thank Peng Ge, Qin Cao, and David Boyer for their help and support with cryoEM. Without this method and their help, I would never get the structures of the LCD fibrils, thus the answer to my project. Thanks to them, I learnt solidly and quickly about the technical aspects of cryoEM.

Many thanks to my colleagues Kevin Murray and Romany Abskharon, who made significant contributions to my thesis work. K.M. started and guided me into the world of small molecule disaggregators. Without him, none of the current small molecule projects would exist. R.A. helped me tirelessly with any antibody-related experiments. I would also like to thank Michael Collazo, Genesis Falcon, Daniel Anderson, and Alex Lisker. M.C. and G.F. gave me support to set up crystal trays. D.A. and A.L. gave me technical help with lab chemicals and computers, respectively. In addition, thank you to the Eisenberg lab community for creating a warm and efficient environment to work in.

Special thanks to my collaborators from UT southwestern, Professor Steven McKnight, Masato Kato, and Siheng Xiang for their generosity to share the hnRNPA2 constructs with me and their help and advice on protein purification. I also want to thank collaborators from UCLA, Shino Magaki, Christopher Williams, and Ramiro Diaz for helping us acquire patient brains and advance amyloid research.

I would like to acknowledge the Whitcome Pre-doctoral Fellowship in Molecular Biology, Departmental Fellowship, Audree V. Fowler Fellowship, and Dissertation Year Fellowship for financial support.

Many thanks to my friends: Yiwen Yang, Yawen Zheng, Xinyi Cheng, Prashansa Ratan, Carolyn Hu, Chenxiang Wang, Bosi Peng, Yao Gong, Jun Shen for making my PhD journey a fun and fulfilled one.

Lastly, I want to dedicate my most genuine gratitude and appreciation to my parents Zugen Lu and Weiping Wang for raising me with immense care and love, and allowing me to pursue science at my free will in the US.

Thank you all for making my PhD a happy and memorable one. Thank you all for making me a good scientist. Thank you all for the help of advancing our knowledge of amyloid diseases.

VITA

Education

University of California, Davis, B.S. in Biochemistry and Molecular Biology 2017. 06

Research experience

Graduate Student Researcher, David Eisenberg lab, UCLA 2017.09—present

Research Assistant, Andrew Fisher lab, UC Davis 2016. 01 –2017. 08

Awards

NextGen Award for FASEB Protein Aggregation Conference 2019. 06

Whitcome Pre-doctoral Fellowship in Molecular Biology, UCLA 2019. 09 – 2020. 08

Departmental Fellowship, UCLA 2020. 01 – 2020. 03

Audree V. Fowler Fellowship, UCLA 2020. 09 – 2021. 09

Dissertation Year Fellowship, UCLA 2022. 07 – 2023. 06

Daniel E. Atkinson and Charles A. West Dissertation Award 2023.06

PUBLICATIONS

- Lu, J., Cao, Q., Hughes, M.P. et al. CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nat Commun* 11, 4090 (2020)
- Lu, J., Ge, P., Sawaya, M.R. et al. Cryo-EM structures of the D290V mutant of the hnRNPA2 low-complexity domain mask the nuclear import signal required for function (In preparation)
- Guenther, E.L. et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat Struct Mol Biol* 25, 463–471 (2018)
- Cao, Q. et al. Cryo-EM structures of hIAPP fibrils seeded by patient-extracted fibrils reveal new polymorphs and conserved fibril cores. *Nat Struct Mol Biol* 28, 724–730 (2021)
- Tayeb-Fligelman, E. et al. Low complexity domains of the nucleocapsid protein of SARS-CoV-2 form amyloid fibrils. *Nat. Commun.* 14, 2379 (2023)
- Murray, K. A. et al. Small molecules disaggregate alpha-synuclein and prevent seeding from patient brain-derived fibrils. *Proc. Natl. Acad. Sci.* 120, e2217835120 (2023)

OVERVIEW

This thesis work started with the finding made by McKnight's group that many LCDs of the RNPs can form a fibril network that macroscopically manifests as a hydrogel, and with the discovery made by Hughes et al. of Low-complexity Amyloid-like Kinked Segments (LARKS), a structural motif that has kinked backbones and acts as adhesive regions for LCD interactions. The functional RNP hnRNPA2 can form a hydrogel of fibril networks and a disease-related mutation D290V is found in its LCD. This thesis mainly details the story of the structure determinations for both wildtype and the variant hnRNPA2 LCD and how the comparison of both structures helps us better understand their functions and the pathogenic mechanism caused by the missense mutation. Peptide structures of hnRNPA2, TDP-43, and Sars-CoV-2 nucleoprotein are also studied in this thesis work to facilitate inhibitor design. My thesis work also involves finding small molecule disaggregators for amyloid proteins such as tau and α -synuclein. Former lab member Kevin Murray has identified a series of small molecules that can enter the blood brain barrier. Among them, CNS11G stands out for its active dissociation of both recombinant and brain-derived α -synuclein fibrils.

Chapter 1 is a reprint of the published manuscript that describes the cryoEM structural characterization of wildtype hnRNPA2 LCD fibril core and a six-residue peptide containing the disease-related mutation D290V. We found that wildtype hnRNPA2 fibrils have only one morphology, and this contains a single protofilament. The backbone of hnRNPA2 is highly kinked in the fibril core and enriched with LARKS, allowing it to have non-covalent crosslinking with itself and other LCDs. Thermostability assay and solvation energy calculations suggest hnRNPA2 fibril is less stable compared to pathogenic amyloid fibrils. We found that these properties are

shared with other RNPs such as FUS. The X-ray crystallographic structure of the D290V-containing peptide (GNYNVF) suggests that the missense mutation fundamentally alters the fibril structure to a more stable energetic state. The mCherry-tagged hnRNPA2 LCD constructs were kindly provided by Steven McKnight's group. Fibrillization, X-ray fibril diffraction, cryoEM preparation, data processing and structure determination were done by me with invaluable assistance from colleagues: cryoEM experts Qin Cao and David R. Boyer, crystallography and model building experts Michael R. Sawaya and Duilio Cascio. I used an in-house script written by Michael R. Sawaya to calculate the solvation energy of hnRNPA2 fibril core. I also did a 3-D environment profiling, originally developed by James U. Bowie and David S. Eisenberg, suggesting that wildtype hnRNPA2 fibril core has a higher percentage of solvent-exposed residues and a lower percentage of nonpolar residues compared to pathogenic amyloids. The D290V-containing six-residue peptide was identified in Kato et al., 2012. The crystal structure of GNYNVF forms a highly energetically stable steric zipper. This work was published in Nature Communications in 2020.

Chapter 2 is a reprint of the manuscript currently under review that describes how the cryoEM structures of the D290V variant of hnRNPA2 LCD fibrils explain their pathogenic mechanisms. I, with the aid of Michael P. Hughes and Einav Tayeb-Fligelman, designed and performed thermostability and phase separation assays which indicate that the variant hnRNPA2 LCD forms fibrils faster and is more stable than wildtype. The cryoEM structures of the variant hnRNPA2 LCD are fundamentally different from wildtype, and more similar to pathogenic amyloids. The variant structure shows polymorphism, each morphology with two or more protofilaments. There is a large increase in the content of β -sheet conformation residues in the variant structures comparing to wildtype. Solvation energy calculations show that the variant structure is in the

middle of the spectrum of solvation energy values between FUS and pathogenic Tau. This result implies that stability of the variant is not the sole reason for its pathogenicity, which leads us to seek other reasons. We found the PY-NLS is hidden in all three polymorphs of the variant hnRNPA2 structure, which will disrupt hnRNPA2's designated localization. We also noticed that half of the fibril core-facing residues in wildtype structure turns solvent-facing in the variant structures. This residue facing-flip, together with the solvent exposed disease-causing mutation Valine 290, contributes to new surface exposed epitopes, which could potentially lead to the variant's toxicity in cells. Fibrillization, cryoEM preparation, data processing and structure determination were done by me, with crucial help from cryoEM experts Peng Ge and David R. Boyer. Peng Ge was essential in determining the variant structures. Model building and refining were done by me, with invaluable assistance from Michael R. Sawaya and Duilio Cascio. Structure analysis and energy calculations were done with the help of Michael R. Sawaya and Michael P. Hughes. This work is submitted to (?).

Chapter 3 is a reprint of the published manuscript that describes the discovery of small molecules CNS11 and CNS11G and their active dissociation of α -synuclein fibrils both in vitro and in vivo. This work was primarily done by Kevin A. Murray and Carolyn J. Hu. My contributions were to test in vitro CNS11G's effect on α -synuclein fibrils dissociation, both MSA brain-derived and recombinant. The discovery of this paper is promising for synucleinopathy drug development because it provides therapeutic leads for targeting α -synuclein fibrils.

Chapter 4 is a reprint of the published manuscript which describes steric zippers and LARKS structures from an RNP TDP-43 associated with stress granules. This work was primarily done by Elizabeth Guenther and Qin Cao. My contributions were to determine the six-residue peptide structures of AALQSS and AMMAAA, two sequences from the LCD of TDP-43. This paper

determines 10 peptide crystal structures and suggests a hypothetical pathway of how familial ALS-related mutations convert LARKS to steric zippers. This work is important because it provides an explanation for a transition of reversible to irreversible pathogenic aggregation.

Chapter 5 is a reprint of the published manuscript which describes the cryoEM structures of recombinant IAPP seeded by brain-derived human IAPP. This work was primarily done by Qin Cao and David R. Boyer. My contributions were to help with cryoEM data processing. This paper details the characterization of four different polymorphs of IAPP, and comparison to the S20G variant IAPP structure. This work is important because three polymorphs identified are previously unreported and comparisons of wildtype to S20G hIAPP fibril structures suggest a mechanism for how S20G facilitates fibrillization.

Chapter 6 is a reprint of the published manuscript that describes the inhibition of amyloid formation of Sars-CoV-2 nucleocapsid protein. This work is a collaborative work led by Einav Tayeb-Fligelman during Covid. This paper involves the identification of an LCD in the Sars-CoV-2 nucleocapsid protein (NCAP), fibrillization of NCAP LCD, identification of six-residue peptides that are critical for fibrillization, and inhibitor designs. My contributions were to determine the crystal structures of six-residue peptides. This paper describes the inhibitors that we designed can decrease NCAP LCD fibrillization. This work is unique for showing that NCAP can not only phase separate, but also fibrillize.

Chapter 1

CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid

Abstract

hnRNPA2 is a human ribonucleoprotein (RNP) involved in RNA metabolism. It forms fibrils both under cellular stress and in mutated form in neurodegenerative conditions. Previous work established that the C-terminal low-complexity domain (LCD) of hnRNPA2 fibrillizes under stress, and missense mutations in this domain are found in the disease multisystem proteinopathy (MSP). However, little is known at the atomic level about the hnRNPA2 LCD structure that is involved in those processes and how disease mutations cause structural change. Here we present the cryo-electron microscopy (cryoEM) structure of the hnRNPA2 LCD fibril core and demonstrate its capability to form a reversible hydrogel in vitro containing amyloid-like fibrils. Whereas these fibrils, like pathogenic amyloid, are formed from protein chains stacked into β -sheets by backbone hydrogen bonds, they display distinct structural differences: the chains are kinked, enabling non-covalent cross-linking of fibrils and disfavoring formation of pathogenic steric zippers. Both reversibility and energetic calculations suggest these fibrils are less stable than pathogenic amyloid. Moreover, the crystal structure of the disease-mutation-containing segment (D290V) of hnRNPA2 suggests that the replacement fundamentally alters the fibril structure to a more stable energetic state. These findings illuminate how molecular interactions promote protein fibril networks and how mutation can transform fibril structure from functional to a pathogenic form.

Introduction

The protein hnRNPA2 functions in RNA stabilization, splicing, trafficking, and translation^{1,2,3,4}. In its role in protecting mRNAs, it is a component of cytoplasmic stress granules, one of the so-called membraneless organelles (MLOs)^{1,5}. Its 341 amino-acid-residue sequence contains an RNA-binding domain (RBD) with two RNA recognition motifs (RRMs)

and a 161-residue C-terminal low-complexity domain (LCD). The LCD is rich in Gly, Tyr, Phe, Asn, and Asp residues, and poor in hydrophobes, with Val, Ala, Leu, and Ile entirely absent. Regions of biased amino acid composition, such as this, allow transient interactions between the many known MLO-forming proteins^{1,5,6,7,8}.

Aggregated hnRNPA2 has been found to have nuclear clearance and cytoplasmic inclusions under cellular stress², and in vitro its LCD forms fibrils with the capacity to interact with each other to make hydrogels of the sort observed in MLOs^{1,6}. In vitro studies have shown evidence of liquid droplets over time turning into hydrogels containing fibrils^{9,10}. The hnRNPA2-LCD hydrogel can trap other functional proteins such as hnRNPA1 and CIRBP¹, suggesting a functional role for the hydrogel.

The fibrils of MLOs are functional in contrast to pathogenic fibrils associated with neurodegenerative conditions. One such condition is the disease MultiSystem Proteinopathy (MSP) with symptoms indistinguishable from ALS and FTD. Patients with this disease present with aggregated hnRNPA2 in cytoplasmic inclusions² and the variant sequence D290V in the LCD of hnRNPA2. Other mutations found in the RNP LCDs were previously shown to impede MLO and hydrogel formation^{1,11,12}.

Informing both functional MLO-associated fibrils and pathogenic disease-associated fibrils, hnRNPA2 is similar to other RNA-binding proteins, including hnRNPA1¹³, FUS^{7,9,12,14}, TIA1¹¹, and TDP-43^{15,16}. For this reason, knowledge of the structures of these proteins, and of their LCDs in particular, may uncover general principles of how proteins can form both functional and pathogenic fibrils. Near-atomic structures are already available for the ordered fibril cores of the LCDs of FUS¹⁴ and TDP-43¹⁵.

Here we determined the cryoEM structure of the LCD of hnRNPA2 and the crystal structure of the segment containing the variant sequence D290V. These structures help to answer the questions: (1) What sequence features account for the formation of the functional fibrils? (2) What structural features of the functional fibrils drive them to bind non-covalently with each other to form the networks that underlie hydrogels? And (3) How does a single residue mutation convert functional to pathogenic fibrils?

Results

Hydrogel and fibril formation of mC-hnRNPA2-LCD

Purified recombinant mC-hnRNPA2-LCD was concentrated to $\sim 60 \text{ mg ml}^{-1}$ and incubated at 4°C to test its ability to form a hydrogel. Determination of whether the sample exists in a liquid or gel phase is assayed by the mobility of a bubble through the sample. We inverted a 1.5 ml silicon tube containing freshly purified mCherry-tagged (purple color) hnRNPA2-LCD (termed mC-hnRNPA2-LCD), and then introduced a bubble at the lower surface. The bubble rose to the top of the tube, indicating a homogenous solution rather than a gel (Fig. [1a](#), left). Within a week, the protein solution converted into a gel, indicated by the retention of bubbles at the lower surface (Fig. [1a](#), right). Negative-stain transmission electron microscopy (TEM) of the diluted hydrogel showed a network of uniform amyloid-like fibrils with an average width of 20 nm (Fig. [1b](#)). The X-ray diffraction pattern is consistent with cross- β architecture: two reflections at 4.7 and 10 Å corresponding to the inter-strand and inter-sheet spacing, respectively¹⁷, of amyloid fibrils (Fig. [1c](#)).

Reversibility of mC-hnRNPA2-LCD hydrogel

To test whether the hnRNPA2-LCD hydrogel is reversible, we performed a reversibility assay on purified mC-hnRNPA2-LCD incubated at 4 °C for either 2 days or 2 weeks. After heating the 2-day hydrogel to 45 °C, the bubbles introduced stayed at the bottom of the sample, indicating maintenance of a hydrogel state (Supplementary Fig. [1a](#), left). TEM images showed networks of uniform, twisted, and unbranched amyloid-like fibrils similar to those observed before heating (Supplementary Fig. [1b](#), left). However, after heating to 55 °C, introduced bubbles slowly rose and stalled in the middle of the hydrogel (Supplementary Fig. [1a](#), middle). TEM of these samples showed fragmented fibrils and partial disruption of fibril networks (Supplementary Fig. [1b](#), middle). After heating to 65 °C and above, the hydrogel transitioned to a homogenous protein solution where bubbles introduced quickly rose to the top of the silicon tube (Supplementary Fig. [1a](#), right), indicating the 2-day hydrogel is reversible. TEM revealed that over 80% of the fibrils are dissolved, leaving behind amorphous aggregates (Supplementary Fig. [1b](#), right). For the 2-week hydrogel, bubbles introduced remained immobile up to 75 °C heating (Supplementary Fig. [1c](#)) and TEM images (Supplementary Fig. [1d](#)) revealed a fibril morphology similar to the 2-day hydrogel. Hence, the 2-week hydrogel appears less reversible than the 2-day hydrogel. Moreover, the 2-week fibrils at 65 °C were more bundled than the remnant of 2-day hydrogel fibrils at 65 °C. We infer that the young hnRNPA2 fibrils are mostly reversible, but aging makes the fibrils irreversible, possibly by the evolution of fibril bundling. We also purified and concentrated the mCherry tag alone as a negative control, and we found that fibril networks and the amorphous aggregates were not the result of mCherry itself (Supplementary Fig. [2](#)).

CryoEM structure of the mC-hnRNPA2-LCD fibrils

To answer the question of why the mC-hnRNPA2-LCD hydrogel exhibits reversibility, we determined the cryoEM structure of mC-hnRNPA2-LCD fibrils to a resolution of 3.1 Å (Fig. [2c](#)).

During cryoEM data collection and processing, we found that the mC-hnRNPA2-LCD fibrils have only a single morphology, which consists of one asymmetric protofilament with a pitch of 600 Å, a left-handed helical twist of -2.88° and a helical rise of 4.81 Å (Fig. [2b](#) and Supplementary Fig. [3](#)). Contrary to the globular proteins which form 3D structures, the mC-hnRNPA2-LCD structure is confined to 2D layers which stack on top of each other, forming a twisted in-register β -sheet that runs along the fibril axis. Out of 161 residues in the LCD, 57 residues from Gly263 to Tyr319 form a fibril core (Fig. [2a](#)). The mC-hnRNPA2-LCD fibril encompasses a PY-nuclear localization signal (PY-NLS) and the core hexamer segment containing the site for a disease-causing mutation that was identified by Kim et al. in 2013² (Fig. [2d](#)). Previous studies have shown that mutations in the PY-NLS can lead to ALS¹⁸. An aromatic triad composed of two Tyr (Y275 and Y283) residues and one Phe (F309) residue is buried in the center of the fibril, stabilizing the structure through π - π interactions (Supplementary Fig. [4](#)). Data collection and refinement statistics are summarized in Table [1](#).

Energy analysis explains hnRNPA2-LCD fibrils' reversibility

To quantify the structural features that led to the reversibility of hnRNPA2-LCD fibrils, we calculated the solvation energy for our mC-hnRNPA2-LCD fibril structure (details in “Methods” section). We obtained the solvation energy per layer of the fibril and an averaged solvation energy per residue. We illustrated these energies in a solvation energy map where each residue is colored according to its energy. The solvation energy for mC-hnRNPA2-LCD fibril core ($-19.5 \text{ kcal mol}^{-1}$ per chain and $-0.34 \text{ kcal mol}^{-1}$ per residue) is comparable to another well-known reversible protein fibril, FUS ($-12.2 \text{ kcal mol}^{-1}$ per chain and $-0.20 \text{ kcal mol}^{-1}$ per residue) (Fig. [3a, b](#)). In contrast the solvation energy of pathogenic irreversible amyloid structures such as human serum amyloid A¹⁹ ($-34.4 \text{ kcal mol}^{-1}$ per chain and

$-0.64 \text{ kcal mol}^{-1}$ per residue) has a more negative value per chain and residue (Fig. [3c](#)), indicating that compared to the pathogenic amyloid fibrils, mC-hnRNPA2-LCD fibrils form less stable structures. Solvation energies of other proteins are compared in Table [2](#).

The disease-causing mutant core segment

Muscle biopsies have found atrophic hnRNPA2 fibrils in patients diagnosed with inclusion body myositis (IBM)[2](#). A missense point mutation that converts a conserved aspartic acid to a valine (D290V) in hnRNPA2-LCD was identified in 2013 as being associated with ALS and MSP[2](#). To investigate the structural differences between the wildtype and the mutant, we crystallized the hexamer segment $^{286}\text{GNYN}\underline{\text{V}}\text{F}^{291}$ (Supplementary Fig. [5](#)) containing the disease-causing mutation and determined its structure by X-ray crystallography (Fig. [4a](#) and Statistics in Table [3](#)). As expected, the crystal structure of the mutant segment forms a steric zipper motif and the sidechains of the two β -sheets mate tightly with each other in a dry interface. The area buried (A_b) and shape complementarity (S_c) for the mutant segment are 104 \AA^2 per chain and 0.86, respectively. The mutant segment structure forms in-register, parallel β -sheets. To compare this mutant segment crystal structure with our cryoEM wildtype mC-hnRNPA2-LCD structure, we overlaid the two structures (Fig. [4b](#)). The backbones of the wildtype structure and the mutant segment fit well with each other, whereas the sidechain of Asp290 in the wildtype structure clashes with Asn289 of the mating sheet of mutant steric zipper, disrupting the steric zipper formation. On the other hand, forming the steric zipper in the mutant structure causes Asn287 and Tyr288 sidechains to adopt different conformations, which create steric clashes with nearby Asn282 and Met304 residues in the wildtype fibril structure. Moreover, Phe291 from the mating sheet of the steric zipper clashes with the backbone of Gly281 in the wildtype fibril structure. Based on this structural analysis, we speculate that the D290V mutation converts the reversible

hnRNPA2 fibrils to irreversible fibrils. When the conserved Asp290 residue is mutated to a Val, the mutation enables a more stable steric zipper interaction that was previously prevented by the Asp290 sidechain. Meanwhile, the formation of a steric zipper causes multiple steric clashes and disrupts the wildtype fibril structure, shifting the aggregation of hnRNPA2 from reversible to irreversible.

To compare the effects on aggregate formation of the wildtype core segment GNYNDF and the mutant segment GNYNVF, both were shaken for 4 days and the samples were checked using negative stain TEM (details in “Methods” section). The mutant segment, as we expected, precipitated after shaking (Fig. [4c](#), left), and TEM images showed that the precipitates were composed of micro-crystals (Fig. [4d](#), left). The wildtype segment forms a hydrogel (Fig. [4c](#), right), which to our knowledge is the shortest segment that forms a hydrogel, and TEM images indicated that the gel was composed of thin fibrils (Fig. [4d](#), right). The results suggest that the mutant segment with the pathogenic mutation forms a more stable structure than the WT segment, and converts a gel-promoting segment into an aggregation-promoting segment.

Discussion

mC-hnRNPA2 LCD fibrils reveal distinctly different structural and energetic properties from pathogenic amyloid (Table [4](#)). Our TEM and X-ray diffraction data show that mC-hnRNPA2-LCD forms fibril networks with cross- β structures (Fig. [1](#)). Unlike pathogenic amyloid fibrils, which are usually not crosslinked and are stable at high temperatures^{[15](#)·[16](#)·[20](#)}, mC-hnRNPA2-LCD fibrils form 3-dimensional networks and are sensitive to even mild heat (Supplementary Fig. [1](#)). Heated mC-hnRNPA2-LCD, instead of remaining as a fibril network, is broken into oligomer-like material, and does not reform a hydrogel upon cooling. We hypothesize that in this in vitro experiment, the environment lacks biological factors such as chaperones to achieve

reversibility. Nevertheless, the smaller stabilization solvation energy compared to pathogenic amyloid indicates the reversibility of hnRNPA2 LCD fibrils (Fig. 3). Amyloid-like fibrils have also been reported for other granule-associated, LCD-containing proteins, such as FUS^{1,21} and hnRNPA1¹³, those being heat-sensitive and SDS-sensitive, and being capable of forming a reversible hydrogel.

Our cryoEM structure of mC-hnRNPA2-LCD fibrils reveals local structural motifs that distinguish reversible from pathogenic amyloid: mC-hnRNPA2-LCD fibrils contain multiple weakly binding LARKS motifs²² and only a single short steric zipper motif²³. LARKS are low-complexity amyloid-like, reversible kinked segments which enable amyloid-like assembly with reduced stability compared to steric zippers²². Steric zippers are pairs of β -sheets in which the sidechains of one β -sheet interdigitate with those of the opposing sheet to form a tight, dry interface^{23,24}. Unlike steric zippers, the β -sheets formed by LARKS are often kinked either at a Gly or aromatic residues²², limiting the size of interacting surfaces. Consequently, LARKS are less stabilizing than steric zippers. Within the mC-hnRNPA2-LCD fibril core there are eight predicted LARKS (Hughes et al., in preparation), seven grouped together (Supplementary Fig. 6a and Fig. 2d), each corresponding to a kink in the protein backbone. The structure contains only four short β -sheets, reducing the possibility of steric zippers (Fig. 2d). The multiple LARKS motifs introduce kinks into the backbone of the structure, disrupting stabilizing β -sheets into segments shorter than the six residues of most steric zippers, thus contributing to the reversibility of the hnRNPA2-LCD fibrils. Pathogenic amyloid typically houses a high content of β -sheets capable of forming steric zippers, for example Tau PHFs²⁵ (78%) and A β ₄₂²⁶ (74%) in contrast to reversible amyloid-like fibrils hnRNPA2 (23%) and FUS¹⁴ (26%). We note the location of the short β -sheets in hnRNPA2-LCD are consistent with the secondary structure prediction of

Murray et al.[27](#) (Fig. [2d](#)). We regard LARKS and steric zippers as the structural basis of “stickers”[28:29](#) in reversible, amyloid-like fibrils and pathogenic amyloid fibrils, enabling inter-protein and intra-protein interactions.

A further feature that distinguishes reversible amyloid-like fibrils from irreversible, pathogenic fibrils is the enrichment of polar residues in the fibril core. As shown in the mC-hnRNPA2-LCD residue polarity map, almost 50% of the residues are hydrophilic and only 12% are hydrophobic (Supplementary Fig. [6b](#)), and the core contains a high content of Gly (35%), Asn (18%), and Tyr (16%) compared to tau PHFs[25](#) (Gly:12%, Asn:4%, Tyr:1.4%) and A β 42[26](#) (Gly:14%, Asn:2%, Tyr:2%). Furthermore, hnRNPA2-LCD lacks any instance of the hydrophobic residues Val, Ala, Ile, and Leu. In contrast, tau PHFs[25](#) and A β 42[26](#) contain 24% and 36%, respectively, of these residue types (Table [5](#)). The high content of Gly contributes to hnRNPA2-LCD flexibility. 3D environment profiling[30](#) also suggests a higher percentage of residues exposed to solvent (35%) and a lower content of nonpolar residues (4%) compared to tau PHFs[25](#) (12% and 13%) and A β 42[26](#) (14% and 18%) (Table [6](#)). These differences are also manifested in solvation energy maps (Fig. [3](#)). Our results indicate that hnRNPA2-LCD adopts a cross- β fibril assembly that morphologically resembles pathogenic amyloid. However, quantitative examination shows that reversible amyloid-like fibrils differ from pathogenic amyloid by having a smaller solvation stabilization energy and a greater enrichment of LARKS and polar residues. These features facilitate hydrogen-bonding with water molecules and contribute to LCD flexibility and capacity for assembly and disassembly in stress granules.

HnRNPA2 and FUS[14](#), both proteins that function in stress granules, form fibrils which share four structural features. First, both are mainly stabilized by hydrogen-bonding and polar interactions rather than hydrophobic effects. Both hnRNPA2-LCD and FUS-LCD fibril cores are

enriched in Tyr residues that stabilize structures by π -stacking interactions^{14:24:31}. Similar to hnRNPA2-LCD, the FUS-LCD fibril core has ample polar residues (Table 5, columns 2 and 3), and FUS-LCD has a high content of glutamine residues that are either exposed to solvent or form stabilizing hydrogen-bonded ladders along the fibril axis. The abundance of residues with hydroxyl groups in both hnRNPA2 and FUS (Table 5, column 3) allows post-translational modifications, such as phosphorylation³², to modulate the stability of the fibrils, affecting stress granule assembly. Compared to hnRNPA2-LCD, the FUS-LCD has a relatively high content of Ser and Thr residues, allowing additional stabilization of the core structure through sidechain hydrogen bonding.

Second, solvation energy calculations suggest that the stability of the mC-hnRNPA2-LCD fibril structure is nearly as poor as FUS-LCD. Because the LCDs in hnRNPA2 and FUS are reversible fibrils associated with stress granules, we expect both to have lower stability than pathogenic amyloid fibrils. And this is confirmed by our solvation energy calculation (Fig. 3b and Table 2). The relatively small solvation stabilization energy explains why hnRNPA2 and FUS are sensitive to mild heat and denaturing condition such as 2% SDS¹⁶.

Third, hnRNPA2 and FUS fibrils exhibit only a single morphology. In contrast, a majority of the studied structures of pathogenic irreversible amyloid have shown polymorphism, whether the fibrils were produced *ex vivo* or *in vitro*. For example, so far, cryoEM and ssNMR studies have identified nine morphologies for different isoforms of pathogenic protein tau^{25:33}, three morphologies for TDP-43¹⁵, and six morphologies for A β 42^{26:34:35:36:37:38}. Moreover, multiple fibril polymorphs have been structurally determined under a single set of conditions from the same EM grid^{15:39}. Unlike the polymorphism seen in the pathogenic irreversible fibril structures, our cryoEM structure of mC-hnRNPA2-LCD fibril shows only a single morphology

under the condition tested. This lack of polymorphism is also seen in the FUS-LCD structure that was determined by ssNMR¹⁴. As suggested previously, monomorphic structure is essential for proper biological functions of proteins¹⁴. Similar to globular proteins, reversible amyloid-like fibrils have apparently been honed by evolution to adopt a single structure to support a given function. Reversible fibrils such as hnRNPA2 can undergo folding and unfolding to find a kinetically accessible global free energy minimum, leading to an optimal structure of lowest accessible free energy. The hnRNPA2-LCD is conserved among distantly related organisms (Supplementary Fig. 7), further evidence of its functional role. Fourth, both FUS-LCD and mC-hnRNPA2-LCD fibrils are composed of a single protofilament rather than two or more, as is the case with more than 70% of pathological fibrils structurally determined.

Enrichment in LARKS motifs also diminishes the tendency for polymorphism. Interactions between two LARKS motifs are more specific than steric zippers since complementarity between kinked surfaces requires a particular sequence, whereas complementarity between two canonical flat β -sheets can be achieved by more sequences. Furthermore, a kinked sheet of a given sequence has few plausible ways to fit with a partner kinked sheet, whereas two flat β -sheets have a range of plausible registrations with respect to each other to achieve different low-energy structures. The polymorphism seen in pathogenic irreversible fibrils represents different structures that are trapped in multiple local free energy minimum. The hnRNPA2 being reversible and functional in vivo implies that it should also follow the Anfinsen's dogma⁴⁰ that the native full-length hnRNPA2 fibrils in vivo should have the same structure as the recombinant protein structure we report here.

Our cryoEM hnRNPA2-LCD structure is consistent with previous biochemical and NMR studies. Xiang et al. in 2015⁶ reported that the hnRNPA2-LCD adopts a similar structure in

hydrogels and in liquid droplets as evidenced by conservation of NAI footprints which probed 23 residues of hnRNPA2-LCD from Ser219 to Ser335. They found the N-terminal portion was unprotected from NAI labeling, but the C-terminal portion encompassing N282–Y324 was protected. This protected region overlaps well with our ordered core spanning residues G263–Y319. The identity of the only unprotected probe residue in this range, Lys305, agrees with our cryoEM structure showing that the sidechain of Lys305 points toward the solvent. There is also agreement between the partial protection of residues Asn282, Tyr288, and Ser312 and our observation that all three are partially buried in the fibril core. Similar consistency is found between the strong protection of Ser285, Ser306, and Ty319 and our observation that these residues are completely buried (Fig. [2c](#)). The structural consistency among different structural studies of hnRNPA2-LCD, specifically our cryoEM, Murray et al.'s NMR²⁷, and Xiang et al.'s NAI footprinting⁶, suggests two important points: (1) the fibril core structure of hnRNPA2-LCD is identical or closely similar among three studies, (2) the interactions of residue sidechains we report are closely similar as Xiang et al.'s observations in both hydrogel and liquid droplets.

We offer a speculative hypothesis of how a single-point missense mutation can alter the fibril structure. This hypothesis is consistent with the data we have. As indicated by its greater stabilization energy, the mutant segment GNYNVF is more prone to form crystals than the wild type segment GNYNDF (Fig. [4](#)). In addition, we calculated the stabilization energy of the mutant segment and all hexamer segments that constitute the wild type hnRNPA2 LCD fibril, FUS LCD fibril, and pathogenic amyloids. Unsurprisingly, the mutant segment is an outlier in terms of having greater energetic stability than any hexameric segment in either hnRNPA2 or FUS (Supplementary Fig. [8a](#)). This energetic analysis supports our hypothesis that when this mutation forms, the mutant segment can drive the assembly of a more stable fibril.

If the wild type segment GNYNDF of full-length hnRNPA2 assembles as it does in the peptide crystal structure, the remaining residues would necessarily adopt a different conformation than the wild type fibril in order to avoid steric clash with the hexapeptide assembly. Our cryoEM images offer evidence that the mC-hnRNPA2-D290V-LCD fibrils are more heterogeneous and thicker than the wild type mC-hnRNPA2-LCD fibrils (Supplementary Fig. [8b, c](#)). This supports our hypothesis that the D290V mutation drives the mutant segment GYNNVF to adopt a homo-zipper conformation and drives the fibril assembly into a different structure with two protofilaments.

Previous studies by Guo et al.[41](#) showed that over half of the LCD-containing RNPs such as FUS, TAF15, hnRNPA1, and hnRNPA2 possess a PY-NLS. They found that the import receptor karyopherin- β 2 (Kap β 2) binds the PYNLS signal of these RNPs and then, once bound, acts as chaperones that dissolve pre-formed fibrils. The PYNLS of hnRNPA2 (residues 296–319), containing two epitopes: a hydrophobic/basic residue stretch (residue 196–312) and a consensus sequence R/K/H-X₍₂₋₅₎-PY (residue 313–316), is located right after a central steric zipper motif (segment ²⁸⁶GNYNDF²⁹¹) that drives hnRNPA2 fibrilization[2](#). Our structure of hnRNPA2 provides insight into how this interaction of Kap β 2 with hnRNPA2 functions to prevent fibrilization from occurring.

The hnRNPA2 fibril core is essentially a highly kinked, branched β -arch. The potential fibril-driving steric zipper motif (286–291) forms the only extended (unkinked) β -strand in the fibril structure and contributes to the upper half of the β -arch (Fig. [2d](#)). The lower half of the β -arch is part of the PYNLS signal (residues 296–319). The PYNLS signal forms a hydrophobic pocket between P298 and N300 that buries the terminal aromatic residue of the steric zipper F291 in the fibril core, yielding a significant amount of the stabilization energy (Fig. [3a](#)). We infer that if

Kap β 2 binds to the PYNLS signal, the bottom half of the β -arch is unavailable to make the hydrophobic environment necessary to sequester F291 to stabilize the central steric zipper and form fibrils. In addition, we propose that owing to the labile nature of the hnRNPA2 fibrils, monomers may more easily dissociate from the fibril and be sequestered by Kap β 2. This is supported by the relatively weak stabilization energy of hnRNPA2 fibrils compared to irreversible amyloid as the energy barrier between monomers associated with the fibril end or freely floating is much lower. The mechanism of action that our structure suggests is that hnRNPA2 fibrils cannot form when associated with the Kap β 2 via the PYNLS⁴¹ because the PYNLS is required to support the steric zipper motif that drives fibril formation².

Our cryoEM mC-hnRNPA2-LCD structure is consistent with and presents more evidence to support Guo et al.'s⁴¹ discovery that the PYNLS is included in the structure yet not completely buried in the middle of the fibril core (Fig. [2c, d](#)), which could allow chaperones such as Kap β 2⁴² to engage and melt the hnRNPA2-LCD hydrogel.

Does the mCherry-tag affect the structure of mC-hnRNPA2-LCD fibrils? We actively sought to characterize the hnRNPA2 LCD fibril structure without the mCherry tag; however, we found that hnRNPA2 LCD alone is highly insoluble and tends to aggregate. We thus included the mCherry tag in our experiments. mCherry is visible during cryoEM processing. As shown by EM images, the width of the mC-hnRNPA2-LCD fibril (~20 nm) (Supplementary Fig. [9a](#)) is significantly larger than that of our hnRNPA2-LCD fibril core model (2.6–6.7 nm) (Supplementary Fig. [9d, e](#)). During cryoEM data processing, we found globular densities surrounding the fibrils in low-resolution 2D classifications (Supplementary Fig. [9b](#)), and these globular densities were averaged into a fuzzy coat in high-resolution 2D classifications and 3D reconstruction (Supplementary Figs. [3c](#) and [9d](#)). The width of the fibril including the globular densities or fuzzy

coat in 2D classification is about 22 nm which matches the width of the fibrils measured in EM micrographs. The average diameter of each globular density is about 2.9 nm and is compatible with the diameter of the crystal structure of mCherry protein⁴³ (2.8 nm, PDB 2H5Q), which suggests these globular densities correspond to mCherry (Supplementary Fig. [9c](#)), as shown by the in-scale fitting of our model of mC-hnRNPA2-LCD fibril and the crystal structure of mCherry to the averaged 2D class image (Supplementary Fig. [9c](#)). There are three reasons that we consider innocuous the effect of such a large tag on the core structure. First, there are 86 amino acids between the mCherry-tag and the start of the fibril core (N-terminal G263), including a four-residue linker and the beginning segment of the LCD. We think that an 86-residue-long flexible peptide allows enough freedom to accommodate the mCherry tags without distorting the fibril core structure. When the fluorescent tag is sufficiently separated from the fibril core, the tag can freely fold around the fibril core as a “fuzzy coat”, so that fibrillization and core packing are not disturbed⁴⁴. Second, the biological, full-length hnRNPA2 contains two globular N-terminal RRM domains totaling 193 residues. Although the mCherry-tag is a little larger than the RRM domains, it is possible that the mCherry tag provides a mimic of the natural RRM domains in the full-length fibril. Protein–RNA interactions are important for assembling RNP granules in that RNA molecules act as scaffolds for multivalent RNA-binding proteins⁴⁵ such as hnRNPA2. These globular densities of mCherry suggest a potential arrangement of the RRM domains as they wrap around the hnRNPA2-LCD fibril in vivo. With enough sequence length between the RRM domains and the fibril core-forming region, the RRM domains could interact with RNA molecules to facilitate the RNP granule assemblies⁴⁶. Third, the mCherry tag “fuzzy coat” is not visible in the cryoEM 3D classification (Supplementary Fig. [9d](#)), suggesting there are no disruptive interactions between

the large mCherry tag and the hnRNPA2 LCD fibril core. Thus, the mCherry tag may suggest a model for fibril function.

Our studies indicate that hnRNPA2 fibrils together with other RNPs such as FUS share intrinsic structural and energetic properties that distinguish them from pathogenic amyloid in two major ways: (1) more LARKS motifs and fewer steric zippers, (2) enrichment of polar residues and scarcity in hydrophobic residues. These functional amyloid-like fibrils are likely the constituents of liquid condensates that contain proteins with LCDs. Experiments to detect the cross- β diffraction signal from fibrils of hnRNPA2 in solution are ongoing and made challenging due to background scattering arising from solvent and the β -sheet-rich mCherry tag (Supplementary Fig. 9). High protein concentration or mutations can drive such functional fibrils into more stable pathogenic amyloid fibrils. Cells need to maintain a delicate thermodynamic balance: biomolecules need to be concentrated enough to overcome the entropic loss associated with assembly into cellular condensates. Yet condensates can be easily converted by mutations, crossing a high kinetic barrier⁴⁷ to form irreversible fibrils in MLOs that feed forward into disease states.

Methods

Materials and purification of mCh-hnRNPA2-LCD fusion protein

The construct for overexpression of mCherry-hnRNPA2-LCD fusion protein was provided by Dr. Masato Kato of University of Texas, Southwestern, and has the sequence shown in Supplementary Fig. 10. Protein overexpression and purification procedure was adapted from the protocol by Kato et al. (2012)¹. Protein was overexpressed in *E. coli* BL21(DE3) cells with 0.5 mM IPTG at 25 °C for overnight. LB media with 0.1 mg ml⁻¹ ampicillin was used for cell

culture. Harvested cells were resuspended in lysis buffer containing 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 M guanidine HCl, 2 mM TCEP, and Halt™ protease inhibitor cocktail (Thermo Scientific) for 30 min on ice, and then sonicated. The cell lysate was centrifuged at 32,000×g for an hour. The supernatant was filtered and loaded onto a HisTrap HP column(GE healthcare) for purification. The HisTrap column was pre-equilibrated with the lysis buffer. After proteins were loaded onto the column, proteins were washed with the lysis buffer until the UV280 spectrum line became flat. The sample then was washed with a gradient from 100% wash buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 M Guanidine HCl, 20 mM imidazole, and 2 mM TCEP to 100% elution buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 M Guanidine HCl, 300 mM imidazole, and 2 mM TCEP. Eluted proteins were flash frozen by liquid nitrogen and stored at -80 °C for future use.

Formation of mCherry-hnRNPA2-LCD hydrogels

Purified mCherry-hnRNPA2-LCD fusion proteins were dialyzed overnight at room temperature against a dialyzing buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA, and 0.1 mM PMSF. The protein solutions were concentrated to 40–80 mg ml⁻¹. The protein solutions (~100 µl) were filled into tightly sealed 1.5 ml silicon tubes, and the tubes were incubated at 4 °C for 1–3 days.

Negative stain TEM and X-ray diffraction

All protein samples for TEM were diluted 10 times using dialysis buffer. Samples for TEM were prepared by applying 4 µl of sample on glow-discharged 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The samples were allowed to adhere for 2 min and 30 s, and washed twice with water. The samples were then stained for 2 min with 2%

uranyl acetate and allowed to dry for 1 min. Each grid was inspected using a T12(FEI) electron microscope.

For fiber diffraction, the procedure followed the protocol described by Rodriguez et al.[48](#). The hydrogel sample placed between two capillary glass rods, and allowed to air dry. When the fibrils are completely dry, the glass rods with fibrils aligned in between were mounted on a brass pin for x-ray diffraction. For homogenous solution, 10 μl of 60 mg ml^{-1} protein solution is pipetted into a PET sleeve and then mounted on a brass pin. Buffer was used as a control and used for background subtraction. X-ray diffraction data were collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA, at a wavelength of 0.979 Å and temperature of 100 K. Data were collected using 2.5° oscillations and 450 mm detector distance with an EIGER detector. The results were analyzed using the Adxv software[49](#).

The 3D classification of mCherry-hnRNPA2-LCD structure indicates the helical rise of the fibrils is 4.8 Å. This is different from the results we got from the hydrogel fiber diffraction, showing the inter-strand spacing to be 4.7 Å. We speculated that this difference is due to the sample preparation. For fiber diffraction, fibrils are dried between the two glass rods at room temperature for hours before being shot with X-ray. Yet CryoEM samples are dried by being blotted for only several seconds. Fibrils being drier might lead to a denser packing of the structures. Therefore in in vitro experiments, attentions should be given to the sample preparation steps.

Reversibility assay of mCherry-hnRNPA2-LCD hydrogels

Bubbles were introduced to the hydrogels as a way to tell if the hydrogels are reversible.

Hydrogels were heated up from 4 to 75 °C with a 10 °C increment, at each temperature samples

were incubated for 10 min in a PCR machine. 2 μl of samples at each temperature were taken and diluted 10 times with the dialysis buffer, and checked by negative stain EM.

CryoEM data collection, reconstruction, and model building

2.5 μl of diluted hydrogel samples at a concentration of 0.43 mg ml^{-1} were applied to glow-discharged Quantifoil Cu R1.2/1.3, 300 mesh carbon grids. Samples were blotted with filter paper to remove excess sample and then plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI). Cryo-EM data were collected on a Gatan K2 Summit direct electron detector on a Titan Krios (FEI) microscope equipped with a Gatan Quantum LS/K2 Summit direct electron detection camera (operated with 300 kV acceleration voltage and slit width of 20 eV). Super-resolution movies were acquired with a pixel size of 1.07 \AA pixel^{-1} (0.535 \AA pixel^{-1} in super-resolution movie frame). Thirty movie frames were recorded each with a frame rate of 5 Hz using a dose rate of 1.15 $\text{e}^{-} \text{\AA}^{-2} \text{ frame}^{-1}$ for a final dose of 34.5 $\text{e}^{-} \text{\AA}^{-2}$ on the sample. Automated data collection was driven by the Leginon automation software package⁵⁰. 2935 micrographs were collected with a defocus range from 0.8 to 5.1 μm .

Micrographs summing all frames were all corrected for gain reference and then micrographs with a group of three frames were used to estimate the contrast transfer function using CTFFIND 4.1.851. Unblur⁵² was then used to correct beam-induced motion with dose weighting and anisotropic magnification correction, leading to a physical pixel size of 1.064 \AA pixel^{-1} .

Micrographs with crystalline ice, severe astigmatism, or obvious drift were discarded. All subsequent data processing were performed using RELION 2.0^{53,54}.

All filaments were picked manually using EMAN2 e2helixboxer.py⁵⁵. Particles were first extracted using a box size of 1024 and 686 pixels with an inter-box distance of 10% of the box

length. 2D classification using 1024 pixel particles was used to estimate the fibril pitch and helical parameters. We also performed 2D classifications with 686 pixel particles to select particles for future 3D classification. We performed Class3D jobs with three classes and manually controlled the tau_fudge factor and healpix_order to obtain resolutions around 6–7 Å, using an elongated Gaussian blob as an initial reference. We selected particles that contribute to the highest resolution class and generated an initial 3D reconstruction by running Class3D with 1 class. To obtain a higher resolution reconstruction, we re-extracted particles with a box size of 224 pixels from the fibril tubes containing the Class3D-selected 686 pixel particles. All 224 pixel particles were used directly for two rounds of 3D classification. The final subset of selected particles was used for high-resolution gold-standard refinement as described previously⁵⁴. The final overall resolution estimate was evaluated to be 3.1 Å based on the 0.143 FSC cutoff⁵⁶. Projections from the final reconstruction closely match the 2D class averages, helping to validate the reconstruction.

The refined map was sharpened using phenix.auto_sharpen at the resolution cutoff⁵⁷, and a near-atomic resolution model was built de novo into the sharpened map using COOT⁵⁸. We generated a five-layer model using the helical parameters from the final 3D refinement and then refined the structure using phenix.real_space_refine⁵⁹. After the last round of refinement, we adjusted the residue's phi/psi angles to facilitate main chain hydrogen-bonding, and the final model was validated using phenix.comprehensive_validation^{60·61}. All the statistics are summarized in Table [1](#).

Solvation energy calculation

The solvation energy was calculated based on work published previously⁶². The solvation energy for each residue was calculated by the sum of the products of the area buried for each

atom and the corresponding atomic solvation parameters. The overall energy was calculated by the sum of energies of all residues. Different colors were assigned to each residue, instead of each atom, in the solvation energy map. The energy reported for FUS in Table [2](#) is the average over 20 NMR models.

3D environment compatibility search

3D environment profiling is performed based on the method described previously³⁰. Briefly, an environment category was assigned for the side chains of each residue, and a profile was created using the model structure and its homologous structures based on three features: (1) the secondary structure the residue is in, (2) the fraction of side chain that is covered by polar atoms, and (3) the area buried. A 3D environment profile was created using our mCherry-hnRNPA2-LCD fibril core structure.

Peptides stabilization energy calculation

The hnRNPA2 fibril structure is divided into β -sheets composed of six-residue long contiguous segments. Then the surface area buried between all pairwise combinations of these sheets is measured. If the area buried in the dry interface is $>80 \text{ \AA}^2$, the pair of segments is kept as a potential zipper. If the area is $<80 \text{ \AA}^2$, the pair of segments is discarded due to a lack of enough overlap. The energy of the overlapping pairwise segments is measured using our solvation energy calculation described above. The histogram summarizes the energies of 1657 potential zippers extracted from 12 PDB entries of 5 disease-related amyloid protein: amyloid- β (PDB: 5KK3), α -synuclein (PDB: 2N0A, 6CU7, 6CU8), Serum amyloid A (PDB:6DSO), tau (PDB: 5O3L, 5O3T, 6GX5), and TDP-43 (PDB: 6N37, 6N3A, 6N3B, 6N3C).

Peptide aggregation formation

Both wildtype and mutant peptide segments are dissolved in water to a final concentration of 10 mg ml⁻¹ in 1.5 ml silicon tubes with parafilm seal. Samples were incubated at 25 °C in a Torrey Pine Scientific shaker at level 9 for 90 h and then examined by TEM.

Segment crystallization

Crystals of the mutant peptide segment were grown by hanging-drop vapor diffusion. The segment was dissolved in water to a final concentration of 9.69 mg ml⁻¹ with 31 mM lithium hydroxide. The reservoir solution contains 0.15 M ammonium acetate, 35% MPD, and 0.1 M Bis-Tris pH 5.5, which is optimized based on the original hit in the commercial kit JCSG 96-well block platescreen from Molecular Dimensions, well H5. Hanging drop contains a 1:1 solution of protein:reservoir.

X-ray data collection, processing, and structural analysis

X-ray diffraction data from GNYNVF segment crystals were collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA, at a wavelength of 0.979 Å and temperature of 100 K. Data were collected using 2.5° oscillations and 150 mm detector distance with an EIGER detector. Indexing and integration of the reflections were done using XDS in space group P2₁ and scaled with XSCALE to a resolution of 1.1 Å⁶³. The structure was solved by direct method using SHELXD⁶⁴. The atomic-resolution model was manually built and adjusted using COOT⁵⁸. The model is then refined by Refmac⁶⁵ with a final $R_{\text{work}}/R_{\text{free}}$ of 7.5/10.7(%) and 100% of Ramachandran angles favored. The crystal structure image was generated using Pymol. Area buried (Ab) and shape complementarity (Sc) were calculated based on work published previously^{66,67}. All the statistics are summarized in Table [3](#).

Data availability

Atomic coordinates and cryoEM map of the hnRNPA2 LCD fibril were deposited in the Protein Data Bank and Electron Microscopy Data Bank with entry codes: [6WQK](#) and EMD-21871. The crystal structure GNYNVF was deposited in the Protein Data Bank with entry code: [6WPQ](#).

Other data are available from the corresponding author upon reasonable request.

Acknowledgements

We thank Peng Ge of the Electron Imaging Center for NanoMachines (EICN) for help with cryoEM data collection and processing; We are grateful to Dr. Masato Kato and Dr. Steven L. McKnight from University of Texas, Southwestern Medical Center for sending us the construct of mCherry-hnRNPA2-LCD and their advice on sample preparation and we thank the Whitcome Pre-doctoral Training Program, UCLA Molecular Biology Institute for funding, NSF MCB 1616265, and NIH RF1 AG054022 for support. This work used NE-CAT beamline 24-ID-E (GM124165) and an Eiger detector (OD021527) at the APS (DE-AC02-06CH11357).

Author Contributions

J.L. designed experiments, purified constructs, crystallized peptides, prepared cryo-EM samples, performed X-ray and cryo-EM data collection and processing, and performed data analysis. Q.C. performed cryoEM data processing. D.R.B. assisted in cryoEM data collection and processing. M.P.H. performed LARKS prediction and assisted in model building. M.R.S. performed solvation energy calculation. M.R.S. and D.C. assisted in X-ray data collection and processing and model building. All authors analyzed the results and wrote the manuscript. D.S.E. supervised and guided the project.

Competing interests

D.S.E. is an advisor and equity shareholder in ADRx, Inc. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-17905-y>.

Correspondence and requests for materials should be addressed to D.S.E.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Figures

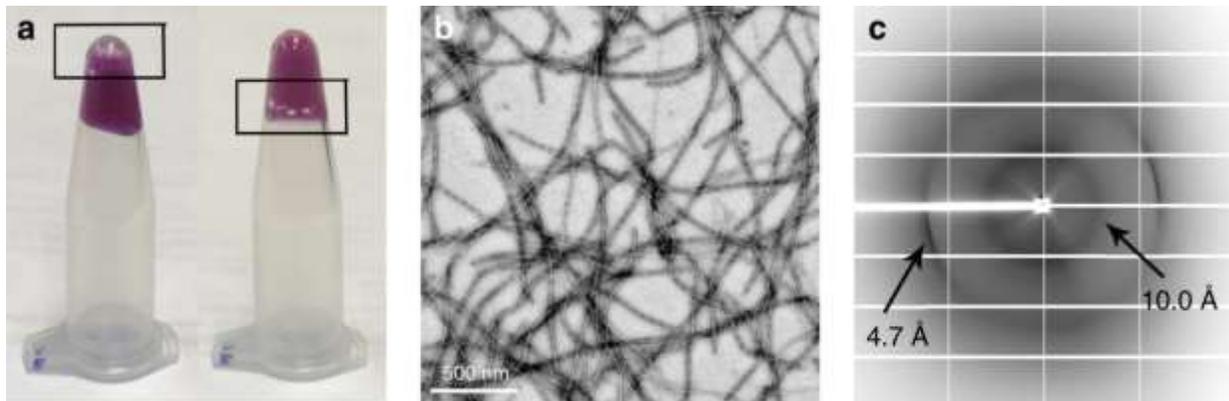


Fig. 1 mC-hnRNPA2-LCD forms a reversible hydrogel that contains amyloid-like fibrils displaying the cross- β diffraction pattern.

a Left microcentrifuge tube holds a homogenous solution of freshly purified mC-hnRNPA2-LCD. A bubble (box) rises to the top of the tube. The right tube contains the hydrogel formed by concentrated mC-hnRNPA2-LCD; the bubble is trapped by the hydrogel. **b** Hydrogel formed by mC-hnRNPA2-LCD visualized by transmission electron microscopy shows uniform, amyloid-like fibrils. Scale bar: 500 nm. **c** X-ray diffraction pattern of the hydrogel formed by mChnRNPA2-LCD. The two reflections characteristic of amyloid are visible at 4.7 and 10.0 Å spacings.

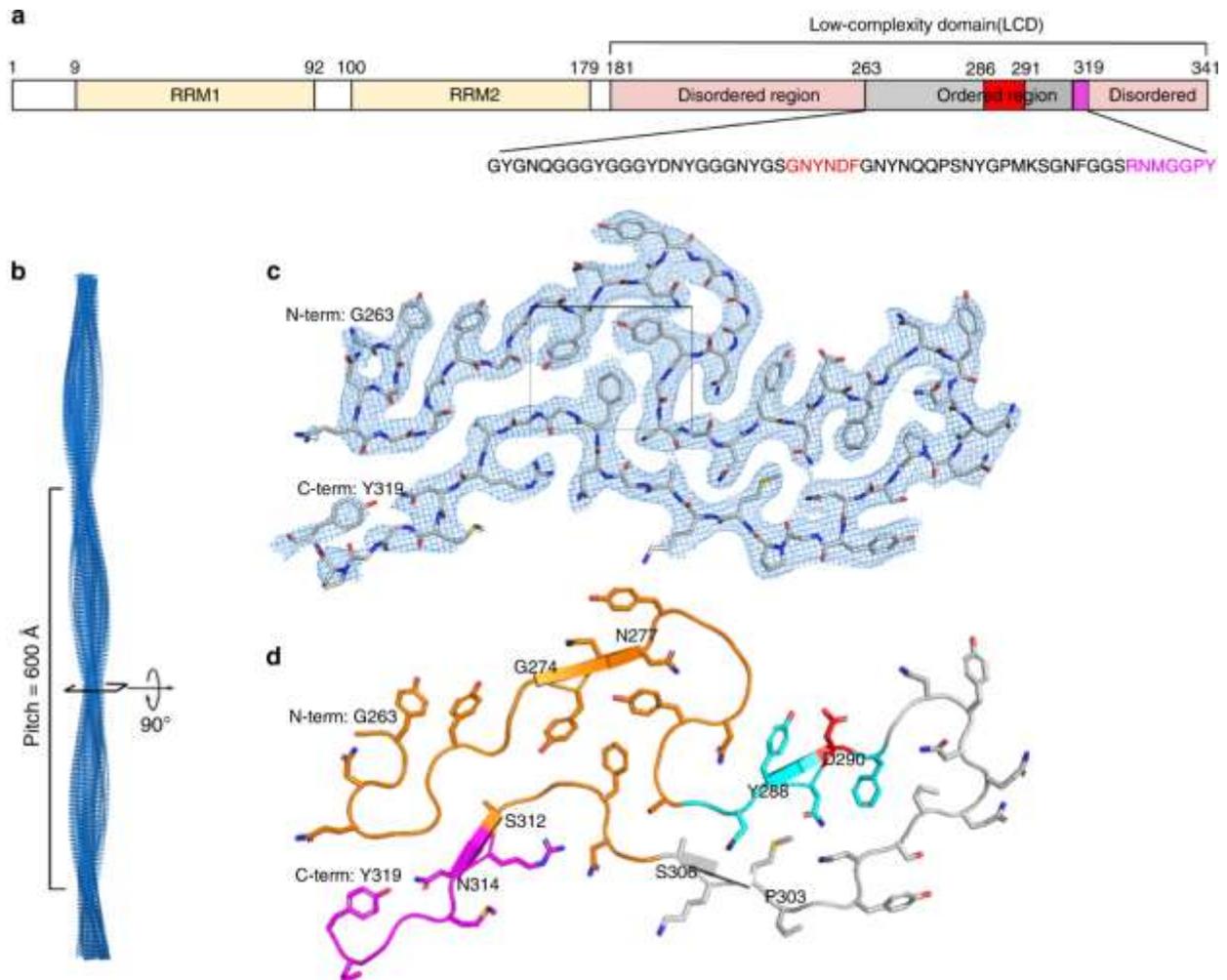


Fig. 2 The Cryo-EM structure of the mC-hnRNP A2-LCD fibril core.

a Domain structure of full-length hnRNP A2. The LCD (residues 181–341) is identified for structural determination. The gray bar shows the range of the ordered fibril core of the cryoEM structure. The red bar shows the core segment (crystal structure described below) containing the site of a disease-causing mutation. The magenta bar shows the nuclear localization signal, PY-NLS. The sequence of the ordered region is shown below with corresponding colors. **b** The mC-hnRNP A2-LCD fibril reconstruction, showing its left-handed twist and pitch. **c** Density and atomic model of one cross-sectional layer of the fibril. The box shows the aromatic triad. **d** Atomic model of one cross-sectional layer of the fibril. The predicted LARKS domain (contains 7 LARKS motifs) (above) and the shorter LARKS motif (below) are colored orange; the core segment is colored cyan, with its disease-causing mutation site colored red; the nuclear localization signal is colored magenta (lower left). β -sheet-forming residues are G274-N277, Y288-D290, P303-S306, S312-N314, indicated by arrows.

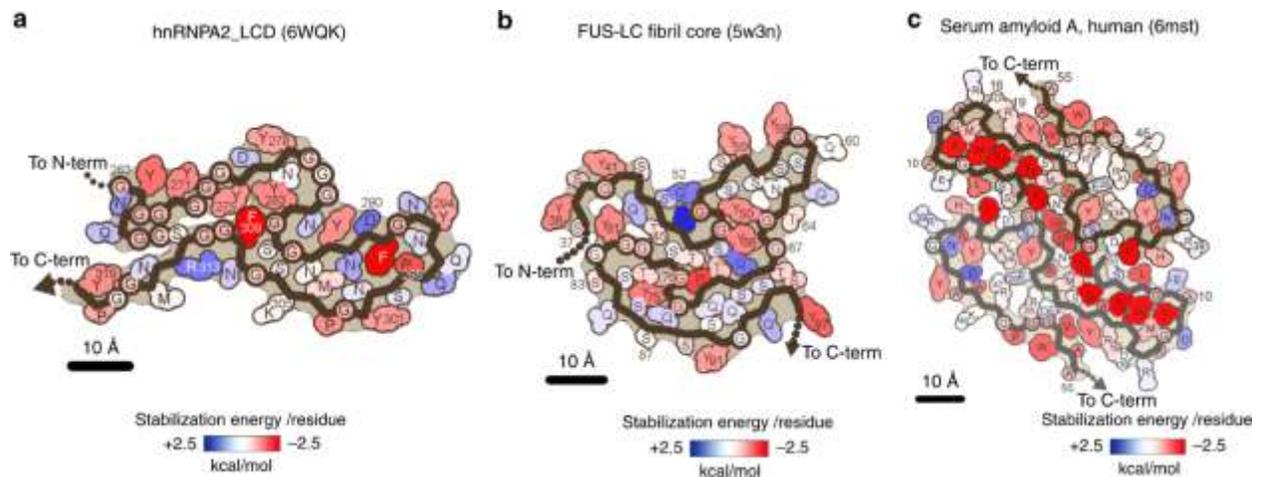


Fig. 3 Solvation energy maps of mC-hnRNPA2-LCD, FUS-LCD, and serum Amyloid A. Residues are colored according to their stabilization energies from unfavorable (blue, +2.5 kcal mol⁻¹) to favorable (red, -2.5 kcal mol⁻¹). **a** Solvation energy map of mC-hnRNPA2-LCD ordered fibril core. **b** Solvation energy map of FUS-LCD¹⁴ ordered region. **c** Solvation energy map of human serum amyloid A¹⁹, ordered region.

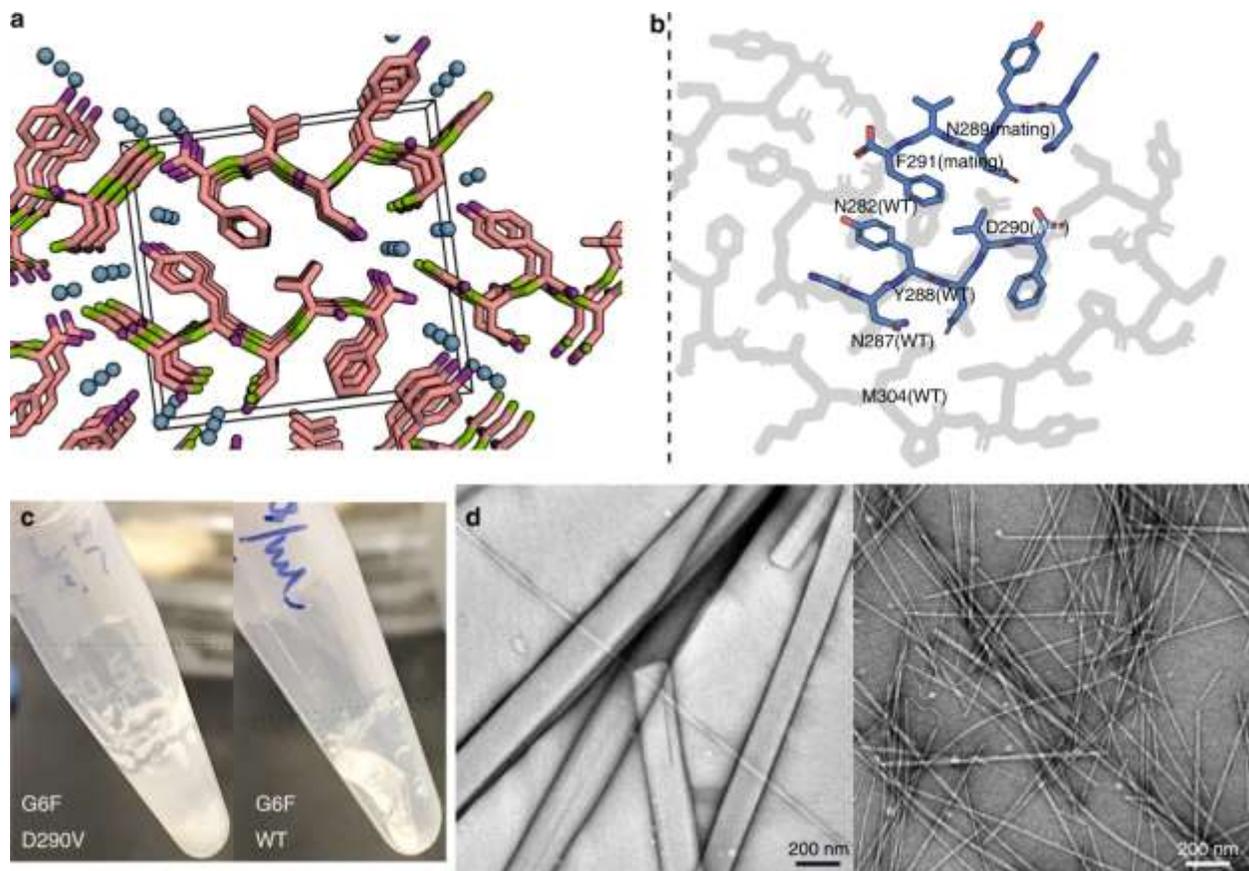


Fig. 4 Mutant core segment and the corresponding wildtype segment show distinct features. **a** The steric zipper structure determined for mutant segment GNYNVE, viewed down the fibril axis. Water molecules within the structures are shown as aqua spheres. The unit cell is shown as

a box. **b** A superimposition of the mutant steric zipper structure of GNYNVF (blue) on the cryoEM mC-hnRNPA2-LCD structure (gray) shows the incompatibility of the mutant structure with the fold of the wildtype fibril. Only the corresponding part of the wildtype structure is shown. The sidechains of Asn287, Tyr288, and Asn289 of the mutant structure experience steric clashes with the sidechains of Asn282, Asp290, and Met304 of the wildtype structure. **c** Left: Mutant segment GNYNVF shaken after 4 days forms a white precipitate. Right: Wildtype segment GNYNDF shaken after 4 days forms a hydrogel. **d** Left: The mutant segment GNYNVF (from **c**) visualized by transmission electron microscopy, shows wide needle crystals. Scale bar: 200 nm; Right: Wildtype segment GNYNDF (from **c**) visualized by transmission electron microscopy, shows thin fibrils.

Tables

Table 1 Statistics of cryoEM data collection, refinement, and validation.	
	HnRNPA2-LCD (EMD-21871) (PDB: 6WQK)
<i>Data collection and processing</i>	
Magnification	×130,000
Voltage (kV)	300
Electron exposure (e ⁻ Å ⁻²)	34.5
Defocus range (μm)	0.8-5.1
Pixel size (Å)	1.064
Symmetry imposed	C1; Helical
Initial particle images (no.)	529,821
Final particle images (no.)	132,571
Map resolution (Å)	3.1
FSC threshold	0.143
Map resolution range (Å)	200-3.1
<i>Refinement</i>	
Initial model used	De Novo
Model resolution (Å)	3.2
FSC threshold	0.5
Model resolution range (Å)	200-3.2
Map sharpening B factor (Å ²)	-120
<i>Model composition</i>	
Non-hydrogen atoms	2070
Protein residues	285
<i>B factors (Å²)</i>	
Protein	55.3
<i>R.m.s. deviations</i>	
Bond lengths (Å)	0.006
Bond angles (°)	0.89
<i>Validation</i>	
MolProbity score	1.70
Clashscore	5.59
Poor rotamers (%)	0.00
<i>Ramachandran plot</i>	
Favored (%)	94
Allowed (%)	6
Disallowed (%)	0.00

Table 2 Comparative values of solvation standard free energies of functional and pathogenic fibrils.

	Ordered residues	Method	Resolution (Å)	Energy/chain (kcal mol ⁻¹)	Energy/residue (kcal mol ⁻¹)
hnRNPA2-LCD (functional)	57	CryoEM	3.1	-19.5	-0.34
FUS-LCD (functional)	61	ssNMR	2.7	-12.2	-0.20
Serum amyloid A Human (pathogenic)	54	CryoEM	2.7	-34.4	-0.64
Tau PHFs AD patient (pathogenic)	73	CryoEM	3.4	-28.9	-0.40
Aβ42 (pathogenic)	42	CryoEM	4.0	-24.8	-0.59

Table 3 Statistics of X-ray crystallography data collection and refinement.

	GNYNVF PDB: 6WPQ
<i>Data collection</i>	
Space group	P2 ₁
<i>Cell dimensions</i>	
a, b, c (Å)	4.78, 19.00, 20.74
α, β, γ (°)	90.00, 95.71, 90.00
Resolution (Å)	1.1
R _{merge} (%)	11.6 (21.8)
I / σI	10.4 (5.7)
Completeness (%)	86.5 (41.7)
Redundancy	5.6 (3.5)
<i>Refinement</i>	
Resolution (Å)	20.64-1.10
No. of reflections	1196
R _{work} /R _{free} (%)	7.5/10.7
<i>No. of atoms</i>	
Protein	51
Water	3
<i>B-factors (Å²)</i>	
Protein	4.8
Water	7.4
<i>R.m.s. deviations</i>	
Bond lengths (Å)	0.014
Bond angles (°)	1.804

Table 4 Comparative properties of functional amyloid-like fibrils and pathogenic amyloid fibrils.

	Functional amyloid-like fibrils	Pathogenic amyloid fibrils
Variety	Monomorphic	Tend to be polymorphic
Stability	Tend to be labile and reversible	Tend to be stable and irreversible
ΔG° of stabilization	- -0.2 kcal mol ⁻¹ of residue ⁻¹	- -0.4 kcal mol ⁻¹ of residue ⁻¹
Protein chains stack into β -sheets by backbone H-bonds	Yes	Yes
Sheets pair	Yes, by LARKS	Yes, by steric zippers
Residues that drive sheet pairing	Often Asn, Gln, Tyr, Phe, Ser, Gly, Pro in low-complexity domains	Varied, but frequently alternating polar and apolar especially Val, Ile, Val
Fiber diffraction pattern	Cross- β	Cross- β
Morphology	Crosslinked networks of fibrils	Non-crosslinked deposits of fibrils

Table 5 Comparative sequence composition of functional and pathogenic amyloid fibril cores.

	Gly (%)	Asn+ Gln (%)	Tyr+Ser+ Thr (%)	Val+Ala+ Ile+Leu (%)
HnRNPA2-LCD (functional)	35.1	22.8	22.8	0.0
FUS-LCD (functional)	19.7	18.0	59.1	0.0
Tau PHFs (pathogenic)	12.3	8.2	15.1	24.4
A β 42 (pathogenic)	14.3	4.8	7.2	35.7

Table 6 Comparative 3D-1D environmental profiles of hnRNPA2-LCD with pathogenic amyloids.

Protein	Exposed to solvent (%)	Partially buried with high fractional environmental polarity (%)	Nonpolar buried (%)
hnRNPA2-LCD	35	32	4
Tau PHFs	12	30	13
A β 42	14	24	18

References

1. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767 (2012).
2. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 495, 467–473 (2013).
3. Amaya, J., Ryan, V. H. & Fawzi, N. L. The SH3 domain of Fyn kinase interacts with and induces liquid–liquid phase separation of the low-complexity domain of hnRNPA2. *J. Biol. Chem.* 293, 19522–19531 (2018).
4. White, R. et al. Activation of oligodendroglial Fyn kinase enhances translation of mRNAs transported in hnRNP A2-dependent RNA granules. *J. Cell Biol.* 181, 579–586 (2008).
5. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* 294, 7115–7127 (2019).
6. Xiang, S. et al. The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets and nuclei. *Cell* 163, 829–839 (2015).
7. Wang, J. et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* 174, 688–699.e16 (2018).
8. Ryan, V. H. et al. Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. *Mol. Cell* 69, 465–479.e7 (2018).
9. Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* 162, 1066–1077 (2015).

10. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase separated liquid droplets by RNA binding proteins. *Mol. Cell* 60, 208–219 (2015).
11. Mackenzie, I. R. et al. TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* 95, 808–816.e9 (2017).
12. Murakami, T. et al. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* 88, 678–690 (2015).
13. Gui, X. et al. Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* 10, 1–12 (2019).
14. Murray, D. T. et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* 171, 615–627. e16 (2017).
15. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat. Struct. Mol. Biol.* 26, 619–627 (2019).
16. Guenther, E. L. et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* 25, 463–471 (2018).
17. Eisenberg, D. S. & Sawaya, M. R. Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* 86, 69–95 (2017).
18. Zhang, Z. C. & Chook, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline–tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl Acad. Sci. USA* 109, 12017–12021 (2012).

19. Liberta, F. et al. Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nat. Commun.* 10, 1104 (2019).
20. Kollmer, M., Close, W., Funk, L. et al. Cryo-EM structure and polymorphism of A β amyloid fibrils purified from Alzheimer's brain tissue. *Nat Commun* 10, 4760 (2019).
21. Luo, F. et al. Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nat. Struct. Mol. Biol.* 25, 341–346 (2018).
22. Hughes, M. P. et al. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* 359, 698–701 (2018).
23. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* 107, 3487–3492 (2010).
24. Sawaya, M. R. et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447, 453–457 (2007).
25. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* 547, 185–190 (2017).
26. Gremer, L. et al. Fibril structure of amyloid- β (1-42) by cryo-electron microscopy. *Science* 358, 116–119 (2017).
27. Murray, D. T. et al. Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. *Proc. Natl Acad. Sci. USA* 115, E9782–E9791 (2018).
28. Semenov, A. N. & Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* 31, 1373–1385 (1998).

29. Rubinstein, M. & Semenov, A. N. Thermoreversible gelation in solutions of associating polymers. 2. Linear dynamics. *Macromolecules* 31, 1386–1397 (1998).
30. Bowie, J. U., Lüthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170 (1991).
31. Nelson, R. et al. Structure of the cross- β spine of amyloid-like fibrils. *Nature* 435, 773–778 (2005).
32. Ding, X. et al. Amyloid-forming segment induces aggregation of FUS-LC domain from phase separation modulated by site-specific phosphorylation. *J. Mol. Biol.* 432, 467–483 (2020).
33. Falcon, B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* 561, 137–140 (2018).
34. Colvin, M. T. et al. Atomic resolution structure of monomorphic A β 42 amyloid fibrils. *J. Am. Chem. Soc.* 138, 9663–9674 (2016).
35. Lührs, T. et al. 3D structure of Alzheimer's amyloid-beta(1–42) fibrils. *Proc. Natl Acad. Sci. USA* 102, 17342–17347 (2005).
36. Schmidt, M. et al. Peptide dimer structure in an A β (1–42) fibril visualized with cryo-EM. *Proc. Natl Acad. Sci. USA* 112, 11858–11863 (2015).
37. Wälti, M. A. et al. Atomic-resolution structure of a disease-relevant A β (1–42) amyloid fibril. *Proc. Natl Acad. Sci. USA* 113, E4976–E4984 (2016).
38. Xiao, Y. et al. A β (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* 22, 499–505 (2015).

39. Li, B. et al. Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* 9, 1–10 (2018).
40. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* 181, 223–230 (1973).
41. Guo, L. et al. Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell* 173, 677–692.e20 (2018).
42. Lee, B. J. et al. Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* 126, 543–558 (2006).
43. Shu, X., Shaner, N. C., Yarbrough, C. A., Tsien, R. Y. & Remington, S. J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45, 9639–9647 (2006).
44. Kaniyappan, S. et al. FRET-based Tau seeding assay does not represent prion-like templated assembly of Tau filaments. *Mol. Neurodegener.* 15, 39 (2020).
45. Van Treeck, B. & Parker, R. Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell* 174, 791–802 (2018).
46. Protter, D. S. W. et al. Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. *Cell Rep.* 22, 1401–1412 (2018).
47. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* 294, 7128–7136 (2019).
48. Rodriguez, J. A. et al. Structure of the toxic core of α -synuclein from invisible crystals. *Nature* 525, 486–490 (2015).

49. Arvai, A. Adxv. Adxv—A Program to Display X-Ray diffraction Images <https://www.scripps.edu/tainer/arvai/adxv.html> (2015).
50. Suloway, C. et al. Automated molecular microscopy: the new Legimon system. *J. Struct. Biol.* 151, 41–60 (2005).
51. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221 (2015).
52. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* 4, e06980 (2015).
53. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* 198, 163–176 (2017).
54. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530 (2012).
55. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* 157, 38–46 (2007).
56. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryo-microscopy. *Ultramicroscopy* 135, 24–35 (2013).
57. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. Struct. Biol.* 74, 545–559 (2018).

58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* 66, 486–501 (2010).
59. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* 74, 531–544 (2018).
60. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66, 213–221 (2010).
61. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21 (2010).
62. Eisenberg, D., Wesson, M. & Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.*
<https://eurekamag.com/research/032/001/032001963.php> (1989).
63. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* 66, 133–144 (2010).
64. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* 64, 112–122 (2008).
65. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* 53, 240–255 (1997).
66. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946–950 (1993).
67. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760–763 (1994).

Chapter 2

Cryo-EM structures of the D290V mutant of the hnRNPA2 low-complexity domain mask the nuclear import signal required for function

Abstract

hnRNPA2 is a human ribonucleoprotein that transports RNA to designated locations for translation via its ability to phase separate *in vivo*. Its mutated form, D290V, is implicated in MultiSystem Proteinopathy known to afflict two families, mainly with myopathy and Paget's Disease of Bone. Here, we investigate this mutant form of hnRNPA2 by determining cryo-EM structures of the recombinant D290V low complexity domain (LCD). We find that in contrast to the wildtype fibrils, the PY-nuclear localization signals are masked in the fibril cores of all three mutant polymorphs. Also, the mutant fibrils are more stable than wildtype fibrils as judged by phase separation, thermal stability, and energetic calculations. Thus, these structures offer evidence to explain how a D-to-V missense mutation diverts the assembly of reversible, functional amyloid-like fibrils into the assembly of pathogenic amyloid, and may shed light on analogous conversions occurring in other ribonucleoproteins that lead to neurological diseases such as ALS and FTD.

Teaser

Structures of a disease-associated variant of the RNA-binding protein hnRNPA2 reveal partial burial of its nuclear localization signal.

MAIN TEXT

Introduction

MultiSystem Proteinopathy (MSP) consists of a related group of dominantly inherited degenerative disorders including cognitive impairment, myopathy, Paget's Disease of Bone and motor neuron disease¹. Mutations leading to MSP were first recognized in the valosin-containing protein gene (VCP) in 2004². Then in 2013, MSP patients were found bearing missense

mutations in the LCD of heterogeneous nuclear ribonucleoproteins A2 and A1 (hnRNPA2, D290V; hnRNPA1, D262V) and no VCP mutations³. In two MSP-affected families, muscle biopsies showed that hnRNPA2 was absent from nuclei where it normally resides and instead accumulated as fibrils in cytoplasmic inclusions³. This nuclear clearance and cytoplasmic inclusion phenomenon are also observed for ribonucleoproteins FUS⁴ and TDP-43^{5,6}.

The human heterogeneous nuclear ribonucleoprotein A2 (hnRNPA2) belongs to a family of hnRNPs and is expressed in cells such as neurons, oligodendrocytes, and myocytes^{3,7}. hnRNPA2 functions in maintaining RNA metabolism through two functionally distinct domains, the RNA-binding domain and the LCD. hnRNPA2 can bind a variety of RNA/DNA sequences⁸, such as those encoding the A2 response element of the myelin basic protein⁹ and transport it to targeted locations for translation by forming membraneless organelles (MLOs). hnRNPA2 was studied in RNA transport granules in oligodendrocytes⁹ and stress granules in neurons³. Its LCD was also found to undergo phase separation to form liquid droplets similar to cellular MLOs^{10,11}. Studies of protein phase separation, including with FUS, TDP-43, and TIA1, have found that MLOs constantly exchange contents with the cellular environment and are transported between organelles or cells^{4,12-17}.

The LCD of hnRNPA2 can also self-assemble into a network of amyloid-like reversible fibrils that macroscopically manifests as a hydrogel. Our previous study showed that the LCD of hnRNPA2 is rich in structural motifs termed LARKS (Low-complexity, Amyloid-like, Reversible, Kinked Segments) that enable amyloid-like assemblies having lower stability than pathogenic amyloid^{18,19}. LARKS contribute to the underlying transient interactions in hydrogels¹⁹. Using footprinting, Xiang et al. showed that the LCD of hnRNPA2 adopts the cross-

β scaffold of LARKS in liquid droplets and in hydrogels²⁰. However *in vivo*, the relationship of phase separation to hydrogels remains unclear.

So far, over 60 mutations have been identified within the LCDs of ribonucleoproteins that are associated with neurodegenerative diseases such as ALS. Understanding the structure-function relationships of mutated LCDs can better assist in understanding the diseases. Identifying the disease-causing mutation D290V of hnRNPA2, Kim et al., used ZipperDB²¹ to argue that the mutant LCD is more prone to aggregation than the wildtype LCD³. In fact, they showed that the hexapeptide containing this mutated residue (NYNVFG) forms fibrils while the wildtype sequence (NYNDFG) does not. They also showed that recruitment of hnRNPA2 to stress granules was enhanced by the D290V mutation³. Increased formation of cytoplasmic inclusions of the mutant hnRNPA2 was observed in mouse and *Drosophila* models and in human patients³. Thus, structural and biochemical comparison of the wildtype to the mutant D290V hnRNPA2 LCD is likely to suggest how a missense mutation diverts function to pathology.

Here, we illustrate how a missense mutation diverts hnRNPA2 LCD from forming functional fibrils to forming pathogenic fibrils, and we propose a mechanism to explain observations of the mutant's clearance from the nucleus and formation of cytoplasmic inclusions. We support our mechanism with evidence from phase separation studies, Thioflavin-T (ThT)-monitored kinetics of fibril formation, assays of thermostability, negatively-stained EM images, and three cryo-EM structures of the mutant D290V hnRNPA2 LCD fibrils.

Results

hnRNPA2 LCD undergoes phase separation, and the D290V mutant forms branched aggregates faster than wildtype hnRNPA2-LCD

Previous studies have shown that wildtype hnRNPA2 LCD can undergo phase separation^{10,11}. Here, we investigate whether the mutant can also undergo phase separation and whether its biophysical behavior differs from wildtype, using recombinant hnRNPA2 LCDs. Due to its high aggregation propensity, we initially dissolved the hnRNPA2 LCD in 5.7 M guanidinium hydrochloride (GnCl). Both the wildtype and the mutant rapidly underwent phase separation upon dilution to a final concentration of 100uM GnCl. The mutant liquid droplets (mean diameter: ~1.8um) are half the diameter of wildtype droplets (mean. diameter: ~3.4um) [Figure.1a]. We then performed “aging” experiments in which the behaviors of both constructs were monitored over time. The wildtype liquid droplets stayed dynamic for 24 hours during which they settled to the bottom of the plate and fused into larger droplets [Figure.1a, top], and then hardened into a hydrogel-like state after 48 hours. In contrast the mutant formed branched aggregates that look like fibrils within 3 hours [Figure.1a, bottom; Movie.S1]. We then recorded negatively stained EM of different time points of both wildtype and the mutant phase separation. For the D290V mutant, in the first hour, only small droplets could be seen [Figure.1b, bottom]. For the following 1-1.5 hours, branched aggregates emanated from the droplets and connected with nearby droplets, making a dense fibrous network [Figure.1b, bottom]. No branched aggregates were observed for wildtype even after 48 hours [Figure.1b, top]. Note, we use the term “branched aggregates” to distinguish this species from “fibrils” used for cryoEM; they are prepared differently and their morphologies are different (“Branched aggregates” are formed as mentioned above. “Fibrils” are formed by directly dissolving LCD in buffers and shaking).

Branched aggregates emergent from the mutant liquid phase are partially irreversible and ThT-positive

To investigate the structural properties of the phase separation of both constructs, immediately after phase separation we added ThT, a fluorescent amyloid dye that changes conformation and fluoresces at wavelength 482 nm when it binds to cross- β structures²². Both constructs underwent phase separation rapidly, indicating the ThT does not affect their ability to phase separate. The mutant liquid droplets showed ThT signal immediately after forming, the signal grew stronger with time, and after 4 hours, a strong ThT signal was seen, and in some places in the droplets, branched aggregates grew so large that the fibril-like outline can be recognized in the fluorescence channel [Figure.2b]. Unlike the mutant, the wildtype did not show ThT signal at 10 minutes after liquid droplet formation. After 4 hours, ThT signal was seen in the wildtype liquid droplets [Figure.2a]. The phase separation behavior we observed in the first 10 minutes indicates that the mutant hnRNPA2 LCD forms cross- β structure in the liquid droplets faster than wildtype.

We then evaluated the reversibility of the 24-hour wildtype liquid droplets and the mutant branched aggregates by performing a thermostability assay. The wildtype and mutant samples were heated to 55°C, the temperature at which wildtype hnRNPA2 hydrogel melts¹⁸, and subsequently cooled to 4°C. When heated, the mutant branched aggregates were unchanged until 55°C when some of the aggregates melted and reformed small liquid droplets [Fig. S1, top]. In contrast, the hardened wildtype liquid droplets were completely melted at 55°C and dynamic small droplets were observed [Fig. S1, bottom]. When cooled to 4°C, the mutant reformed branched aggregates that were undisturbed by pipetting, whereas the wildtype liquid droplets lost their hydrogel-like characteristics and maintained liquid properties in line with its behavior stated in the earlier section [Fig. S1]. Namely, “de-aged” wildtype liquid droplets re-gain normal liquid properties and harden after a prolonged time.

Phase separation of the mutant hnRNPA2 is less sensitive to pH changes than wildtype

To test whether phase separation of either construct is sensitive to pH, we conducted phase separation experiments at six different pH values ranging from 1.3 to 10.0. Liquid droplets of both constructs with the addition of ThT were incubated at room temperature for 22 hours.

Wildtype hnRNPA2 aggregation morphology varied notably with pH. Values near physiological pH (5.5, 7.5 and 9.0) produced droplet-shaped gels, exhibiting a strong ThT fluorescence signal under the microscope [Fig. S2a]. However, at extreme acidic pH values such as pH 1.3 and pH 4.1, wildtype aggregates appear as branched spikes (perhaps fibrous) protruding from an amorphous center; the centers and the branched spikes are both ThT-positive. In contrast to these branched aggregates the droplets formed at pH 10 appeared smooth and round like a liquid, but their resistance to being stirred and mixed indicated that they are actually gel-like solids [Fig. S2a]. These alkaline aggregates, like the aggregates at lower pH values, exhibited intense ThT signal, suggesting cross- β structure is present at all pH values despite the different morphologies of the aggregates.

The mutant's aggregation morphology appeared more uniform across pH values than did wildtype. The mutant formed branched, ThT-positive aggregates at all pH values sampled [Fig. S2b], not just at low pH values as did the wildtype [Fig. S2a]. Some differences in robustness of aggregation are evident when comparing the mutant aggregation at low and high pH values; droplets of the mutant fused and aggregated faster at acidic pH than basic pH [Fig. S2b]. Nevertheless, the conservation of branched morphology across all pH values indicates that aggregation of the mutant hnRNPA2 LCD is less sensitive to pH changes than wildtype.

Fibrils of the mutant grow faster and are more stable than fibrils of wildtype hnRNPA2 LCD

We compared fibril growth kinetics and thermodynamics of wildtype and the mutant hnRNPA2 LCD produced recombinantly. The LCDs of wildtype and mutant were individually dissolved in buffer containing 20 mM MOPS, pH 7.5, 150 mM KCl buffer and 2 M GnCl to a final concentration of 4-5 mg ml⁻¹ and was shaken at 4 °C for 3 days. ThT assays showed that the mutant formed fibrils after 20 minutes whereas wildtype formed fibrils after 25 hours [Figure.3a]. Wildtype hnRNPA2 LCD fibrils remain intact up to 45°C, but completely melt at temperatures 55°C and above [Fig. S3], consistent with our previously reported results¹⁸. The mutant hnRNPA2 LCD fibrils, however, remain intact until 55°C, and do not melt until 65°C [Figure.3b]. Upon cooling to 4°C, both the mutant and the wildtype hnRNPA2 LCDs re-form fibrils [Figure.3b and Figure.S3].

CryoEM reveals the PY-NLS is buried in all three fibril polymorphs of the mutant hnRNPA2-LCD

Atomic-resolution structures of amyloid fibrils of both mutant and wildtype hnRNPA2 LCDs are essential for understanding why the mutant accumulates in the cytoplasm, not the wildtype. The fibril structure of wildtype hnRNPA2-LCD has been determined and described previously¹⁸. Here, we collected cryo-EM data on fibrils of the D290V mutant. During data collection and 3D classification, we found a total of six polymorphs, yet only three of them (PM1-3) had a sufficient number of particles to be determined to near-atomic resolution [Fig. S4]. Contrary to globular proteins which fold into 3D shapes, molecules of the mutant hnRNPA2 LCD are confined to 2D layers in each of the three fibril polymorphs. These layers stack on top of

identical layers, forming a twisted, in-register β -sheet that runs parallel to the fibril axis. The details of the 3 polymorphic fibrils are as follows:

1) PM1 consists of two protofilaments with pseudo- 2_1 screw symmetry, and is determined to a resolution of 3.3Å. The fibril is a left-handed helix with a pitch of 2335Å, giving rise to a helical twist of 179.63° and a rise of 2.40Å. Out of 161 residues in the LCD, 54 residues, from Gly263 to Gly316, form the fibril core. The fibril core contains the disease-causing D290V mutation which is solvent exposed, and half of the PY-nuclear localization signal (PY-NLS). The PY-NLS is a 19-residue segment extending from Tyr301 to Tyr319²³; 16 of the 19 residues (Tyr301 to Gly316) are ordered in the mutant structures. The remaining three residues of the PY-NLS and 107 LCD residues are disordered in the “fuzzy coat”. The interface between the two protofilaments spans residues Tyr294 to Gly311. The rest of the C-terminal residues are the PY-NLS [Figure.4b].

2) PM2 is structurally similar to PM1. It also consists of two protofilaments with pseudo- 2_1 screw symmetry and is determined to a resolution of 3.2Å. It is a left-handed fibril with a pitch of 2151Å, giving rise to a helical twist of 179.60° and a rise of 2.39Å. The same 54 residues from Gly263 to Gly316 form the fibril core, with the disease-causing D290V mutation exposed to the solvent on the outer surface. The core adopts the same fold as PM1, but PM2 has a smaller and more interdigitated interface between the two protofilaments involving residues Tyr294 to Pro303 [Figure.4c].

3) PM3 has C1 symmetry and is determined to a resolution of 3.9Å. PM3 is a left-handed fibril consisting of three protofilaments with a pitch of 2390Å, giving rise to a helical twist of -0.72° and a rise of 4.78Å. Due to the limited resolution of the PM3 map, we restricted our refinement of PM2 to rigid body fitting, rather than allowing individual atomic refinement. All three chains

span 54 residues from Gly263 to Gly316. The disease-associated mutation D290V of chain A and chain B are solvent exposed on the fibril outer surface. V290 of chain C is buried by van der Waals interactions with Asn293 and Tyr294 of chain B. [Figure.4d]. We acknowledge that the resolution of this map is limited, especially for chain B.

All three polymorphs have a buried PY-NLS where Arg313 is hidden within the fibril core [Fig. S5a-c]. Data and atomic refinement statistics of all three polymorphs are summarized in Table 1.

The fibril core is necessary for hnRNPA2 phase separation

Both wildtype and mutant hnRNPA2 LCD fibril cores span a similar sequence range (Gly263-Tyr319¹⁸ and Gly263-Gly316, respectively). To test the importance of the fibril core in phase separation, we generated a construct of 104 residues of hnRNPA2 LCD without the 57-residue fibril core. Under the conditions that produced phase separation in full-length hnRNPA2, the deletion mutant did not undergo phase separation [Fig. S6]. We thus argue that the residues defining the fibril core are essential for hnRNPA2 phase separation.

The mutant structures are energetically more stable than the wildtype hnRNPA2-LCD

To quantify the structural features that lead to a more stable mutant conformation, we calculated the atomic solvation energy of the PM1-3 fibril structures and compared the values to the wildtype. We obtained the solvation energy per layer of the fibril and an average solvation energy per residue. We illustrated these energies in a solvation energy map where each residue is colored according to its energy. The solvation energy per layer of the wildtype ($-22 \text{ kcal mol}^{-1}$ per layer) has a less negative value than all three polymorphs of the mutant (PM1: $-28 \text{ kcal mol}^{-1}$ per layer; PM2: $-37 \text{ kcal mol}^{-1}$ per layer; PM3: $-55 \text{ kcal mol}^{-1}$ per layer), indicating the wildtype structure is less stable on a per-layer basis than the mutant structures [Figure.5].

The mutant structures have an increased number of β -sheet residues

Steric zippers are pairs of β -sheets where the sidechains mate tightly with each other in a dry interface. Steric zippers are highly energetically stable and are commonly found in pathogenic amyloids²⁴. We compared the number of residues having phi and psi angles within the β -sheet regions of the Ramachandran plots of both wildtype and the mutant structures. We categorized three residue types: S for residues in the β -sheet region, H for residues in either left-handed or right-handed α -helix region, O for other residues that are neither β -sheet nor α -helix. We found an increased number of S residues (WT: 47%, PM1: 56%, PM2: 67%, PM3: 67%) and a decreased number of O residues (WT: 35%, PM1: 26%, PM2: 18%, PM3: 18%) in all three polymorphs of the mutant, and a similar percentage of H residues [Fig. S7]. The mutant structures adopt more steric zipper motifs, namely, more β -sheet residues, than the wildtype which lead to their increased stability.

Discussion

Disease relevance of the structures of the D290V variant of hnRNPA2 low-complexity domain

Advances in cryoEM capabilities have enabled structure determination of amyloid fibrils extracted from various organs, including brain, kidney, and heart. About one third of the amyloid structures determined to date (85/232) originate from ex vivo samples²⁵. The majority; however, originate from recombinant material. For CNS-related diseases, for which structures exist from both ex vivo and in vitro samples, structural differences can be substantial. In other cases, close similarities can be found, extending over the entire fold [Fig. S8a-d], or subregions [Fig. S8e-h]. Accordingly, caution is necessary in ascribing relevance to disease of our structure of

recombinant D290V variant hnRNPA2, which is associated with MultiSystem Proteinopathy (MSP). In the case of MSP, the chance of faithfully reproducing disease-relevant fibrils in vitro may be greater than the faithful reproduction of disease-relevant fibrils of age-related CNS conditions such as Alzheimer's and Parkinson's. MSP, has an age of onset of about 20, which is about half that of Alzheimer's disease. The shorter incubation time for fibril formation may be easier to simulate in a test tube. Moreover, the hnRNPA2 sequence may have evolved to resist amyloid polymorphism; wildtype hnRNPA2 yields only a single definite structure, whereas pathogenic brain amyloids, such as wildtype tau and α -synuclein, adopt several different polymorphs when elicited in vitro.

Relative stability of WT and mutant hnRNPA2 LCD fibrils

We found that the mutant D290V hnRNPA2 LCDs are more stable than the wildtype, as judged both biochemically and computationally, by the following reasons.

(1) The emergence of branched aggregates from the fused liquid droplets of the mutant hnRNPA2 is faster than wildtype. We observed that, after 2 to 3 hours, the mutant hnRNPA2 LCD forms partially-irreversible branched aggregates from its fused liquid droplets. In contrast, after 24 hours the wildtype droplets remain gel-like and reversible. The mutant shows ThT fluorescence immediately when the liquid droplets form and the signal strengthens as the branched aggregates grow, indicating that some of the mutant hnRNPA2 LCDs adopt cross- β structures shortly after the droplets form and over time they assemble into cross- β -rich, branched aggregates. That wildtype droplets fail to display ThT signal until hours later suggests that cross- β structures form gradually in the liquid droplets. At 24 hours, the aggregation we observed in the wildtype droplets resembles what we previously observed, a hydrogel composed of a dense fibril network that is heat-reversible¹⁸. That the wildtype droplets maintain the droplet shape, are

ThT-positive, and are reversible by heating indicates that the wildtype droplets form hydrogels. We cannot rule out the possibility that the ThT fluorescence from the droplets may result from an over-concentration of ThT. But we believe the branched aggregates from the mutant droplets corroborate our idea that there is cross- β structure in the droplets. However, further experiments involving determining the structure of those branched aggregates are required because they are prepared differently. Overall, our results from ThT and phase separation assays imply that the mutant hnRNPA2 LCD has stronger and more stable intra- and inter-molecular interactions than the wildtype.

(2) The mutant hnRNPA2 LCD forms fibrils with a much shorter lag time (~20 minutes) and melts at a higher temperature (65°C) than the wildtype (~25 hours and 55°C). Our results are consistent with previous evidence that, by agitation, the GST-tagged mutant hnRNPA2 LCD forms fibrils faster than the wildtype³, and the mutant hnRNPA2 LCD is thermodynamically more stable than wildtype²⁶. Our results suggest that the mutant hnRNPA2 LCD more easily assembles into fibrils which are more resistant to dissociation than wildtype fibrils.

(3) Aggregation behavior of the mutant hnRNPA2 LCD is less sensitive to pH changes than the wildtype. Aggregates of mutant hnRNPA2 exhibit the same branched morphology over a wide pH range, unlike wildtype. When the pH is below 4, wildtype has more abundant and thicker branched aggregates than the mutant hnRNPA2 LCD, because wildtype Asp290 will be protonated and less likely to form H-bonds with residues Tyr288 and Gly292 amide backbone [Fig. S9]. Thus, we hypothesize that the wildtype may adopt an aggregated form potentially similar to the structures the mutant will adopt at physiological pHs. We conclude that both wildtype and the mutant hnRNPA2 LCDs phase separate and aggregate at high concentrations under physiological conditions, adopting cross- β scaffolds, with the mutant scaffold being more

stable than wildtype, possibly because the interaction motifs of wildtype hnRNPA2 LCDs are mainly LARKS rather than steric zippers.

(4) The mutant hnRNPA2 fibrils are energetically more stable than wildtype according to solvation energy estimates calculated from hnRNPA2 LCD fibril structures. The calculations are consistent with our biochemical experiments. The mutant structure has a more negative value of solvation energy per layer (PM1: $-28 \text{ kcal mol}^{-1}$ per layer; PM2: $-37 \text{ kcal mol}^{-1}$ per layer; PM3: $-55 \text{ kcal mol}^{-1}$ per layer) than the wildtype ($-22 \text{ kcal mol}^{-1}$ per layer), indicating that it is energetically more stable and therefore requires more energy to dissociate it than the wildtype. The mutant structure solvation energy is in the middle of the energy spectrum in terms of solvation energy values²⁵, where it is not as stable as irreversible pathogenic amyloid fibrils such as human amyloid serum A²⁷ ($-69 \text{ kcal mol}^{-1}$ per layer) and Tau²⁸ ($-63 \text{ kcal mol}^{-1}$ per layer), but is much more stable than the reversible functional amyloid-like fibrils such as FUS²⁹ ($-12 \text{ kcal mol}^{-1}$ per layer) and wildtype hnRNPA2¹⁸ ($-22 \text{ kcal mol}^{-1}$ per layer). This implies that the higher stability of the mutant structure might not be the only reason for its accumulation in the cytoplasm.

Possible structural mechanisms of D290V mutant hnRNPA2 accumulation in the cytoplasm

Kim et al., reported in 2013 that the D290V mutant hnRNPA2 is cleared out from the nucleus and forms cytoplasmic inclusions. Given the fibril structures of the wildtype and mutant LCDs of hnRNPA2 and their relative stabilities, what is the mechanism by which a missense mutation diverts the assembly of reversible functional amyloid-like fibrils to assembly of irreversible pathogenic amyloid that leads to its accumulation in the cytoplasm? We ruled out the possibility that a difference in the polarity of amino acid composition of the fibril core contributes to a difference in fibril stability. The mutant and the wildtype hnRNPA2 LCDs have similar fibril

core sequences with almost identical polar residue compositions (47% of polar residues in wildtype and 46% in the mutant). So, we must look further to understand the dramatic biological effect of a single missense mutation.

In the following, we propose four structural mechanisms that could explain why the D290V mutant hnRNPA2 accumulates in the cytoplasm.

1) A loss of function for mutant hnRNPA2 arises because its PY-NLS is obscured in the fibril core, disrupting its designated nuclear localization. Unlike the wildtype hnRNPA2 LCD structure, all three polymorphs of the mutant hnRNPA2 LCD structures have a PY-NLS where the arginine residue, R313, is deeply buried in the fibril core, making it inaccessible to chaperones such as karyopherin- β 2 (Kap β 2) [Figure.4 & Fig. S5]. Previous work has shown that the conserved arginine, acting as an epitope of the RX₂₋₅PY motifs of PY-NLS of hnRNPA1, hnRNPD and FUS binds to the C-terminal arch of Kap β 2 by forming a salt bridge³⁰⁻³⁵. Thus, when R313 of hnRNPA2 is buried inside the D290V mutant fibril core, the PY-NLS loses its function because the conserved arginine can no longer interact with Kap β 2³⁵ [Fig. S5]. In short, the structure of the mutant fibril offers a simple mechanism by which a genetic lesion promotes development of disease.

2) The D290V aggregates are somewhat more stable than the WT aggregates, as discussed in the preceding section. Compared to the wildtype hnRNPA2 structure, there is an increased number of residues adopting β -sheet conformations in the mutant structures. In the wildtype structure, the Asp290 side chain forms a Van der Waals interaction and a hydrogen bond network with the Tyr288 side chain and Gly292 backbone amide [Fig. S10]. The mutant molecule losing this hydrogen bond results in Gly292 adopting a β -sheet conformation rather than a kinked conformation. Due to this loss of hydrogen bond, we noticed a number of structural differences

deeper in the fold that appear linked to residue 290. Notably, the sidechains of half of all the residues flip between inward and outward facing orientations [Fig. S9b], and there is an overall increase in the β -sheet content of the mutant structures compared to wildtype both upstream and downstream of Val290 [Fig. S7b]. Many upstream Glycine residues change into β -sheet conformation, including Gly265, 268-269, 272-273 and Gly285. Multiple downstream polar residues, including Tyr294, Gln297, Asn300, Tyr301, and Met315, also adopt β -sheet conformation in the mutant structures. Pathogenic amyloid fibrils normally have a high β -sheet content (<https://srv.mbi.ucla.edu/AmyloidAtlas/>) creating stability²⁵. We believe the stability of the mutant hnRNPA2 to be rooted in the higher β -sheet content of its LCD.

3) The side chain of the solvent-exposed V290 disease-associated replacement presumably makes non-specific interactions with other proteins in cells. It is unexpected to us that the disease-associated valine side chain is solvent exposed in all three mutant polymorphs because burying this hydrophobic residue would make the structure more stable. In PM3 of the mutant structure, the exposed valine interacts with protofilament B. We thus speculate that the exposed valine can also have non-specific hydrophobic interactions with other proteins or RNAs in the MLOs³⁶⁻³⁸, resulting in a sequestration and a loss of function of other proteins. It is also interesting to us that half of the inward facing residues in wildtype structure become outward facing in the mutant structures and vice versa [Fig. S9a]. These changes lead to new surface epitopes that could potentially contribute to the mutant's toxicity in cells.

4) Similar to other pathogenic amyloid fibrils such as tau²⁸, α -synuclein³⁹, and islet amyloid polypeptide, the hnRNPA2 LCD mutant structures are polymorphic. The mutant structures are composed of two or three protofilaments with the same fold, but the protofilaments are arranged slightly differently from each other [Figure.4]. PM1 differs from PM2 with an 8-residue larger

interface, and PM2 differs from PM3 with an additional protofilament. The other three unsolved structures are also different arrangements of PM1/2 or PM3, with an even smaller interface or an additional protofilament interacting at another position [Fig. S4b]. In principle, all six polymorphs of the mutant hnRNPA2 LCD can have different first and second nucleation rates, fragmentation rates, and extension rates, resulting in some polymorphs (PM1-PM3) predominating over the others. Polymorphism can affect the self-propagation rate by providing more templates to seed and extend⁴⁰. It has been shown that different polymorphs of α -synuclein fibrils have different seeding capacities and cytotoxicity profiles⁴¹. The altered nucleation and seeding rates will likely change the dynamics of RNA-containing MLOs and disrupt cellular RNA metabolism. Most functional amyloids fold into a single morphology presumably through evolution. Due to the sensitivity of amyloids to change structures under different environmental conditions or mutations, it is not surprising that D290V of hnRNPA2 leads to a such dramatic effect.

Our experiments confirm previous ideas that phase separation, formation of hydrogel and fibril formation are connected events in the aggregation of both wildtype and mutant hnRNPA2 LCD, and probably sequential states^{20,42}. Similar phenomenon are also observed for FUS⁴ and TDP-43^{5,6}. Thus, our structures for both wildtype and mutant fibrils establish their clear differences, including monomorphic versus polymorphic fibrils. At the same time, we find the mutant differs from other pathogenic amyloids in that it is not as energetically stable. Our results show that the mutant hnRNPA2 LCD has a medium solvation energy value between functional fibrils and pathogenic amyloid, indicating that stability is not the only factor that contributes to its accumulation in the cytoplasm. We propose four mechanisms in addition to stability to explain:

1) Loss-of-function by the hidden PY-NLS. 2) An increase in the β -sheet content. 3) Polymorphism. 4) Non-specific interactions by exposed Valine 290.

Of importance, our D290V mutant hnRNPA2 structure reveals a difference from the wildtype structure that is consistent with known loss of function of the mutant: hnRNPA2 is an RNA-binding ribonucleoprotein that functions normally in mRNA transport from the nucleus to the cytoplasm and then to the designated locations for translation. However, in two families afflicted with MSP, hnRNPA2 was found in muscle cell cytoplasm rather than in nuclei. This loss of normal function is consistent with the salient point of our mutant structure: that the nuclear localization signal is obscured, which could account for the mis-localization of hnRNPA2 in these cases. Recent studies by Taylor's group have shown that a frameshift in the PY-NLS of hnRNPA2 reduces its affinity to chaperone Kap β 2, leading to hnRNPA2's accumulation in the cytoplasm⁴³. Both their and our results show the biological significance of a fully accessible and intact PY-NLS.

In summary, comparison of the recombinant wildtype and recombinant D290V mutant hnRNPA2 structures offers a hypothesis for how a missense mutation leads to nuclear clearance and cytoplasmic accumulations, which is that the D290V mutant sequence forms relatively stable fibrils that mask the nuclear import signal sequence. Such a structure-based hypothesis for disease origin has not emerged from the disease-related fibril structures of tau, α -synuclein, or amyloid β , despite their abundance.

Materials and Methods

Materials and purification of mCh-hnRNPA2 LCD fusion protein

The construct for overexpression of both the wildtype and the mutant mCherry-hnRNPA2-LCD fusion protein was provided by Dr. Masato Kato of University of Texas, Southwestern. A TEV cleavage site was inserted between mCherry and hnRNPA2 LCD. HnRNPA2 LCD without the fibril core construct was designed using benchling and ordered through Twist Bioscience.

Protein overexpression and purification procedures followed the same protocol reported previously¹⁸. Protein was overexpressed in *E. coli* BL21(DE3) cells with 0.5 mM IPTG at 25 °C for overnight. LB media with 0.1 mg ml⁻¹ ampicillin was used for cell culture. Harvested cells were resuspended in lysis buffer containing 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 M guanidine HCl, 2 mM TCEP, and HaltTM protease inhibitor cocktail (Thermo Scientific) for 30 min on ice, and then sonicated. The cell lysate was centrifuged at 32,000×g for an hour. The supernatant was filtered and loaded onto a 5ml HisTrap HP column (GE healthcare) for purification. The HisTrap column was pre-equilibrated with the lysis buffer. After proteins were loaded onto the column, proteins were washed with the lysis buffer until the UV280 spectrum line became flat. The sample then was washed with a 20 column volume gradient at the speed of 1 ml/min from 100% wash buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 M Guanidine HCl, 20 mM imidazole, and 2 mM TCEP to 100% elution buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 M Guanidine HCl, 300 mM imidazole, and 2 mM TCEP. Purified mCherry-hnRNPA2-LCD fusion proteins were dialyzed overnight at room temperature against a dialysis buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA, and 0.1 mM PMSF. The protein solutions were concentrated to 40–80 mg ml⁻¹ and flash frozen in liquid nitrogen and stored in -80 °C for future use.

TEV cleavage and formation of hnRNPA2 LCD fibrils

Purified and concentrated mCherry-hnRNPA2-LCD was incubated with 1:200 v/v of 10 mg ml⁻¹ TEV protease at 25 °C overnight. Proteins were centrifuged at 13,000×g for 10 min, the supernatant was discarded, and the protein was washed and remixed with dialysis buffer. The same process was repeated three times. The LCD alone was then dissolved in buffer containing 20 mM MOPS, pH 7.5, 150 mM KCl buffer and 2 M Guanidinium Hydrochloride to a final concentration of 4-5 mg ml⁻¹. Samples then were shaken at 4 °C for 3 days using a digital dry bath (Torrey Pines Scientific) at speed 9. Fibrils presence was checked by negative-stain EM.

Negative staining TEM

Samples for TEM were prepared by applying 2.5 µl of sample on glow-discharged 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The samples were allowed to adhere for 2 min, and washed twice with water. The samples were then stained for 2 min with 2% uranyl acetate and allowed to dry for 1 min. Each grid was inspected using a T12 (FEI) electron microscope.

Phase separation and “aging” experiments with ThT

Purified and concentrated mCherry-hnRNPA2-LCDs were incubated with TEV protease and washed as described in the previous section. The LCD alone was dissolved in 5.7 M GnCl, and then diluted to 100 µM GnCl using the buffer containing 20 mM Tris-HCl, pH 7.4 and 150 mM NaCl. Samples immediately turned turbid and underwent phase separation.

For the “aging” experiments, samples were put in a 96-well plate and monitored over time using a differential interference contrast (DIC) microscopy (Zeiss). For phase separation with ThT assays, 20 µM final concentration of ThT was added to the sample and monitored over time using Zeiss. The plate was imaged in phase contrast channel (DIC) with 300 ms exposure time,

in the BO-PRO-1 channel (ThT) with 100 ms exposure time. BO-PRO-1 channel has an excitation and emission of 462nm and 482nm. All images were taken at full camera resolution without any binning. Measurements were taken at every 15 minutes, and representative images were presented in the main texts. Liquid droplets diameter quantification is measured using the software Nano Measure 1.2.

For phase separation experiments with different pHs, buffers we used are: 20 mM sodium acetate, pH 1.3 and 150 mM NaCl; 20 mM sodium acetate with acetic acid, pH 4.1 and 150 mM NaCl; 20 mM sodium acetate with acetic acid, pH 5.5 and 150 mM NaCl; 20 mM Tris-HCl, pH 7.5 and 150 mM NaCl; 20 mM CHES, pH 9.0 and 150 mM NaCl; 20 mM CHES, pH 10.0 and 150 mM NaCl.

For liquid droplets thermostability assays, phase separation was formed in the 96-well plate and sat at 25 °C for 24 hours. The plate was then incubated at 37 °C for an hour and stirred with pipette. The 96-well plate was then incubated at 4 °C overnight, and stirred with pipette the next day. Sample status was checked and recorded using Zeiss.

Thermostability assays of hnRNPA2 LCD fibrils

Both the wildtype and the mutant hnRNPA2 LCD alone fibrils were heated up from 4 °C to 95 °C with a 10 °C increment, at each temperature samples were incubated for 15 min in a PCR machine. 2.5 µl of samples at each temperature were taken and checked by negative stain EM. After heated at 95 °C for 15 min, the samples were incubated at 4 °C overnight and then checked by negative stain EM.

Cryo-EM data collection, reconstruction, and model building

2.5 μl of hnRNPA2 LCD alone fibrils at a concentration of 4 mg ml^{-1} were applied to glow-discharged Quantifoil Cu R 1.2/1.3, 300 mesh carbon grids. Samples were blotted with filter paper to remove excess sample and then plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI). Cryo-EM data were collected on a Gatan K2 Summit direct electron detector on a Titan Krios (FEI) microscope equipped with a Gatan Quantum LS/K2 Summit direct electron detection camera (operated with 300 kV acceleration voltage and slit width of 20 eV). Super-resolution movies were acquired with a pixel size of $1.1 \text{ \AA pixel}^{-1}$ ($0.55 \text{ \AA pixel}^{-1}$ in super-resolution movie frame). Thirty movie frames were recorded each with a frame rate of 5 Hz using a dose rate of $1.2 \text{ e}^{-} \text{ \AA}^{-2} \text{ frame}^{-1}$ for a final dose of $36 \text{ e}^{-} \text{ \AA}^{-2}$ on the sample. Automated data collection was driven by the Leginon automation software package⁴⁴. 4366 micrographs were collected with a defocus range from 0.5 to $5.0 \mu\text{m}$.

Unblur⁴⁵ was used to correct beam-induced motion with dose weighting and anisotropic magnification correction, leading to a physical pixel size of $1.064 \text{ \AA pixel}^{-1}$. Micrographs summing all frames were all corrected for gain reference and then micrographs with a group of three frames were used to estimate the contrast transfer function using CTFFIND 4.1.8⁴⁶. Micrographs with crystalline ice, severe astigmatism, or obvious drift were discarded. All subsequent data processing was performed using RELION 3.1.1^{47,48}.

All filaments were picked using crYOLO⁴⁹, trained with 100 micrographs we picked manually using EMAN2 e2helixboxer.py⁵⁰. Particles were first extracted using a box size of 1024 and 686 pixels with an inter-box distance of 10% of the box length. 2D classification using 1024-pixel particles was used to estimate the fibril pitch and helical parameters. We also performed 2D classifications with 686-pixel particles to select particles for future 3D classification. We performed Class3D jobs with three classes and manually controlled the tau_fudge factor and

healpix_order to obtain resolutions around 8–9 Å, using an elongated Gaussian blob as an initial reference. We selected particles that contribute to the highest resolution class and generated an initial 3D reconstruction by running Class3D with 1 class. To obtain a higher resolution reconstruction, we re-extracted particles with a box size of 320 pixels from the fibril tubes containing the Class3D-selected 686-pixel particles. All 320-pixel particles were used directly for 3D classifications. The final subset of selected particles was used for high-resolution gold-standard refinement as described previously⁴⁸. We apply a lowpass filter of 40 Å (--ini_high 40). The final overall resolution estimates for PM1-3 were evaluated to be 3.3 Å, 3.2 Å, and 3.9 Å, respectively, based on the 0.143 FSC cutoff⁵¹. Projections from the final reconstruction closely match the 2D class averages, helping to validate the reconstruction.

The refined map was sharpened using phenix.auto_sharpen at the resolution cutoff⁵², and a near-atomic resolution model was built de novo into the sharpened map using COOT⁵³. We generated a five-layer model using the helical parameters from the final 3D refinement and then refined the structure using phenix.real_space_refine⁵⁴. After the last round of refinement, the final model was validated using phenix.comprehensive_validation^{55,56}. The same protocols were performed for all three polymorphs. All the statistics are summarized in Table 1.

Atomic Solvation energy calculations

The solvation energy was calculated as described previously^{25,57}. The solvation energy for each residue was calculated by the sum of the products of the area buried for each atom and the corresponding atomic solvation parameters. The overall energy was calculated by the sum of energies of all residues. Different colors were assigned to each residue, instead of each atom, in the solvation energy map. The energy reported for FUS is the average over 20 NMR models.

Acknowledgments

Funding: We are grateful to Dr. Masato Kato and Dr. Steven L. McKnight from University of Texas, Southwestern Medical Center for sending us the construct of mCherry-hnRNPA2-LCD and their advice on sample preparation. We thank Lukasz Salwinski for his help in constructs cloning. We acknowledge the use of instruments at the EICN supported by NIH Grants 1R01AG070895 and R01 AG048120, NSF Grant MCB 1616265, and the California NanoSystems Institute at the University of California, Los Angeles. The authors acknowledge NIH Grant 1RF1AG065407, and Department of Energy Grant DE-FC02-02ER63421 for support. J.L. was supported by the Audree V. Fowler fellowship in Protein Science and Dissertation Year Fellowship.

Author contributions: J.L. designed experiments, purified constructs, performed phase separation and biochemical assays, fibrillized proteins, prepared cryo-EM samples, performed cryo-EM data collection and processing, performed data analysis. P.G. assisted in cryo-EM data collection and processing. M.R.S. assisted in performing solvation energy calculations. M.R.S. and D.C. assisted in model building. M.P.H performed LARKS prediction and assisted in figure making. Q.C. and D.R.B. assisted in data processing. R.A. assisted in constructs design. E.T assisted in phase separation assays. J.L., M.R.S., M.P.H., D.S.E. wrote the manuscript. D.S.E. supervised and guided the project.

Competing interests: D.S.E. is an advisor and equity shareholder in ADRx, Inc.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The cryoEM density maps of PM1-3 of the D290V mutant of

hnRNPA2 have been deposited in the Electron Microscopy Data Bank under the accession code EMD-27713, EMD-27728, and EMD-28014 respectively. The resulting atomic model has been deposited in the Protein Data Bank under the accession code PDB-8DU2, PDB-8DUW, and PDB-8EC7 respectively.

Figures and Tables

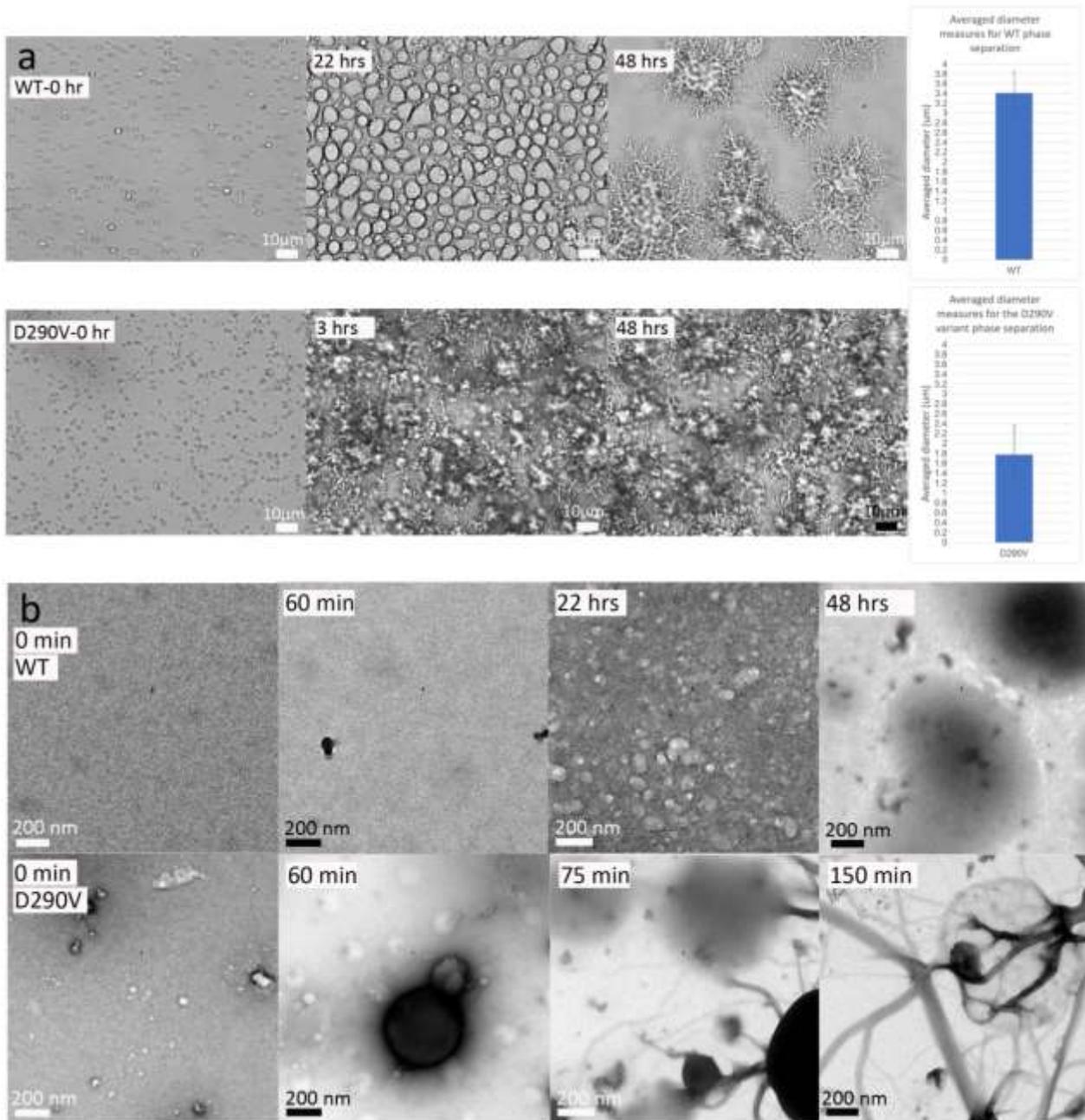


Figure.1 Phase separation of hnRNPA2 LCD and its aggregation

a) Images were taken using differential interference contrast microscopy (DIC). Scale bar: 10 μm .

Top panel: Phase separation of wildtype hnRNPA2 LCD at time points of 0 hour, 22 hours, and 48 hours. Rightmost is a quantification of averaged diameter of WT liquid droplets.

Bottom panel: Phase separation of the mutant D290V hnRNPA2 LCD at time points of 0 hour, 3 hours, and 48 hours. Branched aggregates were observed for the mutant at much earlier time points than for the wildtype. Rightmost is a quantification of averaged diameter of the D290V mutant liquid droplets observed in the left panel.

b) Negative stain EM images of hnRNPA2 LCD phase separation at different timepoints. Scale bar: 200 nm.

Top panel: Wildtype hnRNPA2 LCD phase separation at timepoints of 0 min, 60 min, 22 hours and 48 hours.

Bottom panel: The D290V mutant hnRNPA2 LCD phase separation at timepoints of 0 min, 60 min, 75min and 150min.

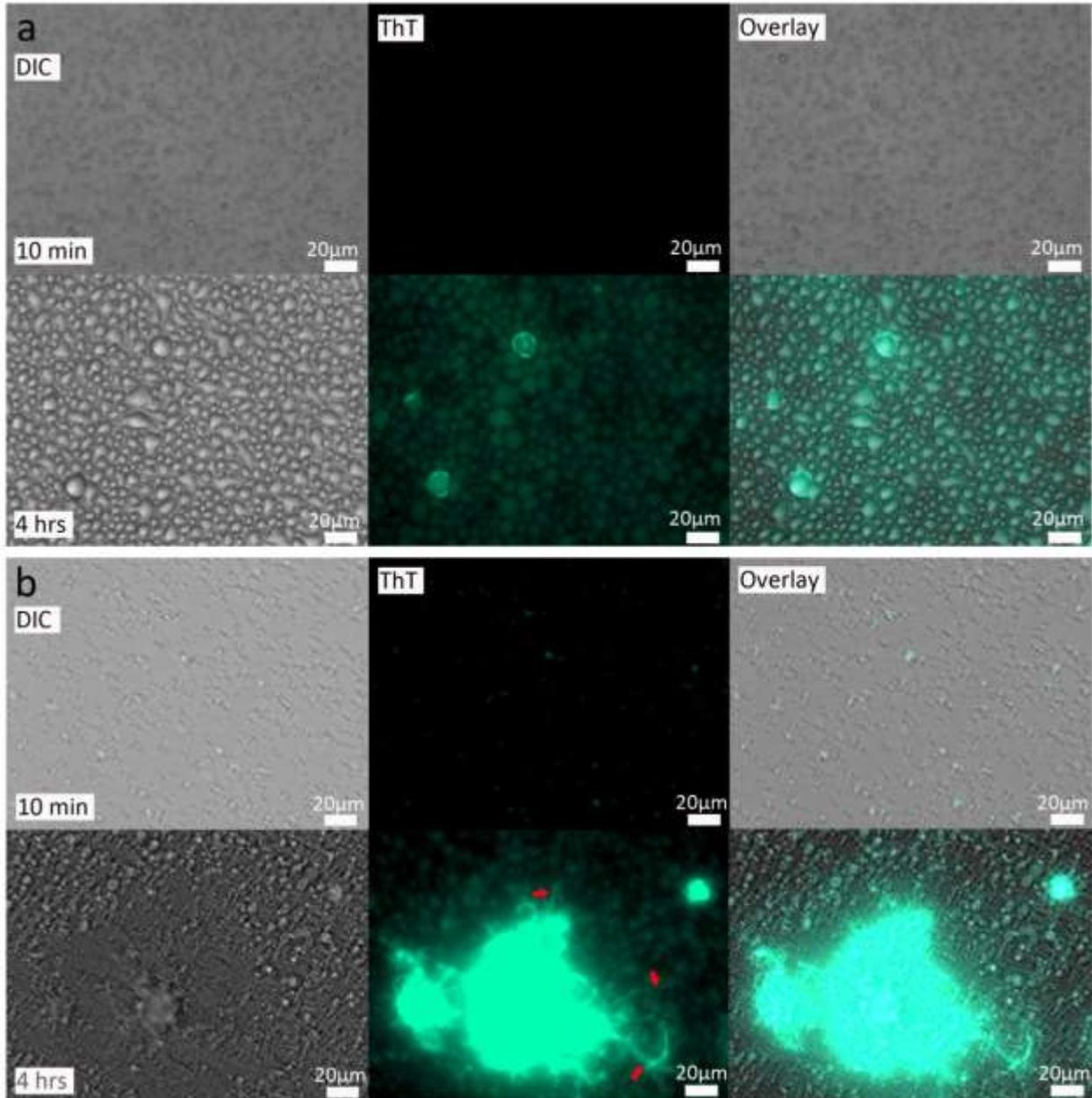


Figure.2 Phase separation of hnRNPA2 LCD is ThT-positive

Phase separation images of hnRNPA2 LCD taken by differential interference contrast (DIC) microscopy (Zeiss). Left panel is phase contrast channel image (DIC), middle panel is BO-PRO-1 channel image (BO-PRO-1 channel has an excitation and emission of 462nm and 482nm, respectively. We use it for ThT), right panel is the overlay of DIC and ThT.

a) Wildtype hnRNPA2 LCD phase separation at timepoints of 10 min and 4 hours. Scale bar: 20 μm.

b) The mutant D290V hnRNPA2 LCD phase separation at timepoints of 10 min and 4 hours. Red arrows show the outlines of fibril-looking aggregates. Scale bar: 20 μm.

a

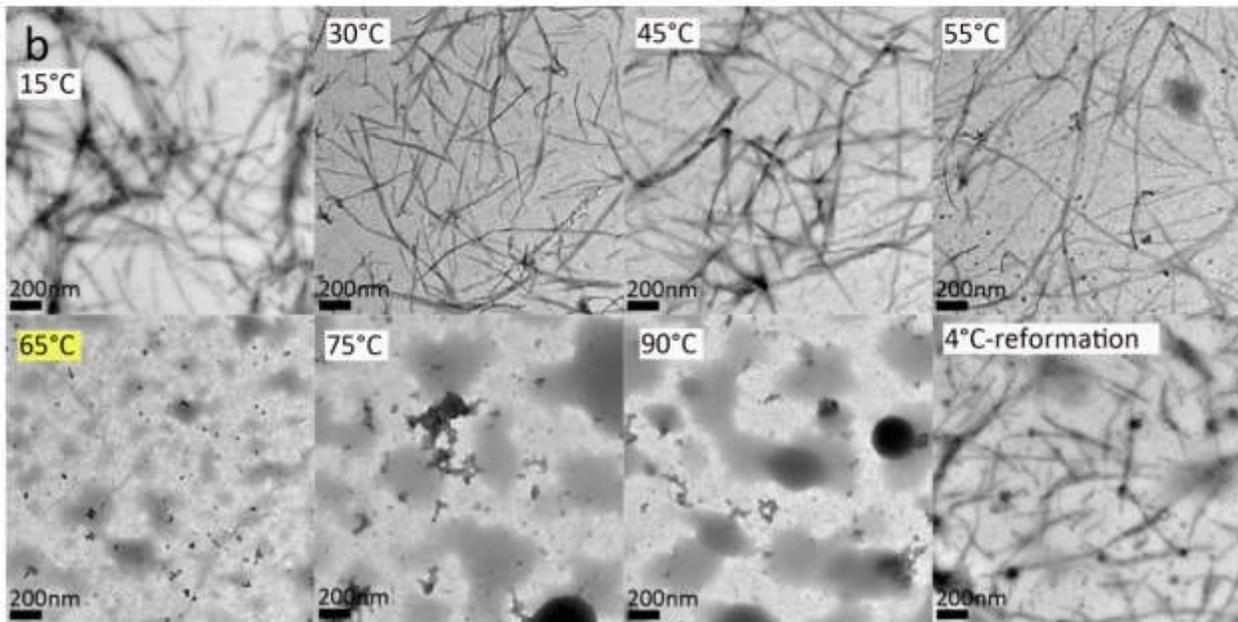
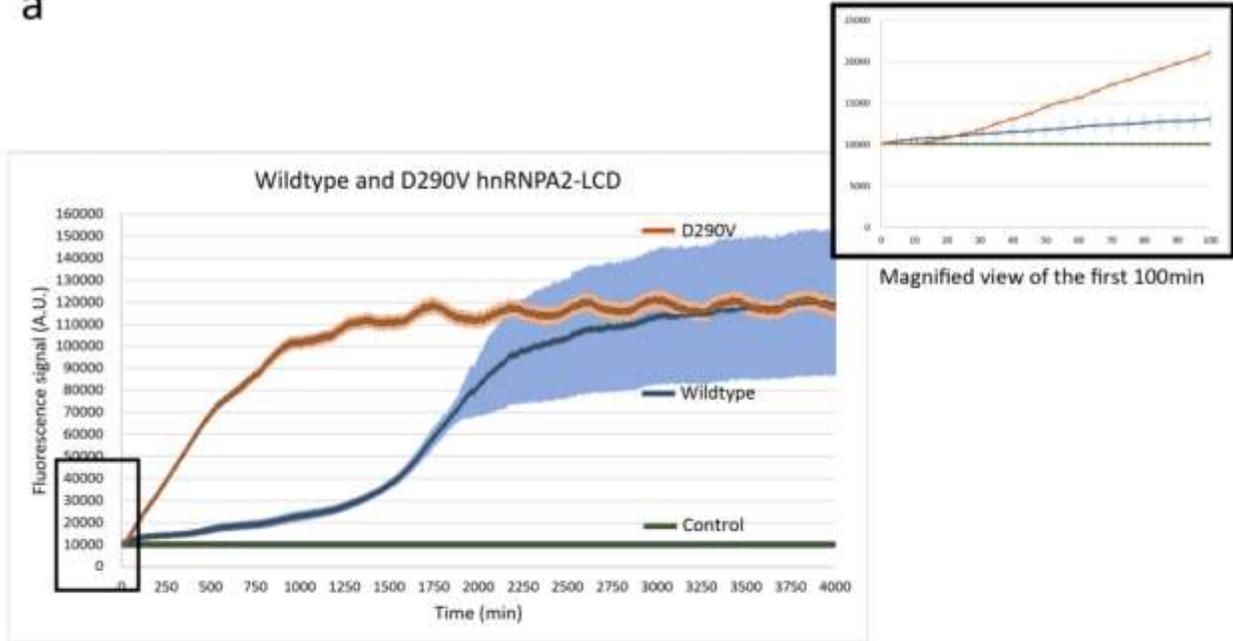


Figure.3 ThT and thermostability assays of hnRNPA2 LCDs

a) ThT assay of wildtype and the mutant D290V hnRNPA2 LCDs. X-axis represents time in the unit of minutes, Y-axis represents fluorescence signal. The wildtype ThT curve is colored blue, the mutant ThT curve is colored orange, the control ThT curve with buffer only is colored green. The box shows a magnified view of the first 100 minutes of the ThT curve.

b) Representative negative stain EM images of the thermostability assay of the mutant D290V hnRNPA2 LCD. The mutant fibrils are melted at 65 °C, which is colored yellow to highlight this transition. Scale bar: 200 nm.

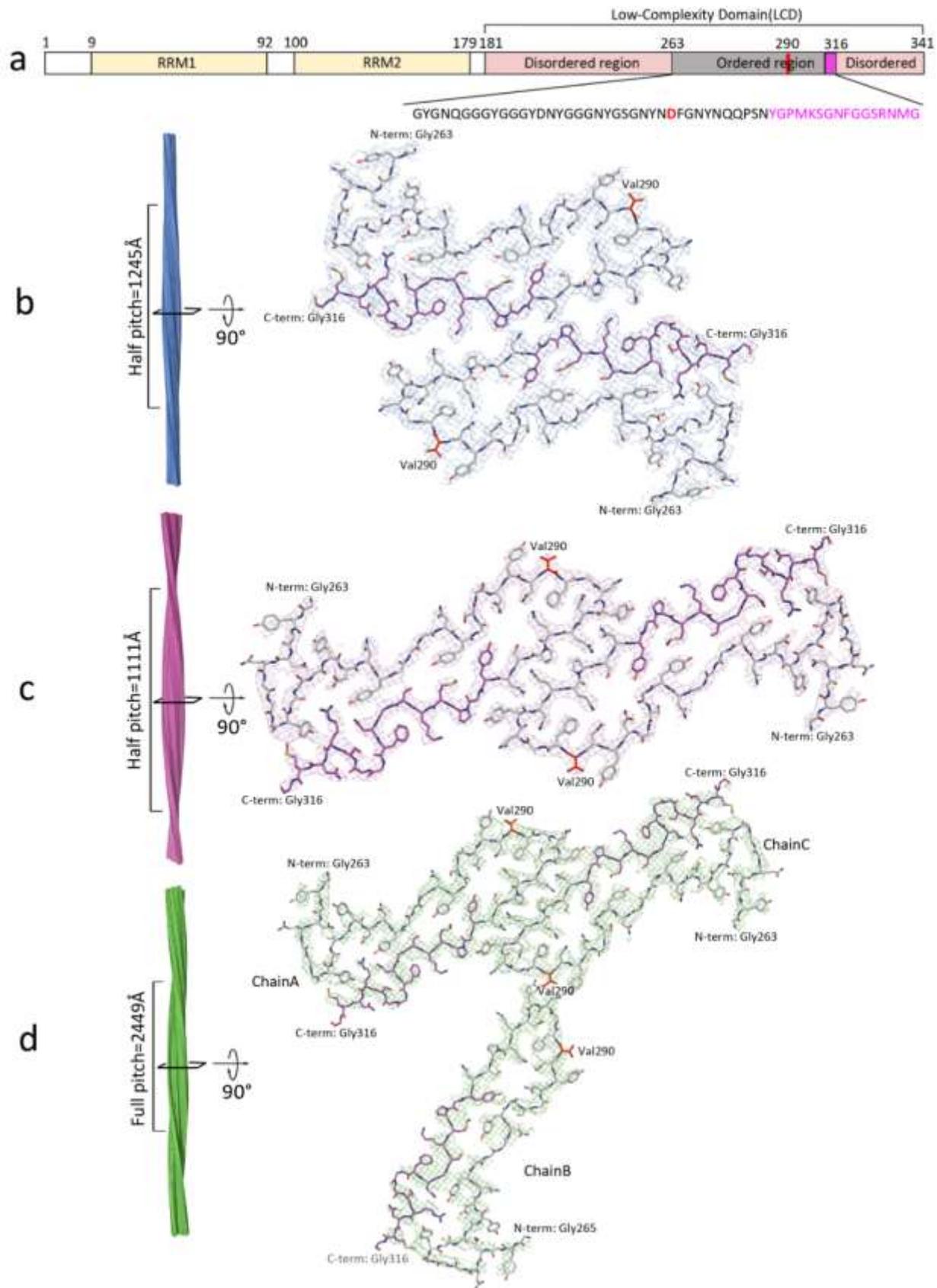


Figure.4 The cryo-EM structures of the mutant D290V hnRNPA2 LCD fibril cores

a) Domain structure of full-length hnRNPA2. The LCD (residues 181–341) was used for fibril structure determination. The gray bar shows the range of the residues that form the ordered fibril core of the cryo-EM structure. The red bar shows the disease-causing mutation. The magenta bar shows the PY-NLS. The sequence of the ordered region is shown below with corresponding colors.

b-d) b,c,d are PM1, PM2, PM3, respectively. Left: Fibril reconstruction showing left-handed pitch and twist. Right: Density and atomic model of one cross-sectional layer of each fibril. Disease-causing mutation is colored red, and PY-NLS residues are colored magenta. N-term: N-terminus; C-term: C-terminus.

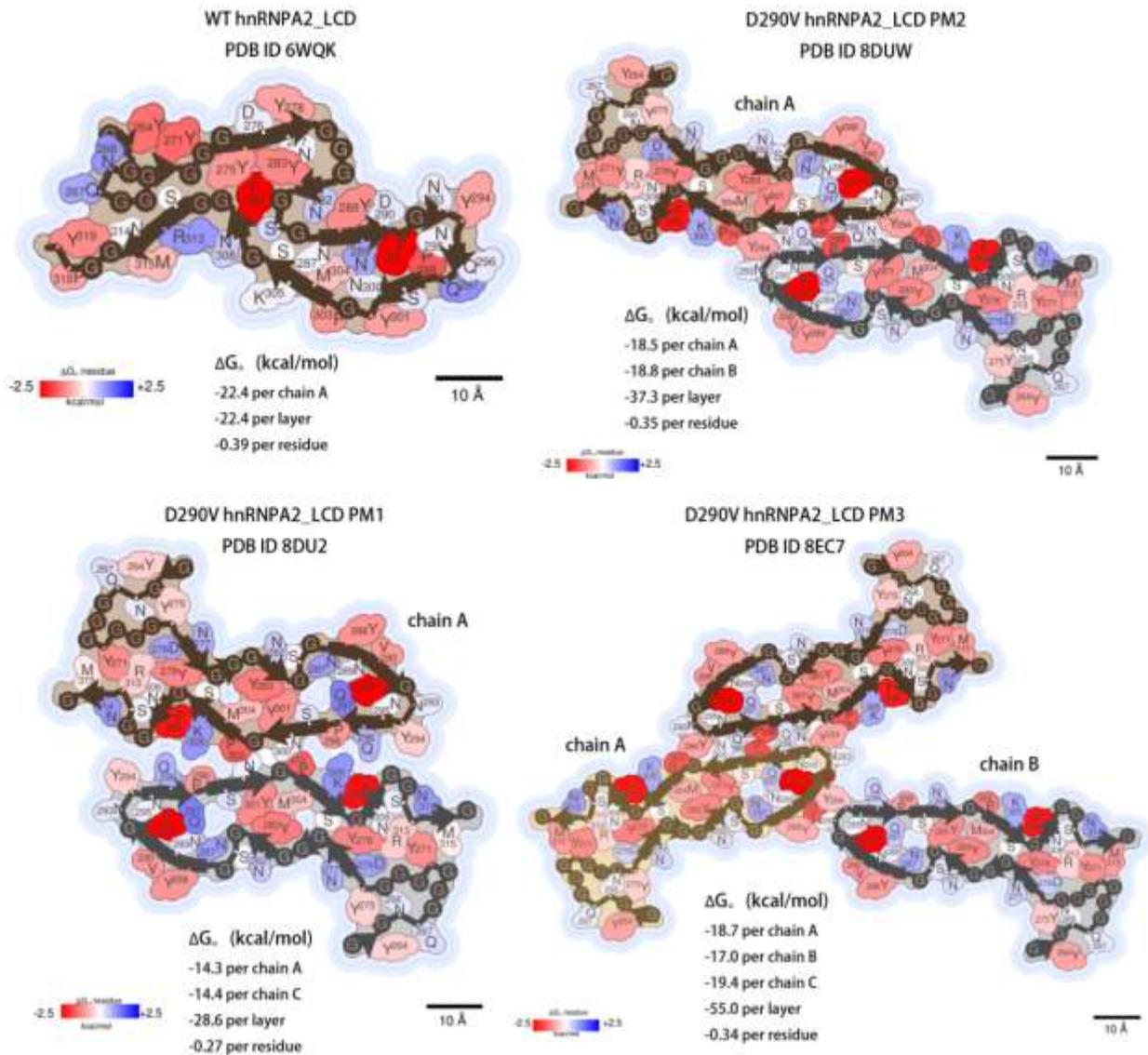


Figure.5 Calculated solvation energy maps for hnRNPA2 LCD fibrils based on atomic solvation parameters

Residues are colored according to their stabilization energies from unfavorable (blue, $+2.5 \text{ kcal mol}^{-1}$) to favorable (red, $-2.5 \text{ kcal mol}^{-1}$). Top left: wildtype hnRNPA2 LCD fibril core; Top right: PM2 of the mutant hnRNPA2 LCD fibril core; Bottom left: PM1 of the mutant hnRNPA2 LCD fibril core; Bottom right: PM3 of the mutant hnRNPA2 LCD fibril core. Scale bar: 10\AA .

	#1 PM1 (EMDB-27713) (PDB 8DU2)	#2 PM2 (EMDB-27728) (PDB 8DUW)	#3 PM3 (EMDB-28014) (PDB 8EC7)
Data collection and processing			
Magnification	X81,000	X81,000	x81,000
Voltage (kV)	300	300	300
Electron exposure (e-/Å ²)	36	36	36
Defocus range (μm)	0.5-5.0	0.5-5.0	0.5-5.0
Pixel size (Å)	1.1	1.1	1.1
Symmetry imposed	C1/P2 ₁	C1/P2 ₁	C1
Initial particle images (no.)	128565	197905	171774
Final particle images (no.)	5675	14902	16114
Map resolution (Å)	3.3	3.2	3.9
FSC threshold	0.143	0.143	0.143
Helical rise (Å)	2.40	2.39	4.78
Helical twist (°)	179.63	179.60	-0.72
Map resolution range (Å)	200-3.3	200-3.2	200-3.9
Refinement			
Initial model used (PDB code)	De Novo	De Novo	Rigid body of PM2
Model resolution (Å)	4.0	3.4	4.5
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)	200-3.6	200-3.4	200-4.2
Map sharpening <i>B</i> factor (Å ²)	108	76	51
Model composition			
Non-hydrogen atoms	3900	3900	5850
Protein residues	540	540	810
<i>B</i> factors (Å ²)			

Protein	79.00	59.21	109.54
R.m.s. deviations			
Bond lengths (Å)	0.003	0.006	0.002
Bond angles (°)	0.469	0.514	0.393
Validation			
MolProbity score	1.39	1.55	1.36
Clashscore	7.17	5.34	6.47
Poor rotamers (%)	0.00	0.00	0.00
Ramachandran plot			
Favored (%)	98.08	96.15	98.08
Allowed (%)	1.92	3.85	1.92
Disallowed (%)	0.00	0.00	0.00
CC (mask) model vs. data	0.81	0.83	0.67

Table.1 Statistics of cryo-EM data collection, refinement, and validation

References

1. Multisystem proteinopathy: Where myopathy and motor neuron disease converge - Korb - 2021 - Muscle & Nerve - Wiley Online Library.
<https://onlinelibrary.wiley.com/doi/10.1002/mus.27097>.
2. Watts, G. D. J. *et al.* Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat. Genet.* **36**, 377–381 (2004).
3. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
4. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
5. Sarcoplasmic redistribution of nuclear TDP-43 in inclusion body myositis - Salajegheh - 2009 - Muscle & Nerve - Wiley Online Library.
<https://onlinelibrary.wiley.com/doi/10.1002/mus.21386>.
6. Fonda, B. D., Jami, K. M., Boulos, N. R. & Murray, D. T. Identification of the Rigid Core for Aged Liquid Droplets of an RNA-Binding Protein Low Complexity Domain. *J. Am. Chem. Soc.* **143**, 6657–6668 (2021).
7. Ryan, V. H. *et al.* Weak binding to the A2RE RNA rigidifies hnRNPA2 RRMs and reduces liquid–liquid phase separation and aggregation. *Nucleic Acids Res.*
[doi:10.1093/nar/gkaa710](https://doi.org/10.1093/nar/gkaa710).

8. The roles of hnRNP A2/B1 in RNA biology and disease - Liu - 2021 - WIREs RNA - Wiley Online Library. <https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1612>.
9. White, R. *et al.* Activation of oligodendroglial Fyn kinase enhances translation of mRNAs transported in hnRNP A2-dependent RNA granules. *J. Cell Biol.* **181**, 579–586 (2008).
10. Amaya, J., Ryan, V. H. & Fawzi, N. L. The SH3 domain of Fyn kinase interacts with and induces liquid–liquid phase separation of the low-complexity domain of hnRNPA2. *J. Biol. Chem.* **293**, 19522–19531 (2018).
11. Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell* **69**, 465-479.e7 (2018).
12. Conicella, A. E. *et al.* TDP-43 α -helical structure tunes liquid–liquid phase separation and function. *Proc. Natl. Acad. Sci.* **117**, 5883–5894 (2020).
13. Ding, X., Gu, S., Xue, S. & Luo, S.-Z. Disease-associated mutations affect TIA1 phase separation and aggregation in a proline-dependent manner. *Brain Res.* **1768**, 147589 (2021).
14. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *eLife* **5**, e18413 (2016).
15. Buratti, E. *et al.* TDP-43 Binds Heterogeneous Nuclear Ribonucleoprotein A/B through Its C-terminal Tail: AN IMPORTANT REGION FOR THE INHIBITION OF CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR EXON 9 SPLICING*. *J. Biol. Chem.* **280**, 37572–37584 (2005).

16. Qi, X. *et al.* Familial Early-Onset Paget's Disease of Bone Associated with a Novel hnRNPA2B1 Mutation. *Calcif. Tissue Int.* **101**, 159–169 (2017).
17. Forman-Kay, J. D., Kriwacki, R. W. & Seydoux, G. Phase Separation in Biology and Disease. *J. Mol. Biol.* **430**, 4603–4606 (2018).
18. Lu, J. *et al.* CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nat. Commun.* **11**, 4090 (2020).
19. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018).
20. Xiang, S. *et al.* The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets and Nuclei. *Cell* **163**, 829–839 (2015).
21. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci.* **107**, 3487–3492 (2010).
22. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1804**, 1405–1412 (2010).
23. Springhower, C. E., Rosen, M. K. & Chook, Y. M. Karyopherins and condensates. *Curr. Opin. Cell Biol.* **64**, 112–123 (2020).
24. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
25. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R. & Eisenberg, D. S. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873 (2021).

26. Murray, D. T. *et al.* Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9782–E9791 (2018).
27. Lu, J., Yu, Y., Zhu, I., Cheng, Y. & Sun, P. D. Structural mechanism of serum amyloid A-mediated inflammatory amyloidosis. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5189–5194 (2014).
28. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer’s disease. *Nature* **547**, 185–190 (2017).
29. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627.e16 (2017).
30. Zhang, Z. C. & Chook, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline–tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl. Acad. Sci.* **109**, 12017–12021 (2012).
31. Soniat, M. *et al.* Crystal structure of human Karyopherin β 2 bound to the PY-NLS of *Saccharomyces cerevisiae* Nab2. *J. Struct. Funct. Genomics* **14**, 31–35 (2013).
32. Imasaki, T. *et al.* Structural Basis for Substrate Recognition and Dissociation by Human Transportin 1. *Mol. Cell* **28**, 57–67 (2007).
33. Cansizoglu, A. E., Lee, B. J., Zhang, Z. C., Fontoura, B. M. A. & Chook, Y. M. Structure-based design of a pathway-specific nuclear import inhibitor. *Nat. Struct. Mol. Biol.* **14**, 452–454 (2007).
34. Lee, B. J. *et al.* Rules for Nuclear Localization Sequence Recognition by Karyopherin β 2. *Cell* **126**, 543–558 (2006).

35. Soniat, M. & Chook, Y. M. Karyopherin- β 2 recognition of a PY-NLS variant that lacks the proline-tyrosine motif. *Struct. Lond. Engl. 1993* **24**, 1802–1809 (2016).
36. Muslimov, I. A. *et al.* Interactions of noncanonical motifs with hnRNP A2 promote activity-dependent RNA transport in neurons. *J. Cell Biol.* **205**, 493–510 (2014).
37. Kamma, H. *et al.* Interaction of hnRNP A2/B1 Isoforms with Telomeric ssDNA and the in Vitro Function. *Biochem. Biophys. Res. Commun.* **280**, 625–630 (2001).
38. Nguyen, E. D., Balas, M. M., Griffin, A. M., Roberts, J. T. & Johnson, A. M. Global profiling of hnRNP A2/B1-RNA binding on chromatin highlights LncRNA interactions. *RNA Biol.* **15**, 901–913 (2018).
39. Guerrero-Ferreira, R. *et al.* Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife* <https://elifesciences.org/articles/48907/figures> (2019) doi:10.7554/eLife.48907.
40. Tycko, R. Amyloid Polymorphism: Structural Basis and Neurobiological Relevance. *Neuron* **86**, 632–645 (2015).
41. Li, B. *et al.* Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9**, 1–10 (2018).
42. Babinchak, W. M. *et al.* The role of liquid–liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* **294**, 6306–6317 (2019).
43. Kim, H. J. *et al.* Heterozygous frameshift variants in HNRNPA2B1 cause early-onset oculopharyngeal muscular dystrophy. *Nat. Commun.* **13**, 2306 (2022).

44. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
45. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
46. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
47. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
48. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
49. Wagner, T. *et al.* Two particle-picking procedures for filamentous proteins: SPHIRE-crYOLO filament mode and SPHIRE-STRIPER. *Acta Crystallogr. Sect. Struct. Biol.* **76**, 613–620 (2020).
50. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
51. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
52. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 545–559 (2018).

53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
54. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 531–544 (2018).
55. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
56. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
57. Eisenberg, D., Wesson, M. & Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chemica Scripta*
<https://eurekamag.com/research/032/001/032001963.php> (1989).

Chapter 3

Small molecules disaggregate alpha-synuclein and prevent seeding from patient brain-derived
fibrils

Significance

Parkinson's disease (PD) and the related condition multiple system atrophy (MSA) are associated with the aggregation of the protein alpha-synuclein into fibrils within the brain. Disaggregation of these fibrils could possibly slow or reverse the progress of these diseases. As a step toward testing this hypothesis, we have identified two small molecules that disassemble preformed fibrils of alpha-synuclein in the test tube, including fibrils from postmortem patient brains. These two compounds penetrate the mouse brain and reduce aggregation of alpha-synuclein in the worm *C. elegans*; thus, they may merit development into therapeutics for PD and MSA.

Abstract

The amyloid aggregation of alpha-synuclein within the brain is associated with the pathogenesis of Parkinson's disease (PD) and other related synucleinopathies, including multiple system atrophy (MSA). Alpha-synuclein aggregates are a major therapeutic target for treatment of these diseases. We identify two small molecules capable of disassembling preformed alpha-synuclein fibrils. The compounds, termed CNS-11 and CNS-11g, disaggregate recombinant alpha-synuclein fibrils in vitro, prevent the intracellular seeded aggregation of alpha-synuclein fibrils, and mitigate alpha-synuclein fibril cytotoxicity in neuronal cells. Furthermore, we demonstrate that both compounds disassemble fibrils extracted from MSA patient brains and prevent their intracellular seeding. They also reduce in vivo alpha-synuclein aggregates in *C. elegans*. Both compounds also penetrate brain tissue in mice. A molecular dynamics-based computational model suggests the compounds may exert their disaggregating effects on the N terminus of the fibril core. These compounds appear to be promising therapeutic leads for targeting alpha-synuclein for the treatment of synucleinopathies.

Parkinson's disease (PD), multiple system atrophy (MSA), and dementia with Lewy bodies (DLB) are neurodegenerative disorders characterized by abnormal accumulation of the protein alpha-synuclein (1). Known as synucleinopathies, these diseases are hallmarked by the fibrillar aggregation of alpha-synuclein in either neurons or glial cells (2). Alpha-synuclein aggregation is potentially causative of disease progression as variants in alpha-synuclein that promote aggregation are associated with early-age disease onset and familial forms of PD and DLB (3). In PD, alpha-synuclein aggregation occurs primarily in dopaminergic neurons, while in MSA, aggregation is primarily in oligodendrocytes (4, 5). Natively, alpha-synuclein functions as a vesicle transport protein and more recently found to be involved in P-body and mRNA stability (6, 7). Alpha-synuclein is an intrinsically disordered protein, which has made the determination of an atomic structure of its soluble form challenging. It has been shown that soluble alpha-synuclein adopts a helical form when bound to cell membranes, while its aggregated form adopts the cross-beta structure typical of other amyloid fibrils (8–10). Multiple structures of fibrillar alpha-synuclein have been determined both of recombinant (11–16) and brain-derived fibrils (17, 18).

In addition to primary aggregation, one mechanism by which alpha-synuclein pathology spreads throughout the brain is the prion-like seeding of alpha-synuclein aggregates (19). Alpha-synuclein fibrils can seed the aggregation of soluble native protein, and Lewy body pathology is observed to spread through connected brain regions (19). Aggregation of alpha-synuclein, both primary and seeded, has become a main therapeutic target for synucleinopathies. Antibodies that sequester alpha-synuclein aggregates are currently under development, as well as small molecules that bind monomer and prevent primary aggregation (20). We recently identified several rationally designed peptides and small proteins capable of binding to the growing ends of

alpha-synuclein fibrils and preventing fibril elongation and seeding (21, 22). In this current study, we present small molecules capable of disaggregating preformed alpha-synuclein fibrils. We demonstrate the efficacy of these molecules with in vitro studies, cell culture models, and in vivo models using both recombinant and patient brain-derived alpha-synuclein fibrils. We also present a structure-based model to describe their possible mechanism of fibril disaggregation.

Results

The flavonoid epigallocatechin gallate (EGCG) has been previously demonstrated to disaggregate preformed amyloid fibrils (23, 24). In our recent work, we determined the structure of EGCG in complex with tau paired helical filaments (PHFs) extracted from Alzheimer's disease patient brains (25). From this structure, we identified the pharmacophore of EGCG and proposed EGCG's possible mechanism of disaggregation. Using the EGCG pharmacophore as a docking site, we computationally screened a library of ~60,000 small molecules predicted to have favorable central nervous system penetration based on their biophysical properties (i.e., polar surface area and number of rotatable bonds, etc.). From this screen, we identified compounds able to disassemble tau PHFs, among them, the compound we term CNS-11. The ability of EGCG to disaggregate amyloid fibrils is nonspecific as the compound can act upon multiple different amyloid proteins, including tau, alpha-synuclein, and amyloid-beta (23). Thus, we sought to determine whether CNS-11 specifically disaggregates tau fibrils or whether it affects amyloid fibrils more broadly. We observed that CNS-11 and its chemical analog CNS-11g have a robust effect on alpha-synuclein fibrils. In this current study, we characterize these effects using in vitro, in cellulo, in vivo, and in silico approaches.

CNS-11 and CNS-11g Disaggregate Recombinant Alpha-Synuclein Fibrils In Vitro.

Following an initial observation that CNS-11 may disaggregate alpha-synuclein fibrils, we tested 10 chemical analogs of CNS-11 to assess whether they have similar properties ([SI Appendix, Fig. S1A](#)). CNS-11 has a central amide-like backbone within its structure, and the 10 analogs tested (CNS-11a through CNS-11j) also share this core motif. From our preliminary screen using recombinant alpha-synuclein fibrils, we observe that fibrils treated with the analog CNS-11g have a large reduction in insoluble alpha-synuclein as assessed by western blot ([SI Appendix, Fig. S1B](#)).

We next aimed to better characterize the ability of CNS-11 and CNS-11g to disaggregate alpha-synuclein fibrils in vitro. Both compounds were incubated with equimolar ratios of recombinant alpha-synuclein fibrils (100 μ M:100 μ M) for 48 h. After incubation, thioflavin T (ThT), a fluorescent marker of amyloid fibrils, was added to the samples, and ThT fluorescence was measured. EGCG was used as a positive control for comparison. A reduction in ThT signal is observed for all three compounds tested ([Fig. 1A](#)), indicating a reduction in fibril amount. Next, transmission electron microscopy (EM) was performed for each sample. Representative EM images shown in [Fig. 1B](#) demonstrate a qualitative reduction in fibril count for samples treated with CNS-11 and CNS-11g. To obtain a quantitative evaluation of fibril reduction, fibrils were incubated for 48 and 72 h with compound, then micrographs from random grid points were taken for each sample, and the numbers of fibrils per image were counted ([Fig. 1C](#)). We observe a reduction in the number of imaged fibrils for both compounds, particularly for CNS-11g. Dot blot analysis, staining for alpha-synuclein, was performed for fibril samples incubated with compounds for 48 h ([Fig. 1D](#)). Fibrils treated with EGCG, CNS-11, or CNS-11g were centrifuged to separate insoluble fractions, and a significant reduction in insoluble alpha-synuclein is observed for all compounds. Chemical structures of both CNS-11 (N-mesityl-2-(3-

oxoindeno[1,2,3-de]phthalazin-2(3H)-yl)acetamide) and CNS-11g (2-(4-benzyl-1-oxo-2(1H)-phthalazinyl)-N-(2,6-dimethylphenyl)acetamide) are shown in [Fig. 1E](#), highlighting their similar chemical structures, low number of hydrogen bond acceptors and donors, and low number of rotatable bonds ([SI Appendix, Table S1](#)). These physical and chemical properties are more predictive of oral bioavailability and central nervous system penetration compared to EGCG, making them more viable candidates for further drug development.

To assess whether either CNS-11 or CNS-11g influences primary alpha-synuclein aggregation, we performed a ThT aggregation kinetics assay ([SI Appendix, Fig. S2](#)). Little difference is seen between alpha-synuclein allowed to aggregate in the presence of either CNS-11 or CNS-11g. Preformed alpha-synuclein fibril seeds were added to the alpha-synuclein sample, which resulted in a faster aggregation time. When the fibril seeds were pretreated with CNS-11 or CNS-11g, the increase in aggregation time was partially reduced, presumably mitigating the effects of the preformed fibrils on aggregation time.

CNS-11 and CNS-11g Prevent Intracellular Seeding of Alpha-Synuclein Fibrils and Mitigate Alpha-Synuclein Cytotoxicity.

Having established the ability of CNS-11 and CNS-11g to disaggregate alpha-synuclein fibrils *in vitro*, we next sought to test both compounds in cellular models. It is thought that alpha-synuclein pathology propagates throughout the brain via a prion-like seeding mechanism ([19](#), [26–28](#)). Alpha-synuclein aggregates in one cell can fragment and spread to neighboring cells, where they seed the aggregation of soluble protein. Molecules able to disaggregate preformed fibrils may be able to interrupt this process. To that end, we assessed whether both compounds could mitigate intracellular seeding of alpha-synuclein fibrils using alpha-synuclein biosensor cells—HEK293T cells expressing A53T mutant alpha-synuclein fused with yellow

fluorescent protein (YFP) (29). At baseline, the fluorescently labeled alpha-synuclein remains soluble within the cell visible as diffuse fluorescence. However, the liposome-mediated transduction of exogenous alpha-synuclein fibrils into the cells induces aggregation of the endogenous protein, and the aggregates can be visualized and quantified as fluorescent puncta within the cell.

CNS-11 and CNS-11g at various concentrations were incubated with recombinant alpha-synuclein fibrils for 48 h and then transduced into the biosensor cells. We imaged the cells 48 h after transduction of the samples and quantified the number of fluorescent puncta per sample well. For samples treated with fibrils alone (Fig. 2A), numerous bright puncta can be seen throughout the cells (white arrows). Conversely, without any fibrils added, no puncta can be seen. Incubation of compounds with fibrils at a final compound concentration of 100 nM to 10 μ M resulted in a dose-dependent reduction in aggregates. We used EGCG as a positive control, which also showed a dose-dependent reduction (Fig. 2B–D). For CNS-11, we observed an initial increase in seeding at low compound concentrations and then a sharp reduction in seeding at higher compound concentrations. The mechanism of this is unclear but may be a result of incomplete disaggregation/fragmentation of the fibrils at the low concentration producing a greater amount of seeding, whereas more complete disaggregation is occurring at the higher concentrations.

Alpha-synuclein aggregates, particularly oligomers, are known to be cytotoxic to neurons (4). Thus, we next determined whether disaggregation of alpha-synuclein aggregates by our compounds mitigates their cytotoxicity (Fig. 2E and F). Recombinant alpha-synuclein fibrils were again incubated with various concentrations of CNS-11 and CNS-11g for 48 h. The fibril and compound mixtures were then added to cultured Neuro-2a (N2a) neuronal cells to a final

alpha-synuclein fibril concentration of 1 μM . A 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) dye reduction cell viability assay of the treated N2a cells reveals that CNS-11g recovers the cytotoxicity of the alpha-synuclein fibrils at substoichiometric concentrations. However, CNS-11 showed no rescue of toxicity at any concentration.

Molecular Dynamics (MD) Simulations of CNS-11 and CNS-11g Reveal Possible Mechanism of Alpha-Synuclein Fibril Disaggregation.

Having found that CNS-11 and CNS-11g disaggregate recombinant alpha-synuclein, we sought to determine a putative binding site for CNS-11 and CNS-11g to recombinant alpha-synuclein fibrils and to elucidate the compounds' possible mechanism for fibril disaggregation. To accomplish this, we performed multiple MD simulations of each compound bound to four possible binding sites along the fibril surface, as well as for an unbound control fibril ([SI Appendix, Fig. S3](#)), using the alpha-synuclein recombinant fibril structure (PDB: 6cu7) ([11](#)). MD simulations were carried out for a total of 40 ns. For each simulation, we measured the distance between the top layer of the alpha-synuclein fibril and the layer just beneath it. In most amyloid fibril structures, a distance of $\sim 4.8 \text{ \AA}$ is seen between fibril layers corresponding to beta-strand separation within the beta-sheet. Thus, if fibril disaggregation occurs, we expect the interlayer spacing to increase beyond 4.8 \AA . For each of the four binding sites assessed, we analyzed the interlayer distance between $\text{C}\alpha$ atoms of the residues adjacent to the compound binding site for both CNS-11 and CNS-11g ([SI Appendix, Fig. S3B](#)). For both site 1 and site 2 found at the N terminus of the fibril core, we observe relative stability of the unbound fibril structure at residues 41 to 45, with little fluctuation between interlayer distance. However, with addition of CNS-11 at site 1, the interlayer distances quickly increase ([SI Appendix, Fig. S3C](#)). Similarly, both CNS-11 and CNS-11g greatly disrupt intrastrand spacing when bound to site 2. In contrast, we do not

observe large interstrand distance increases in any of the residues adjacent to site 3 or site 4 ([SI Appendix, Fig. S3C](#)) when bound to CNS-11 or CNS-11g. From these findings, we conclude that a possible mechanism of fibril disaggregation by CNS-11 and CNS-11g is through destabilization of the N-terminal residues (Gly41–Lys45) of the fibril core, either bound to site 1 or site 2, and less likely through the C terminus (site 3 and site 4). While this provides an initial putative mechanism by which these compounds are functioning, further structural and biophysical characterization will be needed to fully understand their effects on fibril architecture.

Compounds Disaggregate MSA Brain-Derived Alpha-Synuclein Fibrils and Prevent Their Seeding.

Having found that CNS-11 and CNS-11g disaggregate recombinant alpha-synuclein fibrils, we then investigated the efficacy of each compound on brain-derived alpha-synuclein fibrils. Postmortem alpha-synuclein fibrils were extracted and purified from the brain of a patient with MSA. The presence of fibrils in the extract was confirmed by EM ([Fig. 3A](#)). The brain-derived fibrils were confirmed to be alpha-synuclein using immunogold labeling ([SI Appendix, Fig. S4](#)). As with the recombinant fibrils, we next aimed to quantitatively assess brain-derived fibril disaggregation by EM. Compounds were incubated with MSA fibrils for 72 h, and fibrils were quantified by EM daily, with 15 images taken per experimental replicate. We observed a reduction in the average number of imaged fibrils for both CNS-11 and CNS-11g compared to control ([Fig. 3B](#)). We also observe a reduction in fibril length. For both CNS-11 and CNS-11g, an ~25% reduction in mean fibril length is observed for treated MSA fibrils ([Fi](#)

Next, we tested the efficacy of the compounds to mitigate the intracellular seeding of MSA fibrils in biosensor cells. MSA brain-derived fibrils robustly seed intracellular alpha-synuclein in HEK293T biosensor cells ([Fig. 3D](#)) ([30](#)). CNS-11, CNS-11g, and EGCG at various

concentrations were incubated with MSA brain-derived fibrils for 48 h, sonicated, and then transduced into cells. Cells were imaged after 48 h, and the number of fluorescent puncta per sample well was quantified ([Fig. 3E](#)). A significant reduction in intracellular seeding is observed for all three compounds at submicromolar concentrations, with both CNS-11 and CNS-11g showing efficacy comparable to the EGCG control. Together, these data demonstrate that these compounds can disaggregate MSA brain-derived fibrils in vitro, and fibrils pretreated with CNS-11 and CNS-11g are no longer competent at seeding the intracellular aggregation of alpha-synuclein.

CNS-11 and CNS-11g Prevent Aggregation of Alpha-Synuclein in *C. elegans*.

Given our findings that both CNS-11 and CNS-11g can reduce alpha-synuclein aggregates in vitro and in cellular models, we next studied their efficacy using an in vivo model of alpha-synuclein pathology. The DDP1 *C. elegans* strain overexpresses alpha-synuclein fused with YFP and cyan fluorescent protein (CFP) ([31](#), [32](#)). FRET-positive fluorescent alpha-synuclein aggregates can then be visualized and quantified in the adult worms. This particular model typically shows only modest phenotypic changes or changes in life span. Thus, aggregate count was used as the primary metric to assess compound efficacy. DDP1 worms were synchronized (*Methods*), and then, the L1 larvae were grown on plates treated with CNS-11 or CNS-11g at 100 μ M. PBS was used as a vehicle control. At day 6 of adulthood, the worms were visualized, and the number of aggregates was quantified by fluorescent microscopy. As shown in [Fig. 4A](#), numerous fluorescent aggregates are visible in the head region of vehicle-treated (“no inhibitor”) worms. Treatment with either compound resulted in a reduction of total aggregates, as quantified in [Fig. 4B](#). Treated worms were also homogenized for protein analysis by western blot. Homogenate for each experimental condition was separated into insoluble pellet fractions (P)

and soluble supernatant (S) fractions. For untreated worm homogenate, the majority of alpha-synuclein is found in the insoluble pellet. Treatment with CNS-11 and CNS-11g each results in solubilization of the alpha-synuclein, presumably from disrupting the formed aggregates ([Fig. 4C](#)).

Detection and Quantification of CNS-11 and CNS-11g in Serum and Brain Tissue in Mice.

Based on their biophysical properties, both CNS-11 and CNS-11g were predicted to have favorable brain permeability. To validate their brain permeability, the compounds were administered by tail vein injection at a dose of 1 mg per kg of body weight to C57BL/6J mice (n = 6 for each compound). One hour after dosing, the mice were killed by cardiac perfusion, and brain and plasma samples were collected. A liquid chromatographic–tandem mass spectrometric multiple reaction monitoring (LC-MS/MS-MRM) assay was used to detect and quantify the drug levels in each tissue sample. The sample extraction protocol for the plasma and brain was optimized with spiking experiments in which the authentic compounds were added to the plasma and brain from drug-naive mice.

One hour following administration, CNS-11 and CNS-11g were measured in the plasma of treated wild-type mice with a range of concentrations of 2.5 to 11.7 ng/mL for CNS-11 and 5.2 to 23.5 ng/mL for CNS-11g ([Fig. 4D](#)). In addition, both CNS-11 and CNS-11g demonstrated brain permeability. They were measured in brain tissue with a range of concentrations of 5.7 to 17.8 ng/g of brain for CNS-11 and 3.5 to 25.5 ng/g of brain for CNS-11g ([Fig. 4E](#)). The limit of detection in plasma was determined to be 3.2 ng/mL for CNS-11 and 1.4 ng/mL for CNS-11g. The limit of detection in brain was determined to be 2.9 ng/g for CNS-11 and 2.6 ng/g for CNS-11g. The amount of plasma collected from two mice treated with CNS-11g was an insufficient volume for analysis by LC-MS/MS-MRM.

Discussion

Current treatment options for synucleinopathies, including PD, MSA, and DLB, are only capable of symptom management; there are no available therapies for any synucleinopathy that modify disease progression. PD is the second most common neurodegenerative disease, and with the aging of our population, the need for therapeutic options is becoming increasingly urgent. Alpha-synuclein was determined to be the primary component of Lewy bodies over 20 y ago (33).

Since that discovery, the aggregation of alpha-synuclein has been identified as the key component in the pathology of these diseases (6). Considerable effort has been invested in the screening and identification of compounds able to either inhibit the formation of or disassemble alpha-synuclein aggregates (20). High-throughput small-molecule screens have identified several promising compounds, including SynuClean-D (34), anle138b (35), and BIOD303 (36).

Recently, the carotenoid crocin was demonstrated to both inhibit aggregation and disassemble mature alpha-synuclein fibrils (37).

As previously mentioned, CNS-11 was originally identified from a screen of compounds capable of disaggregating tau PHFs, the primary tau aggregate in Alzheimer's disease (25). This screen was guided by the structure of EGCG bound to disaggregating tau PHFs. EGCG is a known potent disaggregator of amyloid fibrils, including tau (24) and alpha-synuclein (38).

Unfortunately, its highly polar chemical composition greatly reduces its bioavailability, diminishing its promise as a potential therapeutic (39). Our aim was to identify compounds capable of disaggregating fibrils like EGCG but with more drug-like biophysical properties.

CNS-11 and CNS-11g were identified from a library of compounds with increased probability of blood–brain barrier penetration, a key hurdle in the development of therapeutics for neurodegenerative disease. Both CNS-11 and CNS-11g lack the numerous hydroxyl groups of

EGCG, but have low polar surface areas and molecular weights predictive of central nervous system penetration (40), making them much more promising therapeutic leads than EGCG. Both CNS-11 and CNS-11g also satisfy the Lipinski and Veber rules of drug-like compounds, which EGCG violates (*SI Appendix, Table S1*), and are demonstrated to penetrate brain tissue in mice following tail vein injection (*SI Appendix, Fig. S5*).

Here, we have demonstrated that CNS-11 and CNS-11g can disaggregate alpha-synuclein and that treatment of fibrils with either compound also reduces the seeded aggregation of alpha-synuclein. This has important implications as the spread of alpha-synuclein pathology in the brain is thought to occur through a templated seeding mechanism. A possible pitfall of disaggregation as a therapeutic mechanism is that it may fragment fibrils, producing more seeds that can propagate and worsen disease pathology instead of mitigating it. For example, Nachman et al. recently demonstrated that disassembly of tau fibrils by the chaperone protein Hsp70 can generate seeding-competent species (41). Thus, for the investigation of any new potential therapeutic that modifies alpha-synuclein aggregates, it is essential to assess its effects on seeding. In the case of CNS-11, we do observe an increase in seeding of recombinant alpha-synuclein fibrils (*Fig. 2*) at low compound concentrations, which may represent incomplete disaggregation of the fibrils. This is not seen for CNS-11g, which may indicate that CNS-11g converts the fibrils into a monomeric form or small multimeric form incapable of seeding. We do not observe any enhancement of seeding from MSA brain-extracted fibrils for either compound and instead see a dose-dependent reduction in seeding even at low concentrations. Nonetheless, effects on seeding at a range of doses should be an important consideration during advancement of molecules that target alpha-synuclein, including the compounds presented in this work. We also observe that the toxicity protection by CNS-11g happens at a different concentration range

than the reduction in cellular seeding ([Fig. 2](#)). This is likely a reflection of different experimental conditions between the two cell-based experiments as different cell lines are used (HEK293 vs Neuro-2a cells).

The cell has existing machinery to combat protein aggregation and disassemble pathologic fibrils once they form. Chaperone proteins Hsc70 with DNAJB1 and Apg2 are able to revert fibrillar alpha-synuclein back to soluble monomer and Hsp70 with DNAJB1 ([42](#)). Recent mechanistic studies have revealed that this disassembly of fibrils may occur through the removal of monomer units directly from the fibril ends ([43](#)). Our MD experiments of CNS-11 and CNS-11g bound to the recombinant alpha-synuclein fibril core may reveal a similar mechanism of disassembly ([Fig. 4](#)). First, our simulations of the compounds docked to four potential binding sites reveal areas near the N terminus of the fibril to be the most destabilized by the compounds. The N terminus of alpha-synuclein has been previously demonstrated as having a strong influence on fibrillization ([44](#)). For the simulations that showed fibril disaggregation, we observe that the top layers of the fibrils are the most destabilized even though the simulations are initiated with the compounds bound centrally along the side of the fibril. Thus, CNS-11 and CNS-11g may be acting in a similar way as the chaperone assemblies, peeling off the unstable end layers of the fibrils ([SI Appendix, Fig. S6](#)). Based on our calculations, the mechanism of binding for either compound to the fibril appears to be driven primarily by interactions with charged side chains, as well as hydrogen bonding with backbone amide groups. However, further structural and biochemical studies will be needed to fully deduce their precise mechanism of action. It should also be noted that several dozen different polymorphic structures of alpha-synuclein have been reported in the literature thus far. Common among many of these structures is a conserved protein fold in the core region of the alpha-synuclein fibril present in both the recombinant and

MSA folds. Thus, we sought to use the recombinant fibril structure as a generalized model for our MD simulations as it possesses this conserved core region.

Here, we have demonstrated the ability of two compounds to disassemble alpha-synuclein with both in vitro and in vivo models. We also show the compounds are effective on alpha-synuclein fibrils directly extracted from patient brain tissue and are capable of penetrating living brain tissue in mice. Further validation will be needed to establish their therapeutic efficacy, but these preliminary results demonstrate the potential of these compounds as leads for future drug development toward the treatment of synucleinopathies. Given that the compounds also have effects on tau aggregates and the biophysical properties of a drug-like compound, CNS-11 and its related analogs may also have future promise in treatment of other diseases involving aberrant protein aggregation.

Methods

Expression and Purification of Alpha-Synuclein.

The alpha-synuclein construct was transformed into BL21(DE3) Gold *E. Coli*. Protein was expressed following inoculation of a 30 mL starter with colonies after selection with ampicillin. 6L LB medium was inoculated with the starter culture and grown to an OD600 of 0.6 to 0.8 shaking at 220 rpm at 37 °C. Protein expression was induced by the addition of 500 μM IPTG, and then, cells were grown for an additional 3 h. Cells were centrifuged at 4,000 rpm for 5 min, and cell pellet was resuspended in 100 mL of lysis buffer (100 mM Tris-HCl pH 8.0 and 1 mM EDTA), and the pellet was sonicated on ice to lyse cells. Cell lysate was centrifuged for 30 min at 15,000 rpm. Ammonium sulfate of 0.22 g/mL was added to lysate for 30 min and then centrifuged for 30 min at 15,000 rpm. The pellet was resuspended in 80 mL of 20 mM Tris pH

8.0. Solution was dialyzed in 20 mM Tris pH 8.0 overnight. A HiPrep Q HP column (GE Healthcare) was then used to purify the protein. A gradient 0-100% of buffer A (20 mM Tris pH 8.0) to buffer B (20 mM Tris pH 8.0 and 500 mM NaCl) was used over a volume of 100 mL. Fractions were collected and then ran on a size-exclusion G3000 column (Tosoh Bioscience) using a buffer of 100 mM sodium sulfate, 25 mM sodium phosphate, and 1 mM sodium azide at pH 6.5. The protein was then dialyzed in 100 mM sodium sulfate and 25 mM sodium phosphate overnight, with two exchanges of buffer. Protein was concentrated and flash frozen for storage using liquid nitrogen. Concentration of protein was determined by a Pierce BCA assay (Thermo #23225).

Generation of Recombinant Alpha-Synuclein Fibrils.

First, 50 μ M of recombinant alpha-synuclein in 1 \times PBS was added to several wells of a Nunc black optical bottom plate (Thermo Scientific) to a total volume of 100 μ L per well with a single PTFE bead (0.125 inch diameter) per well. Then, the plate was agitated using a Torrey Pines floor shaker on maximum speed for 96 h at 37 °C. The presence of fibrils within the samples was confirmed using transmission EM (see below) prior to use in experimental assays.

Transmission EM.

For all transmission electron microscopy (TEM), 6 μ L of sample was added onto 400 mesh Formvar/carbon film copper mesh grids (EM Sciences) and then incubated for 4 min. Then, 6 μ L of 2% uranyl acetate solution was used to stain the grids. After 2 min, excess solution was blotted off, and grids were left to dry for a minimum of 30 min. TEM images were taken on a Tecnai 12 transmission electron microscope. For quantitation of fibrils, n = 20 TEM images were

taken per experimental condition from random points throughout the grid. Fibril counts were manually quantified per image in a blinded fashion.

Immunogold Labeling and EM of MSA Brain-Derived Alpha-Synuclein Fibrils.

Carbon/Formvar 400 mesh copper grids were glow discharged at 15 mA for 30 s. MSA brain fibrillar extract diluted 1:6 in PBS was applied to each of two grids for 3 min. After blotting, grids were blocked with 0.1 % gelatin in PBS for 10 min and incubated with a 1:100 dilution of primary antibody, LB509 (Santa Cruz Biotechnology) in 0.1% gelatin in PBS or just 0.1% gelatin in PBS (control) for 1 h. After five washes in 0.1% gelatin in PBS, the secondary antibody, goat anti-mouse IgG gold (6 nm) (Abcam), was applied to the grids for 30 min. After five washes in water, the grids were stained with 2% uranyl acetate and dried. Grids were imaged at 18,500 to 49,000 times magnification on a FEI Tecnai electron microscope operating at 120 kV.

Disaggregation Assays.

For all disaggregation assays, 50 mM stocks of CNS-11 and CNS-11g in DMSO were used to make working concentrations of compound in 1× PBS. Compounds were incubated quiescently with fibril samples (both recombinant and brain derived) for 72 h at 37 °C. For ThT measurements, thioflavin T (ThT) was added to samples at 50 μM, and measurements were taken in a Nunc black 96-well optical bottom plate with a FLUOstar OMEGA plate reader (BMG Labtech).

In Vitro Aggregation Assays.

Thioflavin T (ThT)-based aggregation kinetics assays were performed in Nunc black 96-well optical bottom plates (Thermo Scientific) in a microplate reader taking fluorescence

measurements every 10 min (FLUOstar OMEGA, BMG Labtech). All assays were carried out at 37 °C in 1× PBS buffer and 50 μM ThT at a final well volume of 100 μL; 50 μM full-length αSyn was used. PTFE beads (0.125 inch diameter) were used to agitate the sample, and plates were shaken at 700 rpm with double orbital rotation. For compound-treated samples, 50 μM CNS-11 and CNS-11g dissolved in PBS were used. For fibril-seeded samples, 2.5 μL of fibrils was added to each well (at 50 μM original alpha-synuclein concentration).

Western and Dot Blot Analysis.

Samples were loaded onto a NuPAGE 12% Bis-Tris precast protein gel and ran for 35 min at 200 V. To transfer protein from the gel to a nitrocellulose membrane, an iBLOT2 dry blotting system was used. The membrane was blocked in TBST with 5% milk for 1 h and then washed three times with TBST. The membrane was incubated with the primary antibody (anti-α-synuclein MJFR1 (Abcam, catalog #ab138501); 1:5,000 dilution in 5% milk/TBST solution) for 1 h, washed three times with TBST, incubated with the horseradish peroxidase-conjugated secondary antibody (goat anti-rabbit IgG H + L (Invitrogen; catalog #A27036, lot #2116291); 1:4,000 dilution in 5% milk/TBST), and washed three times in TBST. Signal was detected with a Pierce ECL Plus Western Blotting Substrate (catalog # 32132), and blot was imaged with a Pharos FX Plus Molecular Imager.

For *C. elegans* experiments, the same protocol was used, except following the treatment period worms were harvested, flash frozen, and then homogenized for protein extraction. The membrane was washed with TBST overnight before repeating the western blot analysis for beta-actin (primary antibody anti-beta-actin (C4) (Santa Cruz Biotechnology, catalog #sc-47778, lot #J1119); 1:500 dilution in 5% milk/TBST) and horseradish peroxidase-conjugated secondary

antibody (goat polyclonal anti-mouse IgG (Abcam; catalog #ab205719, lot #GR3271082-2); 1:5,000 dilution in 5% milk/TBST).

For dot blots alpha-synuclein fibrils were incubated with equimolar ratios of compounds (100 μ M:100 μ M) for 48 h at 37 °C in PBS for various lengths of time. Samples were spotted onto the nitrocellulose membrane, and 10 μ L was spotted for each condition, spotting 2 μ L at a time, and allowed to dry in between. The membrane was blocked in TBST with 5% milk for 1 h and then washed three times with TBST. The membrane was then incubated with primary antibody (anti-alpha-synuclein MJFR1 (Abcam, catalog #ab138501); 1:4,000 dilution in 2% milk/TBST solution for 1 h), washed three times with TBST, incubated with the horseradish peroxidase-conjugated secondary antibody (goat anti-rabbit IgG H + L (Invitrogen; catalog #A27036, lot #2116291); 1:5,000 dilution in 2% milk/TBST), and washed three times in TBST. Signal was detected with a Pierce ECL Plus Western Blotting Substrate (catalog # 32132), and the blot was imaged with a Pharos FX Plus Molecular Imager. Densitometry analysis was performed using ImageJ.

3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium Bromide (MTT) Dye Reduction Cell Viability Assay.

Neuro-2a cells (ATCC catalog # CCL-131) were cultured in MEM media (Life Technologies catalog # 11095-080) with 10% FBS (Life Technologies catalog # 10437010) and 1% penicillin-streptomycin (Life Technologies catalog # 15140122) in a 5% CO₂ incubator at 37 °C. N2a cells were plated onto clear 96-well plates (Costar catalog # 3596) at 5,000 cells/well in 90 μ L culture media for 24 h. Recombinant alpha-synuclein fibril samples were coincubated with and without CNS-11 or CNS-11g in 10 μ L volume overnight at 37 °C and then added to the N2a cells (final fibril concentration of 1 μ M). All experiments were performed in triplicate. After incubation for

24 h, 20 μ L thiazolyl blue tetrazolium bromide MTT dye (Sigma; 5 mg/mL stock in DPBS) was added to each well and then incubated for 3.5 h at 37 °C. Removal from the incubator and replacement of well media with 100 μ L of 100% DMSO halted the assay. Absorbance was measured at 570 nm using a SpectraMax M5 reader. A background reading at 700 nm was subtracted from the 570 nm reading. Well readings were normalized to vehicle-alone-treated cells (designated as 100% viable) and cells treated with 100% DMSO (designated as 0% viable).

Biosensor Cell Seeding Assays.

HEK293T biosensor cells stably expressing YFP-fused A53T mutant alpha-synuclein, developed and provided by the laboratory of Marc Diamond at the UTSW, were used. Cells were grown in DMEM (Life Technologies, catalog 11965092) with FBS (10% vol/vol; Life Technologies, catalog A3160401), penicillin/streptomycin (1%; Life Technologies, catalog 15140122), and GlutaMAX (1%; Life Technologies, catalog 35050061) at 37 °C and 5% CO₂ in a humidified incubator. Compounds were incubated with recombinant or patient-derived alpha-synuclein fibrils for 48 h in Opti-MEM media and then applied to ~70% confluent biosensor cells. Prior to adding to cells, the coincubated compound/fibril solution was sonicated for 5 min in a Cup Horn water bath and then mixed for 20 min with Lipofectamine 2000 in Opti-MEM (1:20 dilution) for 20 min. 10 μ L of the inhibitor/fibril + Lipofectamine mixture was added to 90 μ L of cells plated in black 96-well tissue culture plates in triplicate for each concentration of compound tested. The number of seeded aggregates was quantified using a Celigo Image Cytometer (Nexcelom) in the YFP channel. Images were processed in ImageJ; background fluorescence from unseeded cells was subtracted, and the number of particles per image was counted using the particle analyzer function. The quantity of aggregates in each well was normalized to cell confluence. SD between triplicates and a nonlinear regression curve was used to calculate IC₅₀ values for dose-response

curves. High-quality fluorescent images were obtained using a ZEISS Axio Observer D1 fluorescence microscope in the YFP channel.

Extraction of Alpha-Synuclein Fibrils from Patient Brain Tissue.

Extraction of sarkosyl-insoluble alpha-synuclein fibrils from neuropathologically confirmed brain samples of patients diagnosed with MSA was performed using the method previously described by Schweighauser et al. without any modifications ([17](#)). The presence of fibrils in each extract was confirmed by TEM prior to use in experiments. The presence of alpha-synuclein was confirmed by immunoblotting using anti-alpha-synuclein MJFR1 antibody (Abcam, catalog #ab138501).

C. elegans Experiments.

The DDP1 (uonEx1 [unc-54::αSyn::CFP + unc-54::αSyn::YFP(Venus)]) strain was acquired from the Caenorhabditis Genetics Center (CGC) and used for experiments. *C. elegans* were grown and maintained using standard conditions. Worms were synchronized using hypochlorite bleaching, hatched overnight in M9 media (5 g/L NaCl, 6 g/L Na₂HPO₄, 3 g/L KH₂PO₄, and 1 μM MgSO₄) at 17 °C, and then cultured on plates with nematode growth medium (NGM; 17 g/L agar, 2.5 g/L peptone, 3 g/L NaCl, 1 mM CaCl₂, 1 mM MgSO₄, 25 mM KH₂PO₄ pH 6, and 5 μg/mL cholesterol) seeded with OP50 *E. coli*. Strains were maintained at 17 °C. For inhibitor treatment conditions, CNS-11 and CNS-11g were diluted in 1× PBS to a final compound concentration of 100 μM, and the solution was added to plates. Treatment plates were then seeded with heat-treated OP50 (30 min at 65 °C), according to the “NGM dead” method as previously published ([45](#)). Synchronized L1 worms were then added to the treatment plates and grown for 7 d. PBS was used as a control condition. On the third day of growth, 75 μM FUDR

was added. For imaging, worms were mounted onto 5% agar pads on glass slides, immobilized with 1% NaN₃ solution, and imaged by fluorescent microscopy (GFP channel) using a ZEISS Axio Observer D1 fluorescence microscope. Alpha-synuclein aggregates in the head region were quantified.

Animal Experiments.

All animal experiments were approved by the UCLA Animal Research Committee and performed under oversight of the Division of Laboratory Animal Medicine (DLAM). C57BL/6J mice (Jackson Laboratories: JAX:000664) were housed on a 12-h light–dark schedule.

Sample Preparation for MRM Analysis.

Mice were injected intravenously with CNS-11 (n = 6) or CNS-11g (n = 6) at a concentration of 1 mg/kg and euthanized by perfusion 1 h postinjection. Blood was collected by cardiac puncture, and plasma was recovered as the supernatant after centrifugation (3,000 × g, 10 m) and stored frozen. Brains were collected by standard dissection and immediately frozen.

To measure the average drug concentrations across the entire brains, each left hemisphere (average wet weight 250 mg) was processed. The samples were spiked with the internal standard (IS, CNS-11F, 15 pmol in 15 μL of acetonitrile) and homogenized with a probe sonicator (Kontes microsonic cell disrupter, 30 s) after the addition of acetonitrile/water (4:1, v/v, 2,500 μL). The homogenates were divided evenly into three polypropylene microcentrifuge tubes (mcts) for triplicate measurements of each sample. Following centrifugation (16,100 g, 5 min), the supernatants were transferred to clean mcts and dried in a vacuum centrifuge. The pellets were treated with acetonitrile (25 μL) with vigorous mixing, followed by water (25 μL) with more vigorous mixing, and followed by more water (50 μL) for a total volume of 100 μL. The

samples were mixed again, centrifuged (16,100 g, 5 min), and the supernatants were transferred to HPLC injector vials. With each batch of tissue samples, a series of standards were prepared in which CNS-11 and CNS-11g (0, 1, 2.5, 5, and 10 pmol, each in duplicate) were added to drug-naive perfused murine brain tissue (100 mg per tube) along with the same amount of IS (5 pmoles). These samples were processed as described above.

Plasma samples (50 μ L, in duplicate for each mouse) were spiked with the internal standard (IS, CNS-11F, 5 pmol in 5 μ L of acetonitrile) and vigorously mixed after the addition of acetonitrile (500 μ L). Following centrifugation (16,100 g, 5 min), the supernatants were transferred to clean vials and dried in a vacuum centrifuge. The pellets were treated with acetonitrile (25 μ L) with vigorous mixing, followed by water (25 μ L) with more vigorous mixing, and followed by more water (50 μ L) for a total volume of 100 μ L. The samples were mixed again and centrifuged (16,100 g, 5 min), and the supernatants were transferred to HPLC injector vials. With each batch of plasma samples, a series of standards were prepared in which CNS-11 and CNS-11g (0, 1, 2.5, 5, and 10 pmol, each in duplicate) were added to drug-naive murine plasma (50 μ L per tube) along with the same amount of IS (5 pmoles). These samples were processed as described above.

Combined Liquid Chromatography–Tandem Mass Spectrometry with Multiple Reaction Monitoring (MRM).

Aliquots of each sample (2 μ L) were injected onto a reverse-phase HPLC column (Phenomenex Kinetex C18, 2.6 μ m, 100 \times 2.1mm) equilibrated in solvent A (water/formic acid, 100/0.1, v/v) and eluted (100 μ L/min) with a linearly increasing concentration of solvent B (acetonitrile/formic acid, 100/0.1, v/v; min/%B: 0/5, 5/5, 30/100, 32/5, and 40/5). The effluent from the column was directed to an electrospray ion source (Agilent Jet Stream) connected to a triple quadrupole mass spectrometer (Agilent 6460) operating in the positive ion tandem mass

spectrometric multiple reaction monitoring (LC-MS/MS-MRM) mode in which the intensity of the transition of preselected parent ions to preselected fragment ions was recorded after signal optimization (collision energy, fragmentor voltage, and collision cell accelerator voltage) with instrument manufacturer-supplied software (Mass Hunter). Two transitions were monitored for each drug: m/z 398.1 \rightarrow 249.2 (for quantitation) and 277.2 (for confirmation) for CNS-11G (retention time 23.3 min); m/z 396.2 \rightarrow 233.1 (for quantitation) and 261.0 (for confirmation) for CNS-11 (retention time 24.3 min), and one transition was monitored for the IS; m/z 412.3 \rightarrow 249.2 (retention time 24.3 min). Peak areas were integrated and recorded, and a curve was constructed from the data obtained from the standards in which the ratio of drug peak area/IS peak area was plotted against the amount of drug in each sample. The amount of each drug in the samples was then derived by interpolation from the curve.

Computational Models and MD.

Three-dimensional structures of CNS-11 and CNS-11g were generated using Open Babel. Both compounds were docked to four unique locations along the alpha-synuclein fibril structure (PDB code: 6cu7) using AutoDock vina. Top-scoring bound poses were used as starting trajectories for subsequent MD experiments. MD was performed using GROMACS version 2020 using a CHARMM36 force field. Ligand topologies for CNS-11 and CNS-11g were generated using the CGenFF server. Hydrogen atoms were added to all structures using Avogadro. The fibril alone and fibril/compound complex structures were solvated using a 1-nm dodecahedron water box. Chloride ions were added to generate a charge-neutral system. The system was first energy minimized; then, temperature and pressure equilibrated under and NVT ensemble for 100 ps and then NPT for an additional 100 ps. Production MD runs were performed for 40 ns for each

system. Analysis of molecular distances was performed in GROMACS and manually visualized with PyMOL. Images of molecular models were generated with UCSF Chimera.

Acknowledgments

We thank the laboratory of Marc Diamond at the UTSW for providing alpha-synuclein biosensor cells. We thank Duilio Cascio and Alex Lisker of the UCLA DOE Biocomputing Technology Core. We thank the UCLA Pasarow Mass Spectrometry Laboratory for assistance with LC-MS/MS-MRM experiments. We thank Alexander van der Blik for advice on *C. elegans*. K.A.M. and H.P. were supported by the UCLA–Caltech Medical Scientist Training Program (GM08042) and the UCLA Chemistry–Biology Interface Training Grant (US Public Health Service National Research Service Award 5T32GM008496). Funding for this work was provided by the HHMI and NIH. We thank NIH R01 AG070895, RF1 AG048120, and DOE DE-FC02-02ER63421 for support.

Author contributions

K.A.M., K.F.F., H.V.V., and D.S.E. designed research; K.A.M., C.J.H., H.P., J.L., R.A., C.K.W., G.E., and M.B. performed research; K.A.M., J.L., R.A., J.T.B., G.M.R., C.K.W., K.F.F., H.V.V., and P.M.S. contributed new reagents/analytic tools; K.A.M., C.J.H., H.P., J.L., R.A., J.T.B., G.M.R., C.K.W., G.E., K.F.F., H.V.V., P.M.S., and D.S.E. analyzed data; and K.A.M., C.J.H., H.P., G.E., K.F.F., H.V.V., and D.S.E. wrote the paper.

Competing interests

D.S.E. is a SAB chair and equity holder of ADRx, Inc. All other authors declare no conflicts of interest.

Figures

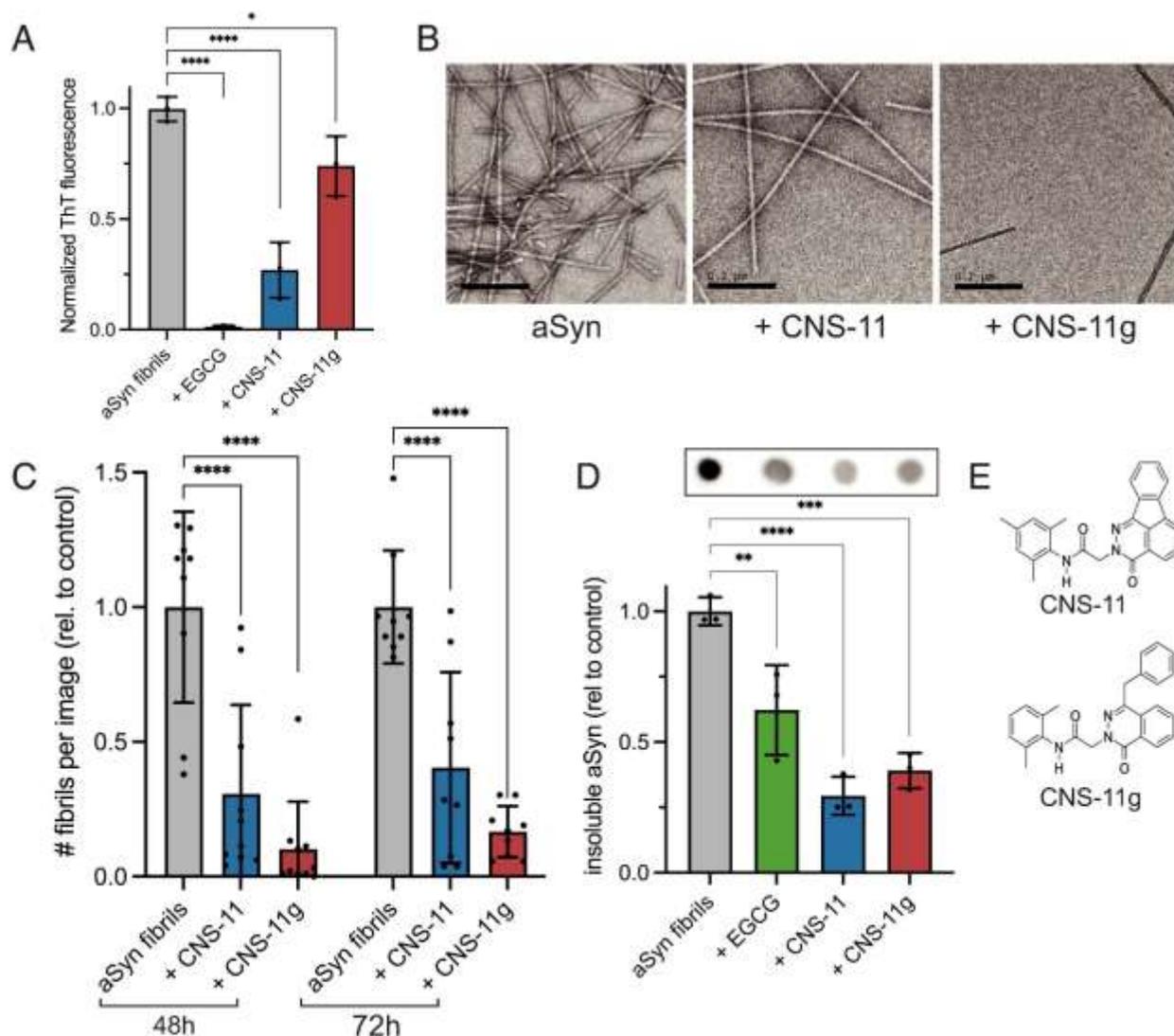


Fig. 1. In vitro characterization of alpha-synuclein fibril disaggregation. (A) Thioflavin T fluorescence of alpha-synuclein fibrils incubated with equimolar EGCG, CNS-11, or CNS-11g for 48 h. A reduction in ThT signal is observed for compound-treated samples, indicating a reduction in alpha-synuclein fibrils. (B) Representative transmission electron micrographs of alpha-synuclein fibrils with CNS-11 and CNS-11g showing a reduction in fibril count after compound treatment. (Scale bar, 200 nm.) (C) Quantification of alpha-synuclein fibrils from TEM images after 48- and 72-h compound incubation. N = 10 images from random regions of TEM grids were quantified for each treatment condition. (D) Alpha-synuclein fibrils treated with CNS-11 and CNS-11g were pelleted into insoluble fractions and then analyzed by dot blot staining for alpha-synuclein. Blots were quantified, showing a reduction in insoluble alpha-synuclein following compound treatment. (E) Chemical structures of CNS-11 and CNS-11g. All error bars represent \pm SD (**P < 0.01, ***P < 0.001, ****P < 0.0001) using a one-way ANOVA with pairwise t test.

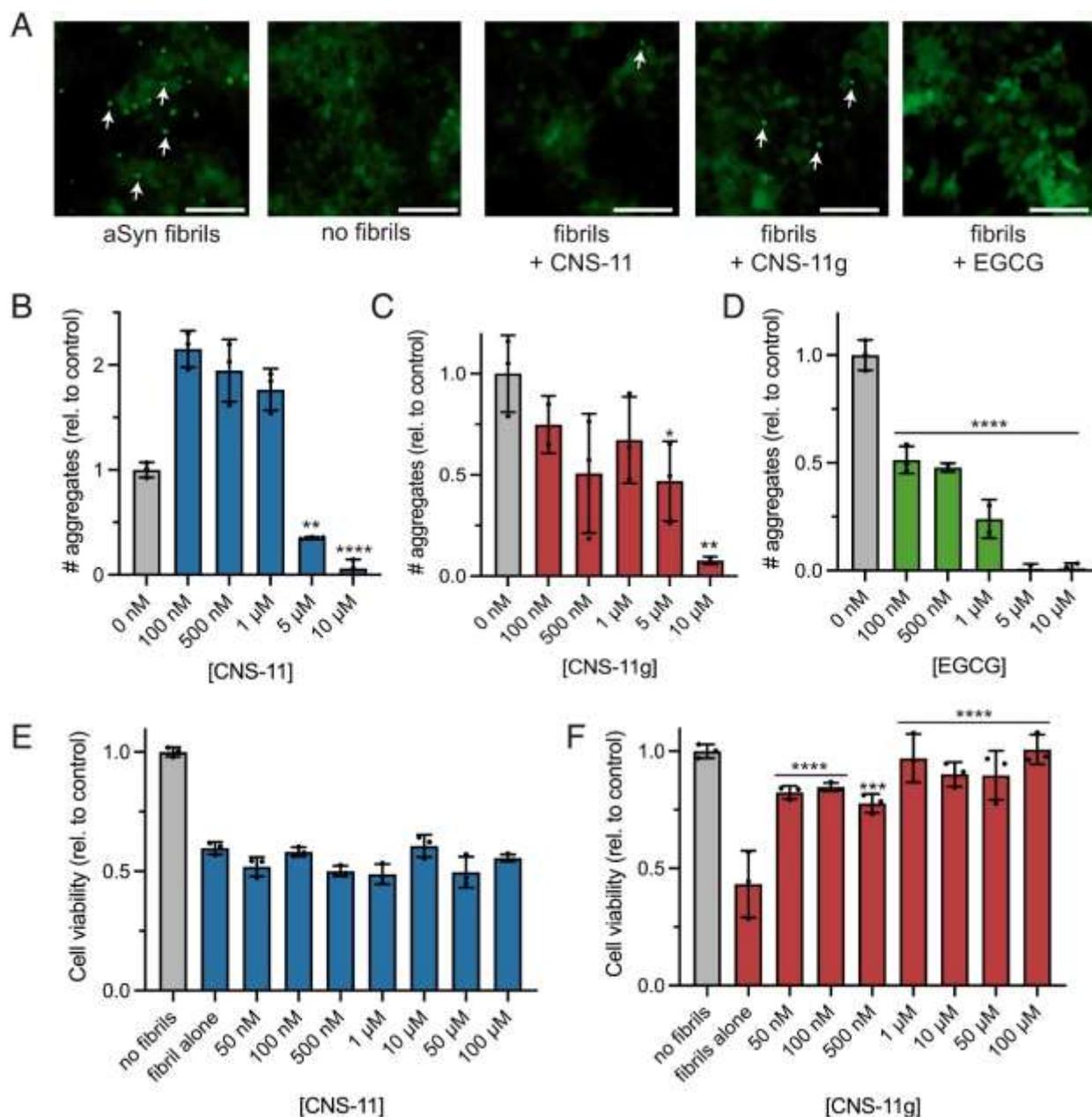


Fig. 2. Inhibition of intracellular seeding in alpha-synuclein biosensor cells and mitigation of alpha-synuclein cytotoxicity by CNS-11 and CNS-11g. (A) Fluorescent microscopy images of alpha-synuclein biosensor cells with and without 10 μ M compound treatment. HEK293T cells expressing YFP-labeled A53T alpha-synuclein, termed biosensor cells, are seeded with exogenous alpha-synuclein fibrils. After seeding, the soluble fluorescent protein is incorporated into intracellular aggregates visible as bright puncta on fluorescence microscopy (“aSyn fibrils,” white arrows). Without addition of fibril seeds, no fluorescent puncta are observed (“no fibrils”). Incubation of CNS-11, CNS-11g, or EGCG with the exogenous fibrils before seeding results in a reduction of visible puncta. (Scale bar, 50 μ M.) (B–D) Quantification of fluorescent puncta from the biosensor cells. Total puncta in each experimental well were quantified from fluorescent microscopy images, which were normalized to total cell confluence. A dose-dependent decrease

in puncta with increasing concentrations of compound pretreatment can be seen for CNS-11 (B), CNS-11g (C), and EGCG (D). “No inhibitor” indicates cells seeded with fibrils not treated with any compound. N = 3 experimental replicates were used for each treatment condition. (E and F) MTT toxicity assay of neuronal cells after treatment with alpha-synuclein fibrils with and without inhibitor. N2a cells were treated with fibrillar alpha-synuclein overnight, resulting in a 40 to 60% reduction in cell viability. CNS-11 was unable to rescue the alpha-synuclein cytotoxicity; however, CNS-11g did show a dose-dependent rescue of cell viability with increasing concentrations of compound. “No fibrils” indicates untreated cells. N = 3 experimental replicates were used for each treatment condition. Error bars represent \pm SD.

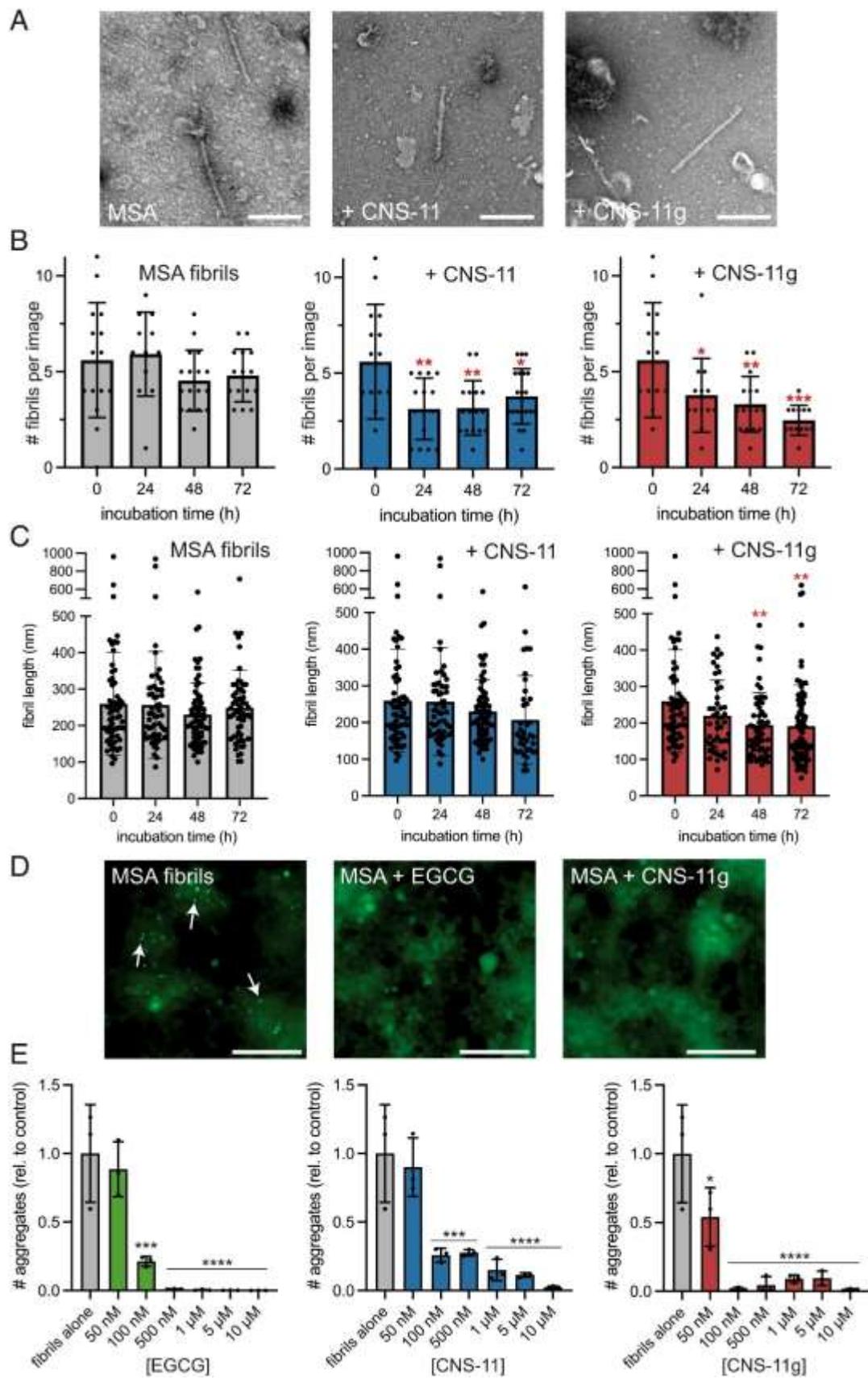


Fig. 3. Effects of CNS-11 and CNS-11g on MSA patient brain-derived alpha-synuclein fibrils. (A) TEM images of alpha-synuclein fibrils extracted from brains of patients with MSA. (B) CNS-11 or CNS-11g was incubated with MSA brain-derived fibrils for up to 72 h. Quantitation of EM images of MSA brain-derived fibrils treated with no compound, CNS-11, or CNS-11g over multiple days. N = 15 images were taken per experimental condition, with each condition performed in triplicate. Average fibril count per image remains relatively stable for the “no inhibitor” control sample, but a reduction in fibrils is seen for both compound-treated samples over the course of 3 d (*P < 0.05, **P < 0.005, ***P < 0.0005). (C) Fibril length analysis of MSA brain-extracted fibrils treated with CNS-11 and CNS-11g. Fibrils were incubated with compounds for 3 d, and fibril length was measured every 24 h using transmission EM. A reduction in average fibril length is observed for both CNS-11 and CNS-11g, statistically significant for CNS-11g at 48 and 72-h incubation time. Bars represent mean fibril length; error bars are \pm SD. N = 15 TEM images were taken for each condition, with two to eight visible fibrils per image (**P < 0.01). (D) Alpha-synuclein biosensor cells seeded with MSA patient-derived alpha-synuclein fibrils. With the addition of the fibril seeds (“+MSA fibrils”), numerous fluorescent puncta are visible (white arrows). Treatment with compound EGCG/CNS-11/CNS-11g greatly reduces the number of puncta visible. (E) Quantification of seeded aggregates from MSA fibril-treated biosensor cells. EGCG, CNS-11, and CNS-11g all show a robust effect on reducing intracellular seeding when preincubated with fibril seeds prior to transduction into the cells. N = 3 experimental replicates were analyzed for each treatment condition. Error bars represent \pm SD (*P < 0.05, ***P < 0.0005, ****P < 0.0001).

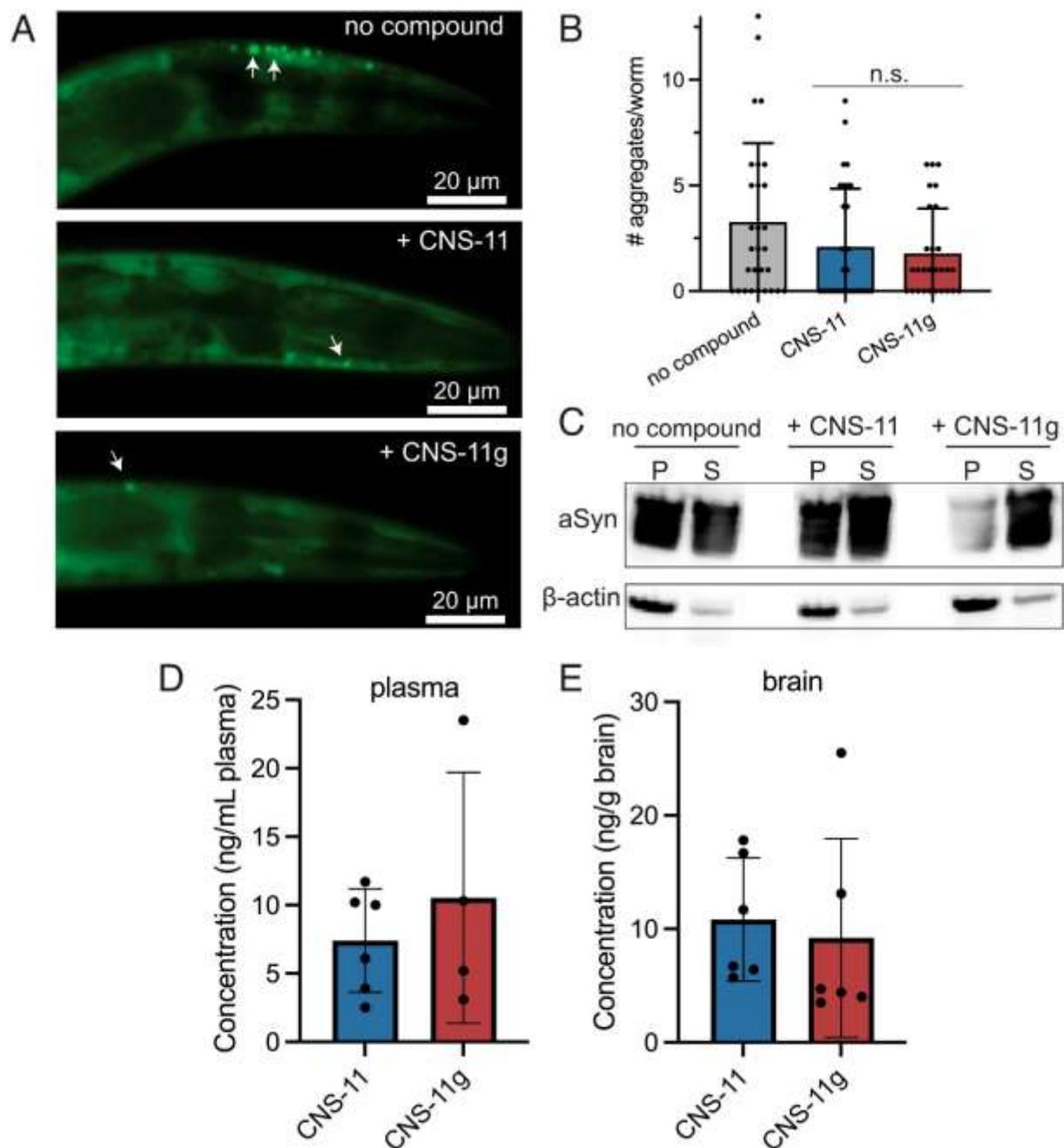


Fig. 4. In vivo effects of CNS-11 and CNS-11g in *C. elegans* and mice. (A) *C. elegans* overexpressing CFP- and YFP-fused alpha-synuclein were treated with CNS-11 and CNS-11g at the L1 larval stage and then imaged by fluorescent microscopy at day 6 adulthood to assess for alpha-synuclein aggregation. Pseudocolored images show numerous punctate aggregates of fluorescent alpha-synuclein in the head region of vehicle-treated control worms (no inhibitor), and a reduction in aggregates is observed for those treated with either CNS-11 or CNS-11g. (B) Quantification of number of aggregates observed in the head region of each treatment group shows a reduction in aggregates for both CNS-11- and CNS-11g-treated worms. (C) Untreated and compound-treated *C. elegans* were homogenized, and western blot analysis of insoluble

alpha-synuclein [found in sample pellet (P)] and soluble alpha-synuclein [found in sample supernatant (S)] was performed. Without treatment, the majority of alpha-synuclein found in the sample homogenate was found in the insoluble (P) fraction. However, with treatment of either compound, particularly CNS-11g, most of the alpha-synuclein is converted to the soluble (S) fraction. β -actin control is shown below. For each treatment condition, n = 30 worms were analyzed for both the microscopy and western blot analysis. Error bars represent \pm SD. (D and E). Brain penetration of CNS-11 and CNS-11g in mice. Mice were injected intravenously with 1 mg/kg of CNS-11 or CNS-11g (n = 6 per compound) and euthanized 1 h after dosing. Compound levels in plasma (D) and brain tissue (E) were analyzed using an LC-MS/MS-MRM method.

References

1. R. A. Barker, C. H. Williams-Gray, Review: The spectrum of clinical features seen with alpha synuclein pathology. *Neuropathol. Appl. Neurobiol.* 42, 6–19 (2016).
2. M. Goedert, Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.* 2, 492–501 (2001).
3. R. M. Meade, D. P. Fairlie, J. M. Mason, Alpha-synuclein structure and Parkinson’s disease – lessons and emerging principles. *Mol. Neurodegener.* 14, 29 (2019).
4. B. Winner *et al.*, In vivo demonstration that α -synuclein oligomers are toxic. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4194–4199 (2011).
5. L. Fellner, K. A. Jellinger, G. K. Wenning, N. Stefanova, Glial dysfunction in the pathogenesis of α -synucleinopathies: Emerging concepts. *Acta Neuropathol.* 121, 675–693 (2011).
6. H. A. Lashuel, C. R. Overk, A. Oueslati, E. Masliah, The many faces of α -synuclein: From structure and toxicity to therapeutic target. *Nat. Rev. Neurosci.* 14, 38–48 (2013).
7. E. Hallaçli *et al.*, The Parkinson’s disease protein alpha-synuclein is a modulator of processing bodies and mRNA stability. *Cell* 185, 2035–2056.e33 (2022).
8. C. C. Jao, A. Der-Sarkissian, J. Chen, R. Langen, Structure of membrane-bound α -synuclein studied by site-directed spin labeling. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8331–8336 (2004).

9. M. Margittai, R. Langen, Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Q. Rev. Biophys.* 41, 265–297 (2008).
10. R. W. Newberry, J. T. Leong, E. D. Chow, M. Kampmann, W. F. DeGrado, Deep mutational scanning reveals the structural basis for α -synuclein activity. *Nat. Chem. Biol.* 16, 653–659 (2020).
11. B. Li *et al.*, Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* 9, 1–10 (2018).
12. R. Guerrero-Ferreira *et al.*, Cryo-EM structure of alpha-synuclein fibrils. *Elife* 7, e36402 (2018).
13. Y. Li *et al.*, Amyloid fibril structure of α -synuclein determined by cryo-electron microscopy. *Cell Res.* 28, 897–903 (2018).
14. D. R. Boyer *et al.*, Structures of fibrils formed by α -synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nat. Struct. Mol. Biol.* 26, 1044–1052 (2019).
15. D. R. Boyer *et al.*, The α -synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *Proc. Natl. Acad. Sci. U.S.A.* 117, 3592–3602 (2020).
16. M. D. Tuttle *et al.*, Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein. *Nat. Struct. Mol. Biol.* 23, 409–415 (2016).
17. M. Schweighauser *et al.*, Structures of α -synuclein filaments from multiple system atrophy. *Nature* 585, 464–469 (2020).

18. Y. Yang *et al.*, Structures of α -synuclein filaments from human brains with Lewy pathology. *Nature* 610, 791–795 (2022), 10.1038/s41586-022-05319-3.
19. K. C. Luk *et al.*, Pathological α -Synuclein transmission initiates parkinson-like neurodegeneration in nontransgenic mice. *Science* 338, 949–953 (2012).
20. C. R. Fields, N. Bengoa-Vergniory, R. Wade-Martins, Targeting alpha-synuclein as a therapy for Parkinson's disease. *Front. Mol. Neurosci.* 12, 299 (2019).
21. S. Sangwan *et al.*, Inhibition of synucleinopathic seeding by rationally designed inhibitors. *Elife* 9, e46775 (2020).
22. K. A. Murray *et al.*, De novo designed protein inhibitors of amyloid aggregation and seeding. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2206240119 (2022).
23. J. Bieschke *et al.*, EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7710–7715 (2010).
24. S. K. Sonawane *et al.*, EGCG impedes human Tau aggregation and interacts with Tau. *Sci. Rep.* 10, 12579 (2020).
25. P. M. Seidler *et al.*, Structure-based discovery of small molecules that disaggregate Alzheimer's disease tissue derived tau fibrils in vitro. *Nat. Commun.* 13, 5451 (2022).
26. M. Masuda-Suzukake *et al.*, Prion-like spreading of pathological α -synuclein in brain. *Brain* 136, 1128–1138 (2013).
27. K. C. Luk *et al.*, Intracerebral inoculation of pathological α -synuclein initiates a rapidly progressive neurodegenerative α -synucleinopathy in mice. *J. Exp. Med.* 209, 975–986 (2012).

28. C. Peng *et al.*, Cellular milieu imparts distinct pathological α -synuclein strains in α -synucleinopathies. *Nature* 557, 558–563 (2018).
29. B. B. Holmes *et al.*, Proteopathic tau seeding predicts tauopathy in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4376–E4385 (2014).
30. S. B. Prusiner *et al.*, Evidence for α -synuclein prions causing multiple system atrophy in humans with parkinsonism. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5308–E5317 (2015).
31. R. Bodhicharla *et al.*, Effects of α -synuclein overexpression in transgenic *Caenorhabditis elegans* strains. *CNS Neurol. Disord. Drug Targets* 11, 965–975 (2012).
32. A. Nagarajan *et al.*, A fluorescence resonance energy transfer assay for monitoring α -synuclein aggregation in a *Caenorhabditis elegans* model for Parkinson's disease. *CNS Neurol. Disord. Drug Targets* 14, 1054–1068 (2015).
33. M. G. Spillantini *et al.*, Alpha-synuclein in Lewy bodies. *Nature* 388, 839–840 (1997).
34. J. Pujols *et al.*, Small molecule inhibits α -synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10481–10486 (2018).
35. J. Wagner *et al.*, Anle138b: A novel oligomer modulator for disease-modifying therapy of neurodegenerative diseases such as prion and Parkinson's disease. *Acta Neuropathol.* 125, 795–813 (2013).

36. B. Moree *et al.*, Small molecules detected by second-harmonic generation modulate the conformation of monomeric α -synuclein and reduce its aggregation in cells*. *J. Biol. Chem.* 290, 27582–27593 (2015).
37. B. Saffari, M. Amininasab, Crocin inhibits the fibrillation of human α -synuclein and disassembles mature fibrils: Experimental findings and mechanistic insights from molecular dynamics simulation. *ACS Chem. Neurosci.* 12, 4037–4057 (2021).
38. J. Bieschke *et al.*, EGCG remodels mature α -synuclein and amyloid- β fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7710–7715 (2010).
39. J. Kanwar, Recent advances on tea polyphenols. *Front. Biosci. Elite Ed.* 4, 111–131 (2012).
40. T. T. Wager, X. Hou, P. R. Verhoest, A. Villalobos, Moving beyond Rules: The development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* 1, 435–449 (2010).
41. E. Nachman *et al.*, Disassembly of Tau fibrils by the human Hsp70 disaggregation machinery generates small seeding-competent species. *J. Biol. Chem.* 295, 9676 (2020).
42. A. S. Wentink *et al.*, Molecular dissection of amyloid disaggregation by human HSP70. *Nature* 587, 483–488 (2020).
43. M. M. Schneider *et al.*, The Hsc70 disaggregation machinery removes monomer units directly from α -synuclein fibril ends. *Nat. Commun.* 12, 5999 (2021).
44. R. P. McGlinchey, X. Ni, J. A. Shadish, J. Jiang, J. C. Lee, The N terminus of α -synuclein dictates fibril formation. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2023487118 (2021).

45. S.-Q. Zheng, A.-J. Ding, G.-P. Li, G.-S. Wu, H.-R. Luo, Drug absorption efficiency in *caenorhabditis elegans* delivered by different methods. *PLoS One* 8, e56877 (2013).

Chapter 4

Atomic structures of TDP-43 LCD segments and Insights into reversible or pathogenic aggregation

Abstract

The normally soluble TAR DNA-binding protein 43 (TDP-43) is found aggregated both in reversible stress granules and in irreversible pathogenic amyloid. In TDP-43, the low-complexity domain (LCD) is believed to be involved in both types of aggregation. To uncover the structural origins of these two modes of β -sheet-rich aggregation, we have determined ten structures of segments of the LCD of human TDP-43. Six of these segments form steric zippers characteristic of the spines of pathogenic amyloid fibrils; four others form LARKS, the labile amyloid-like interactions characteristic of protein hydrogels and proteins found in membraneless organelles, including stress granules. Supporting a hypothetical pathway from reversible to irreversible amyloid aggregation, we found that familial ALS variants of TDP-43 convert LARKS to irreversible aggregates. Our structures suggest how TDP-43 adopts both reversible and irreversible β -sheet aggregates and the role of mutation in the possible transition of reversible to irreversible pathogenic aggregation.

Main

The role of LCDs in proteins linked to neurodegenerative diseases is a topic of intense current investigation^{1,2,3}. LCDs are regions of proteins that exhibit biased compositions of amino acids and frequently no secondary structure^{4,5,6}. In some cases, the LCDs are the result of nucleotide expansions that cause duplications of short sequences, as in Huntington's Disease⁷. In other cases, the LCDs are intrinsic to the protein and believed to have a functional role as in the RNA-binding proteins TAR DNA-binding protein 43 (TDP-43), Fused in sarcoma (FUS) and hnRNPA1 (ref. ⁸), whose LCDs participate in formation of stress granules^{1,9,10}.

Stress granules (SGs) are aggregates of RNA and RNA-binding proteins that form in the cytoplasm when the cell experiences physical, mechanical or chemical stress^{11,12}. This process is often initiated by a halt in translation and puts the cell in a hibernating state until the stress subsides¹³. SGs are reversible, meaning that SG proteins disaggregate when the stress is relieved, the disaggregated proteins and RNA are then free to return to their normal functions in the cell^{1,2,14,15}. Phosphorylation has been shown to play a role in these SG-protein dynamics^{16,17}. Over the last decade, researchers have studied how this process occurs and what biophysical properties these aggregates exhibit. Recent studies on G3BP-1, a SG protein, have demonstrated that the SGs exist in a biphasic state with the mRNA packed at the center and the RNA-binding proteins on the periphery¹⁸. Also, foci composed entirely of RNA have been observed as a product of phase transition in human cells¹⁹. The mechanism of SG assembly and the role of RNA are still being investigated^{1,3,15,20}. Studies on FUS and hnRNPA1 have demonstrated that the LCDs are sufficient for liquid-liquid phase separation (LLPS)^{1,15}. Despite progress, questions remain about the structures of the aggregates and how these may differ between SG assemblies and pathogenic states⁴.

Our recent study determined atomic structures of five segments of low-complexity domains of FUS, hnRNPA1 and nup98 (ref. ²¹). All of these structures displayed kinked β -sheets with weak interactions between sheets. We gave these structures the acronym LARKS (low-complexity aromatic-rich kinked segments) and proposed that these LARKS act as Velcro to provide adhesion between LCDs to form SGs or other membraneless organelle assemblies. These structures contrast with those of steric zippers, which act as molecular glue to fasten amyloidogenic segments in amyloid fibrils (reviewed in ref. ²²). Our study of LARKS offered an atomic-resolution hypothesis about the molecular mechanism of SG formation.

In addition to studies aimed at determining the structure of SG assemblies, researchers have focused on the role of SG-associated proteins in neurodegenerative diseases. A major hallmark of neurodegenerative diseases is the deposition of proteinaceous amyloid fibrils in cells²³. Different from SG assembly, aggregation of proteins into the amyloid state in disease seems a virtually irreversible process^{22:24}. Variants found in TDP-43, FUS and hnRNPA-1 have all been linked to amyotrophic lateral sclerosis (ALS), among other diseases²⁵.

TAR DNA-binding protein (TDP-43) is a 43 kDa protein composed of four domains: an N-terminal domain, two RNA recognition domains and a C-terminal LCD^{26:27:28}. This protein is involved in a number of cellular functions including mRNA shuttling, transcription, miRNA processing and alternative splicing, among others^{29:30}. Over the past decade, research into TDP-43 has found aggregates of the protein in patients with ALS, Alzheimer's disease, Parkinson's disease, frontotemporal lobar degeneration (FTLD) and Huntington's disease^{31:32:33:34}. These aggregates are irreversible, ubiquitinated, hyperphosphorylated, C-terminal cleavage products ranging in size from 25–35 kDa³⁵. Thus, it is hypothesized that the LCD of TDP-43 plays a role in both irreversible disease aggregation and reversible SG aggregation.

Of proteins that participate in SGs and also form pathogenic amyloid fibrils, TDP-43 stands out for four reasons. First, TDP-43 has over 45 variants that are linked to ALS and other neurodegenerative diseases^{25:36:37}. Second, the majority of the variants that cause disease are localized to the LCD^{38:39}. Third, the majority of the aggregates found in the cytoplasm of cells of patients are cleavage products inclusive of the LCD^{40:41}. Finally, TDP-43 is found to aggregate in 97% of sporadic cases and 45% of familial cases of ALS^{42:43}. This makes TDP-43 an intriguing candidate for understanding how its LCD participates in both reversible SG assembly and irreversible, pathogenic assembly.

Through structure determination, we can learn which segments of TDP-43 may drive aggregation and how familial variants affect this process. Here, we use X-ray diffraction and micro-electron diffraction (MicroED) to determine the atomic-resolution structures of ten peptide segments, seven in the reference sequence and three variants, from the LCD of TDP-43. We show that the LCD is capable of forming typical steric-zipper β -sheet structures characteristic of pathogenic aggregation. We also show that segment 312–317 forms a kinked, labile structure—a LARKS, predicted as a protein interaction motif, consistent with properties of SG assemblies²¹. Finally, we show how this structure is perturbed by the familial ALS mutants A315T and A315E, which provides an opportunity for understanding the interplay of the formation of membraneless organelles with pathogenic amyloid.

Results

Six segments from TDP-43 LCD form steric zippers

For structural studies, we targeted segments throughout the LCD because there is no consensus of which region is the amyloidogenic core. The LCD of TDP-43 is an unstructured domain composed of approximately 140 amino acid residues. Researchers have demonstrated that this region plays a role in aberrant mislocalization and aggregation of TDP-43 to the cytoplasm in mice and neuroblastoma cell lines^{44,45}. It was initially suggested that the entire LCD, residues 277–414, served as a prion-like domain and was the minimum fragment required for aggregation⁴⁶. However, additional studies suggested that smaller regions within the LCD are sufficient for aggregation and neurodegeneration. In the first such study, segment 286–331 was identified and shown to form amyloid fibrils and confer neurotoxicity on primary cortical neurons⁴⁷. A second study identified residues 311–360, illustrating through NMR and CD studies that this segment undergoes a helix-to- β -sheet transition that initiates aggregation and

cytoplasmic inclusion^{48,49}. A third study suggested that segment 342–366 is the minimum toxic core that transitions from a random coil to a β -sheet⁵⁰. In short, these studies illustrated that there are several regions of the LCD that can be responsible for amyloid-like aggregation.

Our previous studies demonstrated that adhesive cores of amyloid fibrils are composed of 6–11-residue segments that stack together as steric zippers^{22,51}. To identify which segments from the LCD of TDP-43 might be responsible for TDP-43 aggregation, we consulted ZipperDB⁵² (<https://services.mbi.ucla.edu/zipperdb/>). We selected the top ten hits scored by their adhesive energies (Supplementary Table 1). Additionally, we also selected segments rich in glutamine and asparagine, because the C-terminus of TDP-43 is a prion-like sequence, and these two residues are important for prion aggregation^{53,54}. Lastly, we selected two segments based on a double mutation reported to reduce TDP-43 aggregation⁴⁸ (detailed segment selection in [Methods](#)). In total, we targeted 15 segments in the LCD of TDP-43 (Supplementary Table 1).

Crystal structures revealed that six of the segments formed steric

zippers: ³⁰⁰GNNQGSN³⁰⁶, ³²¹AMMAAA³²⁶, ³²⁸AALQSS³³³, ³³³SWGMMGMLASQ³⁴³, ³⁷⁰GNNSYS³⁷⁵, and ³⁹⁶GFNGGFG⁴⁰² (Table 1, Table 2 and Fig. 1a,b). These structures demonstrated tight side chain interdigitation, similar to previously reported zippers in other amyloidogenic proteins, such as β -amyloid⁵¹ and Tau⁵⁵. All six structures are composed of in-register sheets²².

Segments ³²¹AMMAAA³²⁶ and ³³³SWGMMGMLASQ³⁴³ form antiparallel sheets (Supplementary Fig. 1), whereas the other four segments form parallel sheets.

Segments ³⁰⁰GNNQGSN³⁰⁶, ³²⁸AALQSS³³³ and ³⁷⁰GNNSYS³⁷⁵ all form class 1 steric zippers in which the β -sheets are mated face to face with the same edges of the sheet oriented up the fibril^{22,51} (Fig. 1b). These class 1 structures have two distinct interfaces (labeled as IF1 and IF2 in Fig. 1b). Segments ³²¹AMMAAA³²⁶ and ³³³SWGMMGMLASQ³⁴³ are class 7 steric zippers

with face-to-back and up-up orientation. Segment ³⁹⁶GFNGGFG⁴⁰² forms a class 4 steric zipper with face-to-back β -sheets and up-down orientation⁵¹ (Fig. [1b](#)). These class 7 and class 4 structures have one unique interface. The area buried (Ab) and shape complementary (Sc) of all interfaces are listed in Fig. [1b](#) (ref. [51](#)).

Six zipper segments are important for pathogenic TDP-43 aggregation

To investigate the connection between our steric-zipper structures and the pathogenic aggregates of TDP-43, we designed mutations to disrupt the zipper structures and tested their influence on aggregation of the TDP-43 protein. The segment we chose to represent pathogenic TDP-43 aggregation contains residues 208-414, because it is a pathogenic segment of TDP-43 that was identified from FTLD brains⁴⁰. We term this segment TDP-CTF, where CTF stands for C-terminal fragment. We disrupted steric-zipper formation by substituting tryptophan at the central residue of the sheet-sheet interface; its bulk imposes steric hindrance. We designed eight mutations, Q303W and G304W targeting interface 2 and 1 of ³⁰⁰GNNQGSN³⁰⁶, respectively; A324W targeting the interface of ³²¹AMMAAA³²⁶; L330W and Q331W targeting interfaces 1 and 2 of ³²⁸AALQSS³³³, respectively; M337W targeting the interface of ³³³SWGMMGMLASQ³⁴³; S373W targeting interface 2 of ³⁷⁰GNNSYS³⁷⁵; and G399W targeting the interface of ³⁹⁶GFNGGFG⁴⁰². We did not design a mutation in interface 1 of ³⁷⁰GNNSYS³⁷⁵, because the interaction is relatively weak, as judged by its small area buried.

Our mutagenesis results indicate that all six short segments are important for TDP-43-CTF aggregation. We achieved an aggregate-free starting point by conjugating TDP-CTF wild type and its mutants with a SUMO tag (Fig. [2a](#), top). Aggregation was initiated by ULP1 protease removal of the SUMO tag (Supplementary Fig. [2](#)). After 30 min of incubation on ice, approximately half of the SUMO-free TDP-CTF aggregated, as judged by the pellet fraction

after centrifugation (Fig. [2a](#), middle). In contrast, most of the SUMO-free mutants remained soluble, except G304W. After overnight incubation, cleavage of the SUMO-tag was nearly complete, and all SUMO-free TDP-CTF wild type was in the pellet (Fig. [2a](#), bottom). However, aggregation was incomplete for several mutants, displaying soluble TDP-CTF mutant material in the supernatant, including Q330W, A324W, M337W, L330W, Q331W and G399W. These results indicate that interface 2 of ³⁰⁰GNNQGSN³⁰⁶, the interface of ³²¹AMMAAA³²⁶, interfaces 1 and 2 of ³²⁸AALQSS³³³, the interface of ³³³SWGMMGMLASQ³⁴³, interface 2 of ³⁷⁰GNNSYS³⁷⁵ and the interface of ³⁹⁶GFNGGFG⁴⁰² are important for pathogenic aggregation of TDP-43. Potentially, these segments could facilitate pathogenic aggregation of wild-type TDP-43 by forming the same steric-zipper structures we report here. Three of these zippers are located within the region 318–343, previously identified as the amyloidogenic core essential for TDP-43 aggregation⁴⁸. This correlation is important, given that it covers a relatively small region compared to the whole LCD (26 residues out of 140). Moreover, a previous discovery that the double mutation A324E M337E reduces TDP-43 aggregation⁴⁸ is consistent with our expectation that these substitutions create steric hindrance and electrostatic repulsion, thereby disrupting steric-zipper interfaces of ³²¹AMMAAA³²⁶ and ³³³SWGMMGMLASQ³⁴³. These results suggest that these segments are important for pathogenic aggregation of TDP-43, and the steric-zipper structures we report here may represent the structure of the aggregation core.

To further validate the role of these segments in TDP-43 aggregation, we compared the fibril diffraction of the LCD of TDP-43 with the simulated diffraction of the zipper structures. We acquired fibril diffraction from TDP-LCD, a construct that contains residues 274–414. It was expressed using the same SUMO-tag strategy employed previously (Supplementary Fig. [2](#)). TDP-LCD formed long extended amyloid-like fibrils around 15 nm in diameter (Fig. [2b](#)), similar

to the fibrils formed by TDP-43 in patients with ALS⁴¹. The fibril diffraction pattern shows rings at 10 Å and 4.8 Å, the signature feature of amyloid fibrils²². In comparison, electron micrographs of the six zipper-forming segments reveal microcrystals (Fig. [2b](#)). We have shown that these crystals are composed of cross-β assemblies. That is, the architecture revealed by the crystal structures of the segments is consistent with the cross-β signature of the TDP-LCD. Simulated fibril diffraction patterns calculated from their crystal structures are comparable with experimental diffraction of the TDP-LCD fibril (Fig. [2b](#)); the differences correspond to variation in sheet-to-sheet distance among segments.

A kinked structure disrupted by familial ALS mutations A315T and A315E

In addition to pathogenic aggregation, TDP-43 is also involved in functional aggregation, such as SG formation. In SG assembly, the LCDs of SG proteins are reported to be the key regions responsible for this reversible aggregation^{1,2,3}. To further explore the role of the LCD in the functional aggregation of TDP-43, we focused on segments that have the potential to contribute to SG formation. The McKnight laboratory previously discovered that glycine residues flanked by aromatic side chains are features of hydrogel formation of the LCD of the protein FUS². The hydrogel they studied is a reversible aggregate formed in vitro, composed of reversible amyloid fibrils that trap a large amount of water, and is potentially relevant to SG formation in vivo⁵⁶. The authors showed that replacement of aromatic residues in the LCD disrupted this hydrogel formation², which led to the hypothesis that aromatic residues play a role in SG assembly. For this reason, our recent study focused on the segments of the LCD of SG proteins that contain an aromatic residue and two proximal small or polar residues²¹. Structural studies of these segments in FUS and hnRNPA1 revealed pairs of kinked sheets, defined as LARKS, that display

weak van der Waals interactions between mating sheets. Further research suggests the connection of LARKS with protein adhesion of LCDs in reversible aggregation formation²¹.

In the present study, we find that ³¹²NFGAFS³¹⁷, in the LCD of TDP-43, has similar amino acid composition and also forms labile aggregates and LARKS (Fig. [3](#)). This finding suggests that TDP-43 can adopt the same mechanism of SG formation previously hypothesized for FUS and hnRNPA1. Under EM, ³¹²NFGAFS³¹⁷ forms microcrystals (Fig. [3a](#)). These microcrystals are labile, meaning they fully dissolve when treated with ≥1% SDS and heated to 70 °C for 15 min. Their lability was observed through diminished turbidity (Fig. [3b](#)). This behavior is similar to that of hydrogels formed by LCDs of SG proteins, which resolubilize by treatment with SDS and heat², unlike pathogenic amyloid fibrils, such as β-amyloid fibrils (Supplementary Fig. [3a](#)). The structure of ³¹²NFGAFS³¹⁷ reveals a kinked backbone (Fig. [3c](#)), similar to other LARKS²¹. The kink occurs at Phe316 and has Ramachandran angles of a left-handed α-helix. The kinked backbone makes two phenylalanine residues at positions 2 and 5 face the same side of the sheet, whereas a canonical β-sheet places the odd-numbered side chains on one face and the even-numbered side chains on the opposite face⁵¹. The subsequent packing of mating strands creates two different interfaces. Here, we see that all the aromatic residues pack near to each other, creating a hydrophobic, aromatic interface. The other interface is held together by the hydrogen bonds between asparagine side chains. Both interfaces are weaker than those of most zipper structures, as indicated by the relative low surface area buried, and consistent with the observation that these assemblies are labile under treatment with SDS and heat.

We observed that ³¹²NFGAFS³¹⁷ forms a LARKS structure and that its microcrystalline aggregates are labile, thus providing another example of LARKS with a potential role in SG formation and an opportunity to investigate the influence of pathogenic variants of the putative

key segment of functional aggregation. Within the NFGAFS segment, there are two ALS-linked familial variants: A315T and A315E (refs [47:57](#)). Both of these variants have been shown to increase redistribution of TDP-43 to the cytoplasm and increase neuronal death, as compared to the reference sequence [58:59:60:61](#). Additionally, previous studies have shown that disease-causing mutations in SG proteins are associated with accumulation of persistent SGs [62:63:64](#). However, the molecular mechanism of transformation of SGs to irreversible aggregates has remained elusive. Here, we studied the structure and biochemical behavior of $^{312}\text{NFGTFS}^{31}$, $^{312}\text{NFGFEFS}^{317}$ and $^{312}\text{NFGpTFS}^{317}$ (pT, phosphorylated threonine). We studied $^{312}\text{NFGpTFS}^{317}$, because TDP-43 is found to be hyperphosphorylated in patients with ALS [35](#), and the A315T mutation may create a new phosphorylation site in vivo.

All three of these mutant segments formed microcrystals under EM (Fig. [3a](#)). $^{312}\text{NFGFEFS}^{317}$ and $^{312}\text{NFGpTFS}^{317}$ aggregates were irreversible under the same SDS and heat denaturing assay as $^{312}\text{NFGAFS}^{317}$. $^{312}\text{NFGTFS}^{317}$ aggregates, however, were still labile (Fig. [3b](#)). These results indicated that the stabilities of the mutant segments A315E and A315pT were significantly enhanced as compared to that of $^{312}\text{NFGAFS}^{317}$.

To better understand the biochemical findings, we determined the structures of all three mutant segments. All three structures are composed of kinked, in-register, parallel sheets; however, the kink positions differ. In $^{312}\text{NFGAFS}^{317}$, the kink occurs at Phe316. In the case of $^{312}\text{NFGTFS}^{317}$ and $^{312}\text{NFGFEFS}^{317}$, the kink occurs at Gly314; and the $^{312}\text{NFGpTFS}^{317}$ structure is kinked at two residues, Phe313 and Gly314. Despite the different positions of the kinks, all of the kinks result in the two phenylalanine residues facing the same side of the sheet (Fig. [3c](#)).

Further analysis of the structures illustrates that the packing of sheets in both $^{312}\text{NFGTFS}^{317}$ and $^{312}\text{NFGpTFS}^{317}$ are face to face, as in $^{312}\text{NFGAFS}^{317}$. Both have two

distinct interfaces of mating strands, one dry interface formed by hydrophobic residues (mainly phenylalanine residues, labeled as IF1 in Fig. 3c) and one wet interface formed by mostly hydrophilic residues (labeled as IF2 in Fig. 3c). The packing of $^{312}\text{NFGGEFS}^{317}$ is face to back, so its two interfaces are identical, with hydrophobic interactions in the middle and hydrophilic interactions on both sides. For the $^{312}\text{NFGpTFS}^{317}$ structure, the phosphate is clearly observed in the $F_o - F_c$ omit map (Supplementary Fig. 3b). To our knowledge, this is the first time that a post-translational modification has been reported for an amyloid mimetic structure. The enhanced stability of the A315E and A315pT of $^{312}\text{NFGAFS}^{317}$ assemblies can be explained by the increased interface strengths, indicated by area buried for the dry interface. The area buried for the dry interface increased significantly from 63.0 \AA^2 per chain in $^{312}\text{NFGAFS}^{317}$ to 137.0 \AA^2 per chain in $^{312}\text{NFGGEFS}^{317}$ and 175.6 \AA^2 per chain in $^{312}\text{NFGpTFS}^{317}$ (Fig. 3c). In the case of $^{312}\text{NFGpTFS}^{317}$, increased buried area results from the four phenylalanine residues of two strands perfectly interdigitating with each other. Additionally, on the wet interface, the side chains of Glu315 and pThr315 create a hydrogen bond network, which further strengthens the assembly (Supplementary Fig. 3c). This stacking of negatively charged Glu315 and pThr315 is facilitated by the acidic pH of crystallization conditions, in which the residual TFA from peptide synthesis significantly decreases the pH and results in partially protonated glutamate and P-threonine side chains (Discussion). For $^{312}\text{NFGTFS}^{317}$, the area buried for dry interface and hydrogen bond network for the wet interface is similar that of $^{312}\text{NFGAFS}^{317}$, which is consistent with the observation that $^{312}\text{NFGTFS}^{317}$ has similar behavior in our denaturing assay as $^{312}\text{NFGAFS}^{317}$.

In summary, we find that the $^{312}\text{NFGAFS}^{317}$ segment of TDP-43 forms a kinked β -sheet structure—a LARKS—which is labile. We find that two familial variants within this segment,

A315T and A315E, together with phosphorylation, strengthen the reference sequence assembly, making it irreversible by creating a more stable structure with stronger interaction between each of the sheets. These results further support our hypothesis that LARKS are the potential molecular mechanism for functional, reversible aggregate formation, by showing that LARKS exist in another SG-forming protein, TDP-43. Moreover, these results indicate that the pathogenic mutations and or post-translational modifications can strengthen LARKS and drive SG proteins toward pathogenic aggregation.

Discussion

Our previous studies suggested connections of cross- β structures to both amyloid aggregation through steric zippers forming irreversible pathogenic fibrils²² and LARKS forming reversible hydrogels and possibly membraneless organelles²¹. In this study, we determined structures of segments from the LCD of TDP-43 and found that this domain has the capacity to form both steric zippers and LARKS. This finding is consistent with the dual nature of the TDP-43 LCD, which can drive aggregation of the protein into either functional SG aggregates or into pathogenic amyloid fibrils, as in ALS or FTLD.

We present a speculative model (Fig. 4) that incorporates our findings and explains how SG assembly might lead to pathogenic assembly. In physiological conditions, TDP-43 can execute its normal functions without irreversible aggregation, even though steric-zipper (red bars) and LARKS (kinked blue bars) segments exist in their sequences (Fig. 4, top). In pathogenic conditions, the steric-zipper-forming segments interact with each other to form irreversible assemblies, driving the protein into pathogenic amyloid fibrils (Fig. 4, left pathway). In stress conditions, the LARKS-forming segments interact with each other to form reversible assemblies and drive the protein to functional aggregates such as SGs. Because functional assemblies are

reversible^{1:2:14}, the aggregation of proteins can be reversed when stress is relieved (Fig. 4, right pathway). Numerous factors are believed to influence the fate of the proteins, including intra- and intercellular signals, chaperones, RNA binding, protein folding, microenvironments such as pH, protein concentration, etc. Here, our study suggests that the mutations and/or post-translational modifications can also change the pathway of assembly by directly changing the character of assembly-driving segments (Fig. 4, middle pathway). As exemplified by A315E and A315pT, the mutation and phosphorylation strengthen the ³¹²NFGAFS³¹⁷ assembly, converting a labile LARKS to irreversible aggregates. Our hypothesis is that under conditions that favor SGs, LARKS segments interact with each other, and should one or more of them convert to an irreversible structure, the protein may form a pathological, irreversible assembly.

The mechanism represented in Fig. 4 may apply for other regions of TDP-43 and other SG-forming proteins. The ³¹²NFGAFS³¹⁷ segment is not the only region proposed to be important for reversible aggregation of TDP-43; residues 320–340 have been reported to be involved in granule formation^{3:65}. Applying a threading algorithm for predicting LARKS from amino acid sequence, we predicted additional LARKS in the LCD of TDP-43 (ref. 21). We compared the LARKS predictions with the familial variants reported on TDP-43 LCD and found that 28 out of 42 reported familial variants are in the regions of predicted LARKS (Supplementary table 2). This observation supports our hypothesis that the mechanism of conversion of A315E and A315pT on ³¹²NFGAFS³¹⁷ may be a general mechanism.

We note that both ³¹²NFGEFS³¹⁷ and ³¹²NFGpTFS³¹⁷ crystals were grown at acidic pH, which facilitated the stacking of glutamate and phosphorylated threonine residues. A recent study also shows that decrease in pH drives phase separation in yeast and in in vitro systems⁶⁶. This

observation suggests the importance of pH on SG formation and the role of pH in transition of SG to pathogenic aggregation.

In this study, we also report six steric-zipper structures from LCD of TDP-43 and show the connection between these structures with pathogenic aggregates of TDP-43 by mutagenesis experiments. The observation that all six segments are important for TDP-43 aggregation suggests a segmental polymorphism, meaning the aggregation of TDP-43 can be formed via multiple cores. Although recent structural studies have shown that the amyloid fibril core usually contains tens of residues (e.g., cryo-EM structures of Tau⁶⁷), our study on short peptides facilitates the research of amyloidogenic proteins by identifying the high-resolution structures of core aggregation-prone segments and, more importantly, by providing structural insights for inhibitor designs. By designing peptide inhibitors that bind to the ends of steric-zipper structures and cap their growth, we have successfully developed the inhibitors for aggregation of several amyloidogenic proteins, including p53 (ref. ⁶⁸) and Tau⁵⁵. The zipper structures we reported here will provide the opportunity for structure-based design of TDP-43 inhibitors.

In summary, the present study provides atomic-resolution structures of irreversible and reversible aggregates of segments of TDP-43 and the possible roles of familial mutations and/or post-translational modifications in the interconversion of these two distinct aggregation pathways.

Methods

Selection and generation of TDP-43 peptide targets

We chose our peptides of interest by a combination of computational prediction and literature reports (Supplementary Table ¹). For computational prediction, we selected the segments that contain high-rank predictions from ZipperDB (ranked by the predicted Rosetta energy). From

literature reports, we selected the segments containing the features important for LCD aggregation. First, the LCD of TDP-43 is believed to be a prion-like domain, and glutamine and asparagine residues are found to be important for prion aggregation^{53,54}, so we selected three segments from LCD of TDP-43 that have the highest glutamine and asparagine density. These three segments are GNNQGSN (4 out of 7), ASQQNQ (4 out of 6) and GNNQNNQ (5 out of 6); the latter two were also identified by ZipperDB. Second, aromatic residues are reported to be important in LCD aggregation², so we selected two segments from the LCD of TDP-43 that have multiple aromatic residues close to each other, NFGAFS (two residues between two phenylalanine residues) and GFNGGFG (three residues between two phenylalanine residues). Third, the A324E M337E double mutation is reported to reduce TDP-43 aggregation⁴⁸, so AMMAAA and SWGMMGMLASQ were selected to study the molecular mechanism of this double mutation. All peptides were purchased from Genscript at a purity of 95% or higher. Notice that from 15 peptides we selected, we determined the structures for ten of them. The other segments we selected (Supplementary Table 1) failed to generate structures because of poor crystal formation or poor quality of crystal diffraction.

Peptide aggregates formation

All peptide segments were dissolved in PBS at pH 7.5 to a final concentration of 20 mM. Samples were then incubated at 37 °C in a Torrey Pine Scientific shaker at level 9 for 80 h. Samples were examined by TEM following fibril formation.

Construct design

TDP-CTF (208–414) was designed to mimic the pathological fragment of TDP-43 found in FTLD brain, and the LCD of TDP-43 (TDP-LCD, 274-414) was also designed. Both TDP-43

fragments were conjugated to the carboxyl terminus of SUMO proteins to increase solubility and prevent aggregation during expression and purification. The SUMO protein was also conjugated with (His)₆-tag on its amino terminus for easier purification. The SUMO-TDP constructs were inserted into a pET28a vector with following expressed sequences:

SUMO-TDP-CTF:

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPLRRLMEAFAKRQ GKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
REFFSQYGDVMDVFIPKPFRAFAFVTFADDQIAQSLCGEDLIKGISVHISNAEPKHNSNR
QLERSGRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMA
AAQAALQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSG
AAIGWGSASNAGSGSGFNNGGFGSSMDSKSSGWGM

SUMO-TDP-LCD:

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPLRRLMEAFAKRQ GKEMDSLRF LYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
GRFGGNPGGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMAAAQAA
LQSSWGMMGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSGAAIGW
GSASNAGSGSGFNNGGFGSSMDSKSSGWGM

Protein purification and validation

Both SUMO-TDP-CTF and SUMO-TDP-LCD protein were expressed in *Escherichia coli* BL21 (DE3) strain. Bacterial cells were cultured in LB media supplemented with 50 µg/ml kanamycin at 37 °C to an OD₆₀₀ of 0.6–0.8. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 25 °C for 3 h. Cells were harvested and resuspended in 20 mM

Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, supplemented with 1% (v/v) Halt Protease Inhibitor single-use cocktail (Thermo Scientific). The resuspension solution was sonicated (3 s on/3 s off cycle, 10 min), and the supernatant solution was collected after centrifugation (24,000 *g* for 20 min). The supernatant solution was mixed with homemade NucA nuclease (5000 U per liter of cell culture) and filtered before load to HisTrap HP column (GE health care). The HisTrap column was equalized with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 20 mM imidazole before loading of the supernatant solution and washed with 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 200 mM imidazole after loading. The protein expected was eluted by 20 mM Tris-HCl, pH 8.0, 500 mM NaCl and 500 mM imidazole and concentrated by Amicon Ultra-15 centrifugal filters (Millipore) to at least 2 mg/ml (10-fold higher than the working concentration of aggregation assays). SUMO-TDP-LCD was concentrated to 10 mg/ml (10-fold higher than its fibril formation concentration).

To validate the cleavage of SUMO-tag and existence of TDP fragment, purified SUMO-TDP-CTF and SUMO-TDP-LCD were diluted into 0.2 mg/ml with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μ M DTT, mixed with 100:1 (weight basis) homemade ULP1 protease and incubated on ice. Samples were taken before and 1 h after adding ULP1 protease and separated by NuPAGE 4-12% Bis-Tris gel (Invitrogen). The pellet of SUMO-TDP-LCD was also collected and sampled from fibril formation assay and loaded to the same gel. The SDS-PAGE was stained by Coomassie blue or probed with TDP-43 polyclonal antibody (Invitrogen, catalog number PA5-27221, 1:1,000 dilution) and Anti-Rabbit IgG (whole molecule)-peroxidase antibody (Sigma, catalog number A0545, 1:5,000 dilution) after transfer to a nitrocellulose membrane (Invitrogen). The membrane was visualized with SuperSignal West Pico Chemiluminescent Substrate kit (Thermo Scientific) and BIOMAX Light Film (Carestream).

Fibril formation and x-ray diffraction of TDP-LCD

SUMO-TDP-LCD was diluted into 1 mg/ml with the same dilution buffer used in aggregation assays of TDP-CTF and mixed with 100:1 (weight basis) homemade ULP1 protease. The fibril was formed by shaking the solution at 37 °C for 3 d and keeping it on a bench for another 2 d. The fibril was collected by centrifugation at 9,000 g for 5 min and washed with water twice. The fibril was sampled for TEM analysis and SDS–PAGE, then concentrated to 2% of the original volume by centrifugation and removal of supernatant. The concentrated solution was loaded between two siliconized glass capillaries with their heads facing each other. The solution was then dried out, and the fibril remained on one of the capillary heads in solid form. The capillary with fibril was mounted on an in-house X-ray machine, and the fibril was shot by X-ray for 3 min. The diffraction pattern was recorded on an imaging plate detector.

Fibril diffraction simulation of zipper forming segments

Crystal structures of zipper-forming segments were used for generation of fibril diffraction simulation. To simulate the fibril diffraction patterns, the space group of each crystal structure was expanded into P1, and structure factors were calculated using the CCP4 program [sfall69](#). The diffraction patterns were cylindrically averaged 180° around the fibril axis using custom-written software.

Negative-stain transmission electron microscopy (TEM)

Samples for TEM were prepared by applying 5 µl of sample on hydrophilic 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The fibrils were allowed to adhere for 4 min and washed twice with water. The samples were then stained for 2 min with

2% uranyl acetate, leaving a few nanoliters of stain on the grid. Grids were allowed to dry for 1 min before storage. Each grid was inspected on a T12 (FEI) microscope.

Aggregation assays of TDP-CTF

SUMO-TDP-CTF wild type and its mutants were diluted into 0.2 mg/ml with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 μ M DTT and filtered by 0.1 μ m Ultrafree-MC-VV centrifugal filters (Millipore). The filtered solution was mixed with 100:1 (weight basis) homemade ULP1 protease and incubated on ice to remove the SUMO-tag from TDP-CTF. The samples were taken before and 30 min after adding the protease, and the 30-min samples were divided into supernatant and pellet fractions by centrifugation at 18,000 g for 3 min at 4 °C. The pellet fraction was resuspended with the same buffer used in dilution and with the same volume of the original samples. The samples were mixed with 3:1 (v/v) of NuPAGE LDS sample buffer (Invitrogen) and heated to 100 °C for 10 min. After the 30-min samples were taken, the cleavage solutions were moved to the refrigerator (4 °C) and incubated overnight. The overnight samples were then taken and prepared in the same manner as the 30-min samples. All samples were loaded to and separated by NuPAGE 4-12% Bis-Tris gels (Invitrogen) and stained by Coomassie blue. To make better comparisons, the samples of wild-type TDP-CTF were loaded to both gels when the samples taken from the same time point were split into two gels because of loading well limitation.

Denaturing assays of ³¹²NFGAFS³¹⁷ and its variants

20% (m/v) SDS solution (in water) was diluted to 15%, 10% and 5% (m/v) by water, and 20% (v/v) of SDS–water stocks or water was mixed with 80% (v/v) of PBS to make SDS–PBS stocks with 4%, 3%, 2%, 1% and 0% (m/v) of SDS. SDS–PBS stocks with varies SDS content were

mixed with 20 mM peptide-PBS solution in 1:1 volume ratio to get final peptide solutions with 2%, 1.5%, 1%, 0.5% and 0% (m/v) SDS. All solutions were divided into three PCR tubes (5 μ l per tube), and the solutions containing 2%, 1.5%, 1% and 0.5% (m/v) SDS were heated to 70 °C for 15 min by a thermocycler. The OD₃₄₀ of solutions in each tube were measured by P330 NanoPhotometer (Implen), and the readings from same peptide were normalized to the mean values of the solutions contain no SDS and without heating.

A β fibril denaturing assay

A β purification and preparation was carried out as previously described⁷⁰. A β ₁₋₄₂ was diluted into 50 μ M with PBS buffer and incubated at 37 °C overnight to form A β fibrils. The fibril solution was mixed with 10% (v/v) of 20% (m/v) SDS stock to make 2% (m/v) final SDS concentration and heated to 70 °C for 15 min. The samples were taken before and after adding SDS and heating and analyzed by TEM.

Crystallization

All crystals, except ³¹²NFGEFS³¹⁷, ³¹²NFGpTFS³¹⁷ and ³³³SWGMMGMLASQ³⁴³ were grown by hanging-drop vapor diffusion. ³⁰⁰GNNQGSN³⁰⁶ was dissolved in water to a final concentration of 10 mg/ml. The reservoir solution was optimized around the initial hit of 0.2 M sodium sulfate, 0.1 M Bis Tris propane, pH 7.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein/reservoir.

³²¹AMMAAA³²⁶ was dissolved in water to a final concentration of 10 mg/ml, supplemented with 29.4 mM of LiOH. Crystals were grown over a reservoir solution containing 1.3 M ammonium phosphate and 0.1 M sodium acetate, pH 6.27. Hanging drops contained a 1:2 solution of protein/reservoir.

³²⁸AALQSS³³³ was dissolved in water to a final concentration of 10 mg/ml. Crystals were grown over a reservoir solution containing 0.01 M nickel chloride, 0.1 M Tris, pH 8.5 and 20% (w/v) PDG 2000. Hanging drops contained a 2:1 solution of protein/reservoir.

³⁷⁰GNNSYS³⁷⁵ was dissolved in water to a final concentration of 10 mg/ml. Crystals were grown over a reservoir solution containing 0.2 M sodium nitrate, 0.1 M Bis Tris propane, pH 8.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein/ reservoir.

³⁹⁶GFNGGFG⁴⁰² was dissolved in water to a final concentration of 12 mg/ml. Crystals formed with the reservoir solution: 800 mM sodium phosphate monobasic/ 1,200 mM potassium phosphate dibasic, 100 mM sodium acetate/acetic acid, pH 4.5. Hanging drops contained a 2:1 solution of protein/ reservoir.

³¹²NFGAFS³¹⁷ was dissolved in water to a final concentration of 11 mg/ml. Crystals formed with a reservoir solution of 0.1 M phosphate/citrate, pH 4.2, 40% w/v ethanol and 5% w/v PEG 1,000. The drops contained a 2:1 ratio of protein solution/reservoir.

³¹²NFGTFS³¹⁷ was dissolved in water to a final concentration of 11 mg/ml. Crystals formed with a reservoir solution of 0.2 M ammonium acetate, 0.1 M sodium acetate, pH 4.6, and 30% w/v PEG 4,000. The drops contained a 2:1 ratio of protein solution/ reservoir.

³¹²NFGEFS³¹⁷, ³¹²NFGpTFS³¹⁷ and ³³³SWGMMGMLASQ³⁴³ were dissolved in PBS (pH 7.5) to a final concentration of 20 mM. The peptide solution was shaken at 37 °C in a Torrey Pines shaker at level 9 for 80 h. To better understand the stacking of glutamate and phosphorylated threonine in ³¹²NFGEFS³¹⁷ and ³¹²NFGpTFS³¹⁷ structures, the pH of both peptide solutions was tested by MColorpHast pH-indicator strips (Millipore), and the pH of both solutions was around 4. For ³¹²NFGpTFS³¹⁷, initial crystals growing from shaking diffracted poorly, and in order to get

better diffracting crystals, an additional round of crystal seeding was performed. Fresh-made 20 mM $^{312}\text{NFGpTFS}^{317}$ PBS solution was filtered and mixed with 2% (v/v) of preformed $^{312}\text{NFGpTFS}^{317}$ crystals and incubated at 37 °C without shaking for 120 h. Diffraction data were collected from seeded crystals, and the initial crystals were used in denaturing assay and EM observation in Fig. 3.

X-ray diffraction data collection and processing

All X-ray diffraction data from crystalline specimens were collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA, at a wavelength of 0.971 Å and temperature of 100 K. NFGAFS data were collected using 5° oscillations and 130 mm detector distance. An ADSC Q315 CCD detector was used. Indexing and integration were performed using XDS in space group $P2_12_12_1$ and scaled with XSCALE to 1.1-Å resolution⁷¹. The data were imported using the CCP4 suite⁶⁹. Molecular replacement was performed using Phaser⁷². Initial search models consisting of different idealized polyalanine β-strands did not yield a correct solution. The successful search model used was a kinked strand with the sequence NFGAAS, which was based on the truncated structure of NNFGAIL (ref. ⁷³). The model was manually adjusted over three rounds of refinement to the correct sequence and rotamer orientation in COOT⁷⁴. The atomic model was refined by Phenix⁷⁵ with a final $R_{\text{work}} / R_{\text{free}}$ of 13.3 / 16.0% and 100% of Ramachandran angles favored.

For NFGTFS, data were collected using 5° oscillations and 125 mm detector distance. An ADSC Q315 CCD detector was used. Indexing and integration was performed using XDS in space group $P2_1$ and scaled with XSCALE to 1.0-Å resolution⁷¹. The data were imported using the CCP4 suite⁶⁹. The structure was solved by direct methods using the program SHELX⁷⁶. The model was manually adjusted over three rounds of refinement to the correct sequence and

rotamer orientation in COOT⁷⁴. The atomic model was refined by Refmac⁷⁷ with a final $R_{\text{work}} / R_{\text{free}}$ of 8.7 / 9.3% and 100% of Ramachandran angles favored.

For GFNGGFG, data were collected using 5° oscillations and 150 mm detector distance. An ADSC Q315 CCD detector was used. Indexing and integration was done using Denzo in space group $P1$ and scaled with SCALEPACK to 1.25-Å resolution⁷⁸. The data were imported using the CCP4 suite⁶⁹. Molecular replacement was done using Phaser⁷². Initial search models consisting of different idealized polyaniline β -strands did not yield a correct solution. The successful search model consisted of two kinked strands with the sequence GFAGG, which was based on the truncated structure of SYSGYS (ref. ²¹). The model was manually refined using Phenix⁷⁵ over four rounds of refinement to the correct seven-residue sequence and rotamer orientation in COOT⁷⁴. The atomic model has a final $R_{\text{work}} / R_{\text{free}}$ of 15.2 / 16.8% and 100% of Ramachandran angles favored.

For AMMAAA, data were collected using 5° oscillations and 130 mm detector distance. An EIGER 16 M pixel detector was used. Indexing and integration was done using XDS in space group $P2_12_12_1$ and scaled with XSCALE to 1.2-Å resolution⁷¹. The data were imported using the CCP4 suite⁶⁹. Molecular replacement was done using Phaser⁷² with an idealized polyaniline β -strand (AAAAAA). The model was manually adjusted over 14 cycles to the correct six-residue sequence and rotamer orientation in COOT⁷⁴. The atomic model was refined by Refmac⁷⁷, with a final $R_{\text{work}} / R_{\text{free}}$ of 19.8 / 21.8% and 100% of Ramachandran angles favored.

For AALQSS, data were collected using 5° oscillations and 130 mm detector distance. An EIGER 16 M pixel detector was used. Indexing and integration was done using XDS in space group $P2_12_12_1$ and scaled with XSCALE to 1.1-Å resolution⁷¹. The data were imported using

the CCP4 suite⁶⁹. The structure was solved by direct methods using the program SHELX⁷⁶. The model was manually adjusted over 13 rounds of refinement to the correct six-residue sequence and rotamer orientation in COOT⁷⁴. The atomic model was refined by Refmac⁷⁷, with a final $R_{\text{work}} / R_{\text{free}}$ of 19.5 / 24.3% and 100% of Ramachandran angles favored.

For GNNQGSN, data were collected using 5° oscillations and 125 mm detector distance. An ADSC Q315 CCD detector was used. Indexing and integration was done using XDS⁷¹ in space group C2 and scaled with SCALEPACK⁷⁹ to 1.8-Å resolution. The data were imported using the CCP4 suite⁶⁹. Molecular replacement was performed using Phaser⁷² with a search model consisting of an idealized β -strand of sequence GNNAGA. The model was manually adjusted over 38 rounds of refinement to the correct seven-residue sequence and rotamer orientation in COOT⁷⁴. The atomic model was refined by Refmac⁷⁷ with a final $R_{\text{work}} / R_{\text{free}}$ of 18.3 / 19.4% and 80% of Ramachandran angles favored and 20% allowed.

For GNNSYS, data were collected using 3° oscillations and 125 mm detector distance. An ADSC Q315 CCD detector was used. Indexing and integration was done using Denzo in space group $P2_12_12_1$ and scaled with SCALEPACK to 1.00-Å resolution⁷⁸. The data were imported using the CCP4 suite⁶⁹. Molecular replacement was performed using Phaser with a search model consisting of an idealized five-residue polyalanine β -strand⁷². The model was manually refined using Phenix over two rounds of refinement to the correct six-residue sequence and rotamer orientation in COOT^{74:75}. The atomic model has a final $R_{\text{work}} / R_{\text{free}}$ of 15.5 / 18.1% and 100% of Ramachandran angles favored.

MicroED data collection and processing

The MicroED data were collected using the same protocol as Rodriguez et al.[80](#) with minor differences, as detailed below[80:81:82](#). Data was collected at HHMI Janelia Research Campus using the Tecnai F20 TEM operating at 200 kV recorded using a bottom mount TVIPS F416 CMOS camera.

For NFGFEFS, individual image frames were exposed for 3 s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.2° per second. Diffraction data were collected and merged from four individual crystals. Indexing and integration were done using XDS[71](#) in space group $P2_12_12$ and scaled with SCALEPACK[79](#) to 1.0-Å resolution. The data were imported using the CCP4 suite[69](#). The structure was solved by direct methods using the program SHELX[76](#). The model was manually adjusted over four cycles to the correct sequence and rotamer orientation in COOT[74](#). The atomic model was refined by Refmac[77](#) with a final $R_{\text{work}} / R_{\text{free}}$ of 22.0 / 27.0% and 75% of Ramachandran angles favored and 25% allowed.

For NFGpTFS, individual image frames were exposed for 3 s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.295° per second. Diffraction data were collected on a single crystal. Indexing and integration were done using XDS in space group $P2_12_12_1$ and scaled with XSCALE to 0.75-Å resolution[71](#). The data were imported using the CCP4 suite[69](#). The structure was solved by direct methods using the program SHELX[76](#). The model was manually adjusted over 16 cycles to the correct sequence and rotamer orientation in COOT[74](#). The atomic model was refined by Refmac[77](#) with a final $R_{\text{work}} / R_{\text{free}}$ of 23.2 / 25.1% and 100% of Ramachandran angles favored.

For SWGMMGMLASQ, individual images were exposed for 2 s. During each exposure, the crystals were continuously rotated within the beam at a rate of 0.315° per second. Diffraction data were collected on seven crystals. Indexing and integration were performed XDS in space

group *P1* and scaled with XSCALE to 1.5-Å resolution⁷¹. The data were imported using the CCP4 suite⁶⁹. The structure was solved by molecular replacement using the program Phaser⁷² and a search model consisting of a nine-alanine-long ideal β -strand. The model was manually adjusted in COOT⁷⁴. The atomic model was refined by Refmac⁷⁷ and Buster⁸³ with a final $R_{\text{work}} / R_{\text{free}}$ of 28.0 / 31.3% and 94.5% of Ramachandran angles favored and 5.5% allowed. These *R* factors are higher than expected for a dataset of this quality. We think the discrepancy is due to a disorder in the crystal. The *P1* unit cell dimensions of SWGMMGLASQ allow pseudo-merohedral twinning (pseudo space group *C2*). Also, the native Patterson map suggests that the crystal contains a translocation disorder (peak at 0.379, -0.264, -0.471 with height of 36% of the origin peak). Attempts to correct for these disorders did not significantly improve the *R* factors.

Structure analysis

All images of crystallographic structures were generated in PyMOL. Area buried (Ab) and shape complementarity (Sc) values were calculated based on published methods^{84,85}.

Data availability

Density maps and atomic coordinates have been deposited to the Protein Data Bank (PDB) with accession codes PDB [5WKD](#) (GNNQGSN); PDB [6CEW](#) (AMMAAA); PDB [6CB9](#) (AALQSS); PDB [5WIQ](#) (GFNGGFG); PDB [5WIA](#) (GNNSYS); PDB [5WHN](#) (NFGAFS); and PDB [5WHP](#) (NFGTFS). The MicroED models have been deposited into the EM Data Bank and PDB with accession codes EMD-[7467](#) and PDB [6CFH](#) (SWGMMGMLASQ); EMD-[8857](#) and PDB [5WKB](#) (NFGFEFS); and EMD-[7466](#) and PDB [6CF4](#) (NFGpTFS). All other data are available upon reasonable requests to the authors.

Acknowledgements

We thank S. Sangwan and P. Seidler for discussion, D. Shi and T. Gonen at Janelia for microscope support and M. Collazo at UCLA-DOE Macromolecular Crystallization Core Technology Center for crystallization support. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Eiger 16 M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This research was supported in part by grants from the National Institutes of Health NIH NIA AG029430, NIH NIA AG054022, Howard Hughes Medical Institute and the Janelia Research Campus visitor program. This material is based upon work supported by the National Science Foundation under Grant No. NSF 1616265. We acknowledge the use of instruments at the Electron Imaging Center for Nanomachines supported by UCLA and by instrumentation grants from NIH (1S10RR23057 and 1U24GM116792) and NSF (DBI-1338135). D.R.B. was supported by the National Science Foundation Graduate Research Fellowship Program.

Author contributions

E.L.G., Q.C. and D.S.E. designed the project and wrote the manuscript with input from all other authors, especially M.R.S. E.L.G. and H.T. conducted fibril growth assays and prepared peptide crystals. E.L.G., Q.C. and J.L. cloned and purified TDP-43 constructs and performed the protein aggregation assay and fibril diffraction. M.P.H. predicted putative LARKS. E.L.G., H.T. and M.R.S. processed and solved ³⁰⁰GNNQGSN³⁰⁶. E.L.G. and M.R.S. processed and solved ³¹²NFGAFS³¹⁷, ³¹²NFGTFS³¹⁷, ³⁷⁰GNNSYS³⁷⁵ and ³⁹⁶GFNGGFG⁴⁰². E.L.G., Q.C., J.L.

and M.R.S. processed data and solved the ³²¹AMMAAA³²⁶ and ³²⁸AALQSS³³³. J.A.R. collected MicroED data on ³¹²NFGEFS³¹⁷ and ³³³SWGMMGMLASQ³⁴³, M.R.S., D.C. and E.L.G. processed data and solved the structure of ³¹²NFGEFS³¹⁷ and ³³³SWGMMGMLASQ³⁴³. D.R.B. collected MicroED data for ³¹²NFGpTFS³¹⁷. Q.C., M.R.S., D.C. and D.R.B. processed data and solved the structure of ³¹²NFGpTFS³¹⁷. E.L.G., Q.C., M.R.S., M.P.H. and D.S.E. analyzed structures and designed the model of the LCD in SG formation and pathological aggregation.

Competing interests

D.S.E. is an advisor and equity shareholder in ADRx, Inc.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-018-0064-2>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.S.E.

Figures

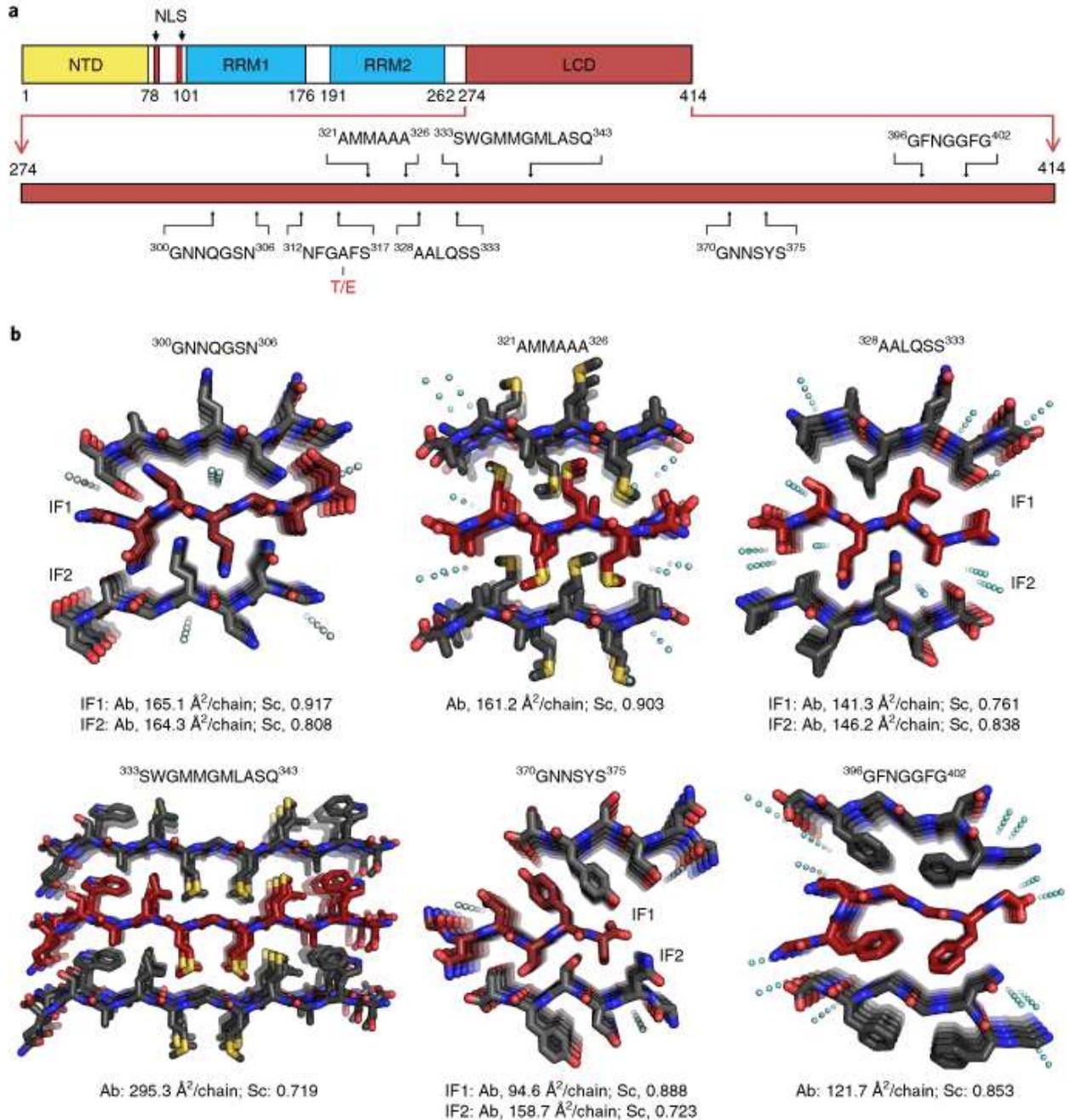


Fig. 1: Segments from the LCD of TDP-43 form steric zippers.

a, Schematic representation of full-length TDP-43. The LCD is enlarged and the sequences, and the relative positions of segments for which we determined structures are shown. NTD, N-terminal domain; NLS, nuclear localization sequence; RRM, RNA-recognition motif. **b**, Six steric-zipper structures we determined in this study. For each structure, three β -sheets and two interfaces between these sheets are shown, and each sheet shows five of its thousands of strands. For ³⁰⁰GNNQGSN³⁰⁶, ³²⁸AALQSS³³³ and ³⁷⁰GNNSYS³⁷⁵, two interfaces (IF) are distinct and are labeled IF1 and IF2; for ³²¹AMMAAA³²⁶, ³³³SWGMMGMLASQ³⁴³ and ³⁹⁶GFNGGFG⁴⁰², all

interfaces are identical because of the symmetry of packing. The area buried (Ab) and shape complementary (Sc) values for each interface are shown below the structure. Water molecules found in the structure are shown as cyan droplets. All of the structures are shown in top view, viewed down the fibril axis.

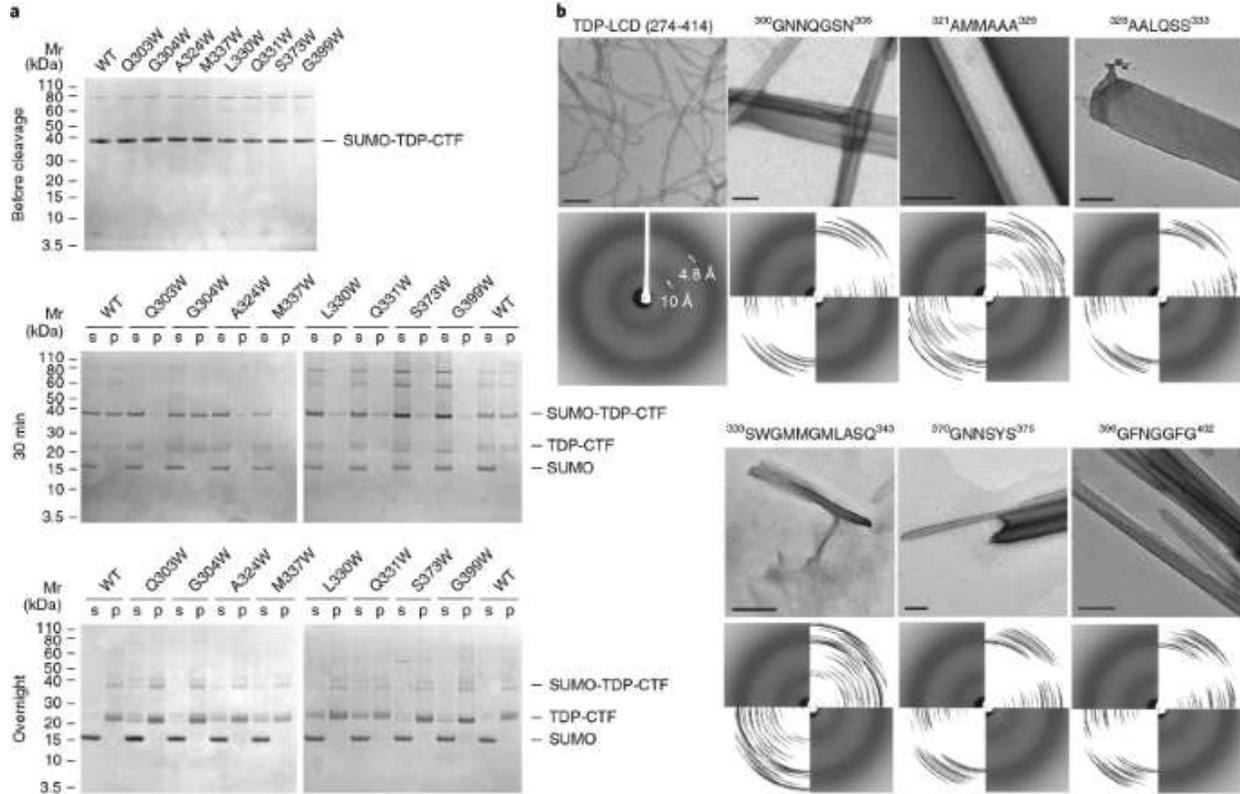


Fig. 2: Validation of steric-zipper-forming segments.

a, Aggregation assays of the pathological fragment TDP-CTF (208–414), both wild type (WT) and with mutations to block steric-zipper formation. TDP-CTF was conjugated with a SUMO protein tag to prevent aggregation, and ULP1 protease was added to remove the SUMO tag. TDP-CTF was incubated on ice for 30 min after cleavage and continued to incubate at 4 °C overnight. Samples were taken after incubation and separated into supernatant (s) and pellet (p) by centrifugation and analyzed by SDS–PAGE. Notice that TDP-CTF G304W has similar behavior to that of wild-type TDP-CTF, whereas other mutants show reduced aggregation compared to the wild type. **b**, Negative-stain EM and fibril diffraction pattern of TDP-LCD and steric-zipper segments. TDP-LCD was conjugated with SUMO tag and mixed with ULP1 protease to remove the SUMO tag. The X-ray diffraction pattern of TDP-LCD fibrils was acquired by experiment, and the diffraction patterns of zipper-forming segments were computationally simulated from their crystal structures. The diffraction pattern of TDP-LCD fibrils was overlaid with the patterns of zipper-forming segments for comparison. Notice that TDP-LCD fibrils produce diffraction rings at 4.8 Å and 10 Å, indicating that they have cross-β spines. Scale bars, 200 nm.

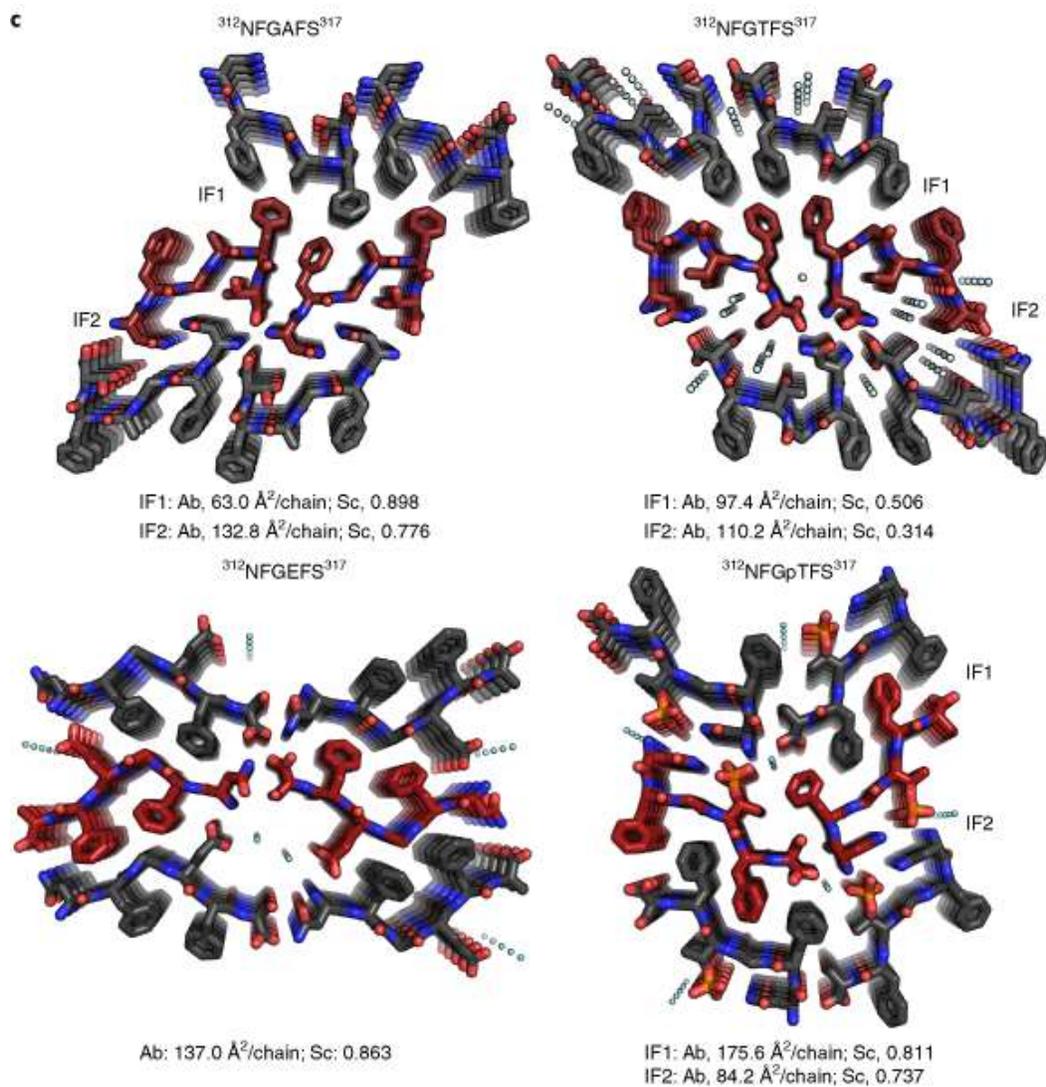
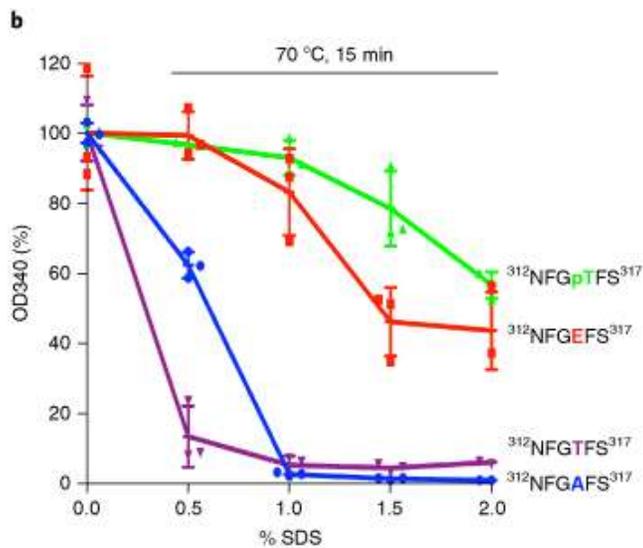
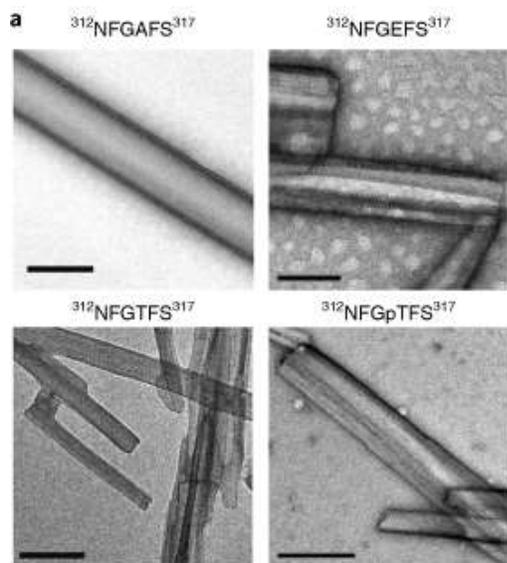


Fig. 3: The $^{312}\text{NFGAFS}^{317}$ segments form a kinked β -sheet that is strengthened by familial variants A315T and A315E.

a, Negatively-stained EM images of aggregates of $^{312}\text{NFGAFS}^{317}$ and its variants after 4 d of shaking at 37 °C. Scale bars, 200 nm. **b**, Denaturing assays of aggregates of $^{312}\text{NFGAFS}^{317}$ and its variants. The aggregates of each segment were treated with the indicated percentages of SDS and heated to 70 °C for 15 min. The absorbance at wavelength 340 (OD340) of the solution after treatment was measured to represent turbidity. The readings from each segment are shown as a percentage of the mean values of the segment treated with 0% SDS, without heat, as mean \pm s.d. ($n = 3$ independent experiments) with individual data overlaid. **c**, Structures of $^{312}\text{NFGAFS}^{317}$ and its variants. For each structure, six sheets are shown with two labeled interfaces (IF1 and IF2). For $^{312}\text{NFGFEFS}^{317}$, the two interfaces between the parallel packed sheets are identical because of its face-to-back packing, so the interfaces are not labeled. The Ab and Sc values for each interface are shown below the structure. Water molecules found in the structure are shown as cyan droplets. All of the structures are viewed down the fibril axis. Notice that the larger interfaces in both $^{312}\text{NFGFEFS}^{317}$ and $^{312}\text{NFGpTFS}^{317}$ are increased over those in $^{312}\text{NFGAFS}^{317}$ and $^{312}\text{NFGTFS}^{317}$, explaining in part their greater stability. Additional hydrogen bond stabilization is shown in Supplementary Fig. 3.

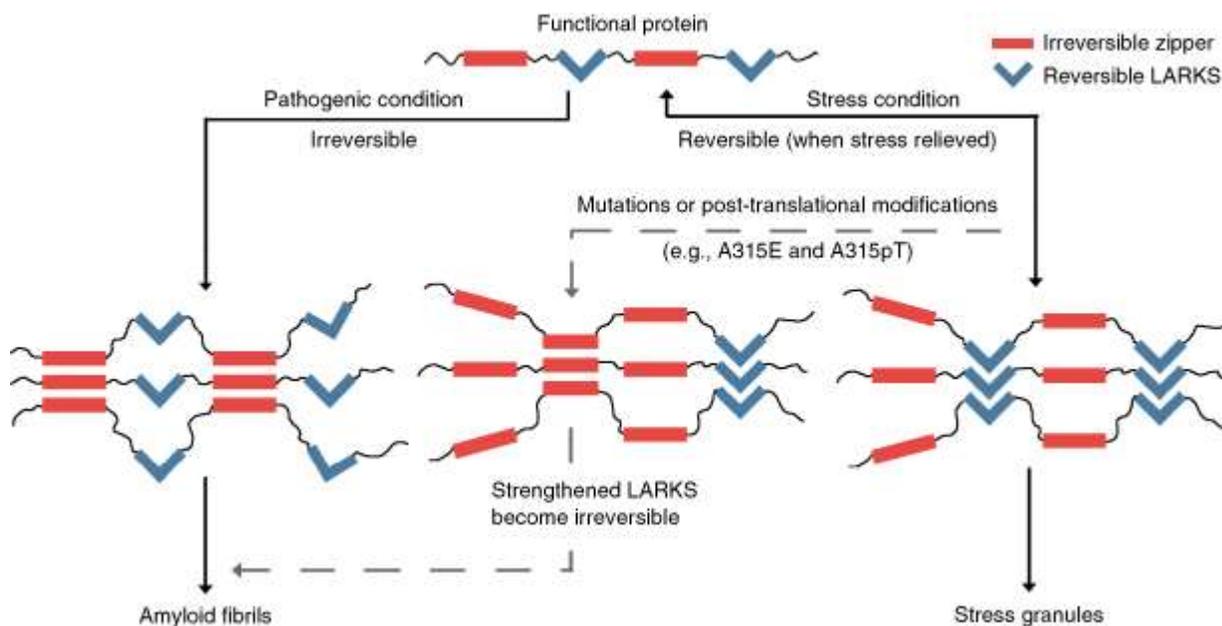


Fig. 4: Speculative model showing the alternative pathways of formation of stress granules with pathogenic amyloid.

TDP-43 forms several irreversible steric zippers (red bars) and several reversible LARKS (blue kinked bars). When performing its normal functions, the protein is not aggregated (top). Under pathogenic conditions (left pathway), the steric zipper segments interact with each other, acting as molecular glue to drive the protein into irreversible aggregation and forming amyloid fibrils; under stress conditions (right pathway), the LARKS interact with each other and, acting as molecular Velcro, provide adhesion to form a hydrogel as in stress granules; these interactions

are reversible. The crystal structures of $^{312}\text{NFGAFS}^{317}$, $^{312}\text{NFGEFS}^{317}$ and $^{312}\text{NFGpTFS}^{317}$ show that A315E and A315T with phosphorylated threonine can strengthen the LARKS to irreversible aggregates. Either mutation or post-translational modifications can act as the agent of conversion (middle pathway). Because the involvement of $^{312}\text{NFGAFS}^{317}$ in SG-pathogenic amyloid transition is yet to be validated, the middle pathway is shown by gray dashed lines, emphasizing that this pathway is speculative.

Tables

Table 1 Data collection and refinement statistics							
	GNNQGSN (PDB 5WKD)	AMMAAA (PDB 6CEW)	AALQSS (PDB 6CB9)	GFNGGFG (PDB 5W1Q)	GNNSYS (PDB 5W1A)	NFGAFS (PDB 5WHN)	NFGTFS (PDB 5WHP)
Data collection							
Space group	C2	P2,2,2 ₁	P2,2,2 ₁	P1	P2,2,2 ₁	P2,2,2 ₁	P2 ₁
Cell dimensions							
a, b, c (Å)	50.35, 4.77, 14.75	9.52, 15.44, 44.1	4.79, 15.97, 43.43	4.81, 16.44, 22.67	4.79, 15.62, 40.68	13.82, 4.85, 46.74	15.33, 4.84, 23.57
α, β, γ (°)	90, 101.73, 90	90, 90, 90	90, 90, 90	90.63, 96.37, 91.26	90, 90, 90	90, 90, 90	90.00, 96.91, 90.00
Resolution (Å)	1.80 (1.86-1.80)*	1.20 (1.23-1.20)	1.10 (1.14-1.10)	1.25 (1.29-1.25)	1.00 (1.04-1.00)	1.10 (1.14-1.10)	1.00 (1.03-1.00)
R_{merge} (%)	15.8 (25.0)	18.1 (68.3)	21.1 (56.2)	16.1 (40.6)	15.5 (55.4)	17.2 (72.6)	8.6 (10.4)
$I/\sigma(I)$	6.9 (3.3)	3.2 (1.4)	5.1 (2.8)	7.1 (2.0)	9.2 (2.1)	9.7 (2.5)	15.0 (7.6)
CC _{1/2}	97.1 (81.6)	99.5 (69.2)	97.0 (91.3)	98.5 (69.0)	99.2 (66.8)	99.1 (82.3)	99.3 (98.0)
Completeness (%)	91.6 (84.8)	96.0 (99.3)	90.7 (72.1)	94.4 (95.4)	95.2 (62.7)	94.0 (75.9)	87.6 (37.0)
Redundancy	2.3 (2.2)	3.5 (3.5)	6.9 (7.3)	4.1 (1.8)	5.4 (1.9)	8.3 (6.4)	4.9 (2.8)
Refinement							
Resolution (Å)	24.32-1.80	22.05-1.20	21.22-1.10	22.53-1.25	20.34-1.00	13.25-1.10	15.22-1.00
No. reflections	345	2,078	1,288	1,796	1,836	1,300	1,639
$R_{\text{int}} / R_{\text{free}}$	18.3 / 19.4	19.8 / 21.8	19.5 / 24.3	15.2 / 16.8	15.5 / 18.1	13.3 / 16.0	8.7 / 9.3
No. atoms							
Protein	48	79	40	94	45	50	48
Water	2	4	4	8	1	0	3
B factors (Å²)							
Protein atoms	8.7	9.1	7.3	3.2	3.5	4.6	2.4
Water	12.2	20.7	20.0	7.6	15.3	-	14.8
R.m.s. deviations							
Bond lengths (Å)	0.021	0.012	0.010	0.005	0.005	0.006	0.019
Bond angles (°)	1.804	1.431	1.422	0.692	1.18	0.95	1.79

*Values in parentheses are for highest-resolution shell.

Table 2 | MicroED data collection and refinement statistics

	SWGMMGMLASQ (EMD-7467, PDB 6CFH)	NFGEFS (EMD-8857, PDB 5WKB)	NFGpTFS (EMD-7466, PDB 6CF4)
Data collection			
Space group	<i>P</i> 1	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	8.56, 9.60, 39.97	42.77, 17.42, 4.90	23.65, 4.72, 30.06
α , β , γ (°)	97.2, 92.9, 105.9	90, 90, 90	90, 90, 90
Resolution (Å)	1.50 (1.55-1.50) ^a	1.00 (1.03- 1.00)	0.75 (0.77- 0.75)
<i>R</i> _{merge} (%)	20.8 (85.5) ^b	28.3 (99.3) ^c	17.2 (66.1)
<i>I</i> / σ (<i>I</i>)	3.3 (0.7)	4.6 (1.1)	3.9 (1.3)
CC _{1/2}	98.7 (72.0)	99.2 (36.0)	98.9(59.9)
Completeness (%)	93.5 (89.4)	88.7 (78.0)	86.6(87.2)
Redundancy	4.2 (2.4)	9.5 (4.8)	3.8 (4.0)
Refinement			
Resolution (Å)	13.17-1.50	21.39-1.00	7.65-0.75
No. reflections	1,819	1,801	4,178
<i>R</i> _{work} / <i>R</i> _{free}	28.0 / 31.3	22.0 / 27.0	23.2 / 25.1
No. atoms			
Protein	162	90	93
Water	0	2	1
<i>B</i> factors			
Protein	18.1	5.3	19.6
Water	-	25.3	3.1
R.m.s. deviations			
Bond lengths (Å)	0.007	0.016	0.021
Bond angles (°)	0.91	1.09	2.04

^aValues in parentheses are for highest-resolution shell. ^bData are from seven crystals. ^cData are from four crystals.

References

1. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
2. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
3. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS mutations disrupt phase separation mediated by α -helical structure in the TDP-43 low complexity C-terminal domain. *Structure* **24**, 1537–1549 (2016).
4. Kato, M. & McKnight, S. L. A solid-state conceptualization of information transfer from gene to message to protein. *Annu. Rev. Biochem.* <https://doi.org/10.1146/annurev-biochem-061516-044700> (2017).
5. Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
6. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
7. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23**, 217–247 (2000).
8. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* **1647**, 9–18 (2016).

9. Bentmann, E. et al. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* **287**, 23079–23094 (2012).
10. Colombrita, C. et al. TDP-43 is recruited to stress granules in conditions of oxidative insult. *J. Neurochem.* **111**, 1051–1061 (2009).
11. Dewey, C. M. et al. TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Mol. Cell. Biol.* **31**, 1098–1108 (2011).
12. Gilks, N. et al. Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell* **15**, 5383–5398 (2004).
13. Dang, Y. et al. Eukaryotic initiation factor 2 α -independent pathway of stress granule induction by the natural product pateamine A. *J. Biol. Chem.* **281**, 32870–32878 (2006).
14. Kedersha, N. et al. Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *J. Cell Biol.* **151**, 1257–1268 (2000).
15. Patel, A. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
16. Reineke, L. C. et al. Casein kinase 2 is linked to stress granule dynamics through phosphorylation of the stress granule nucleating protein G3BP1. *Mol. Cell. Biol.* **37**, e00596–e16 (2017).
17. Murray, D. T. et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615–627.e16 (2017).
18. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *eLife* **5**, 1–25 (2016).

19. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
20. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
21. Hughes, M. P. et al. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018).
22. Eisenberg, D. S. & Sawaya, M. R. Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).
23. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).
24. Meersman, F. & Dobson, C. M. Probing the pressure-temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochim. Biophys. Acta.* **1764**, 452–460 (2006).
25. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474**, 1417–1438 (2017).
26. Chang, C. K. et al. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem. Biophys. Res. Commun.* **425**, 219–224 (2012).
27. Lukavsky, P. J. et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20**, 1443–1449 (2013).
28. Saini, A. & Chauhan, V. S. Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir* **30**, 3845–3856 (2014).

29. Buratti, E. & Baralle, F. E. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biol.* **7**, 420–429 (2010).
30. Lee, E. B., Lee, V. M. Y. & Trojanowski, J. Q. Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci.* **13**, 38–50 (2011).
31. Schwab, C., Arai, T., Hasegawa, M., Yu, S. & McGeer, P. L. Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease. *J. Neuropathol. Exp. Neurol.* **67**, 1159–1165 (2008).
32. Amador-Ortiz, C. et al. TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann. Neurol.* **61**, 435–445 (2007).
33. Nakashima-Yasuda, H. et al. Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathol.* **114**, 221–229 (2007).
34. Higashi, S. et al. Concurrence of TDP-43, tau and α -synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. *Brain Res.* **1184**, 284–294 (2007).
35. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
36. Kabashi, E. et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40**, 572–574 (2008).
37. Chiang, C.-H. et al. Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Sci. Rep.* **6**, 21581 (2016).

38. Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D. W. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19R1**, R46–R64 (2010).
39. Pesiridis, G. S., Lee, V. M. Y. & Trojanowski, J. Q. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **18R1**, R156–R162 (2009).
40. Igaz, L. M. et al. Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* **284**, 8516–8524 (2009).
41. Hasegawa, M. et al. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann. Neurol.* **64**, 60–70 (2008).
42. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
43. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349**, 650–655 (2015).
44. Yang, C. et al. The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. *PLoS One* **5**, e15878 (2010).
45. Chen, A. K. et al. Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* **132**, 1186–1187 (2010).
46. Cushman, M., Johnson, B. S., King, O. D., Gitler, A. D. & Shorter, J. Prion-like disorders: blurring the divide between transmissibility and infectivity. *J. Cell Sci.* **123**, 1191–1201 (2010).
47. Guo, W. et al. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol.* **18**, 822–830 (2011).

48. Jiang, L. L. et al. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* **288**, 19614–19624 (2013).
49. Jiang, L. L. et al. Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Sci. Rep.* **6**, 23928 (2016).
50. Mompeán, M. et al. “Structural characterization of the minimal segment of TDP-43 competent for aggregation”. *Arch. Biochem. Biophys.* **545**, 53–62 (2014).
51. Sawaya, M. R. et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
52. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* **107**, 3487–3492 (2010).
53. Budini, M. et al. Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. *J. Biol. Chem.* **287**, 7512–7525 (2012).
54. Fuentealba, R. A. et al. Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. *J. Biol. Chem.* **285**, 26304–26314 (2010).
55. Seidler, P. M. et al. Structure-based inhibitors of tau aggregation. *Nat. Chem.* **10**, 170–176 (2018).
56. Xiang, S. et al. The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* **163**, 829–839 (2015).
57. Gitcho, M. A. et al. TDP-43 A315T mutation in familial motor neuron disease. *Ann. Neurol.* **63**, 535–538 (2008).

58. Xu, M. et al. Characterization of β -domains in C-terminal fragments of TDP-43 by scanning tunneling microscopy. *J. Struct. Biol.* **181**, 11–16 (2013).
59. Lim, L., Wei, Y., Lu, Y. & Song, J. 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**, e1002338 (2016).
60. Estes, P. S. et al. Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a *Drosophila* model of ALS. *Hum. Mol. Genet.* **20**, 2308–2321 (2011).
61. Walker, A. K. et al. ALS-associated TDP-43 induces endoplasmic reticulum stress, which drives cytoplasmic TDP-43 accumulation and stress granule formation. *PLoS One* **8**, e81170 (2013).
62. Bosco, D. A. et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Hum. Mol. Genet.* **19**, 4160–4175 (2010).
63. Hackman, P. et al. Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. *Ann. Neurol.* **73**, 500–509 (2013).
64. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
65. Schmidt, H. B. & Rohatgi, R. In vivo formation of vacuolated multi-phase compartments lacking membranes. *Cell Reports* **16**, 1228–1236 (2016).
66. Franzmann, T. M. et al. Phase separation of a yeast prion protein promotes cellular fitness. *Science* **359**, eaao5654 (2018).

67. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
68. Soragni, A. et al. A designed inhibitor of p53 aggregation rescues p53 tumor suppression in ovarian carcinomas. *Cancer Cell* **29**, 90–103 (2016).
69. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
70. Cheng, P.-N., Liu, C., Zhao, M., Eisenberg, D. & Nowick, J. S. Amyloid β -sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nat. Chem.* **4**, 927–933 (2012).
71. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
72. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
73. Wiltzius, J. J. et al. Atomic structure of the cross-beta spine of islet amyloid polypeptide (amylin). *Protein Sci.* **17**, 1467–1474 (2008).
74. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
75. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
76. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
77. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

78. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
79. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
80. Rodriguez, J. A. et al. Structure of the toxic core of α -synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
81. Hattne, J. et al. MicroED data collection and processing. *Acta Crystallogr. A Found. Adv.* **71**, 353–360 (2015).
82. Shi, D. et al. The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 (2016).
83. Bricogne, G. et al. BUSTER v.2.9. (Global Phasing Ltd., Cambridge, 2010).
84. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
85. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).

Chapter 5

Cryo-EM structures of hIAPP fibrils seeded by patient-extracted fibrils reveal new polymorphs
and conserved fibril cores

Abstract

Amyloidosis of human islet amyloid polypeptide (hIAPP) is a pathological hallmark of type II diabetes (T2D), an epidemic afflicting nearly 10% of the world's population. To visualize disease-relevant hIAPP fibrils, we extracted amyloid fibrils from islet cells of a T2D donor and amplified their quantity by seeding synthetic hIAPP. Cryo-EM studies revealed four fibril polymorphic atomic structures. Their resemblance to four unseeded hIAPP fibrils varies from nearly identical (TW3) to non-existent (TW2). The diverse repertoire of hIAPP polymorphs appears to arise from three distinct protofilament cores entwined in different combinations. The structural distinctiveness of TW1, TW2 and TW4 suggests they may be faithful replications of the pathogenic seeds. If so, the structures determined here provide the most direct view yet of hIAPP amyloid fibrils formed during T2D.

Main

The tissue deposition of amyloid fibrils is a hallmark of the amyloid family of diseases, including Alzheimer's disease, Parkinson's disease and type II diabetes (T2D). These fibrils are associated with and possibly causal of disease¹. Associated with T2D are fibrils of hIAPP (or amylin), a human protein hormone of 37 residues, which works with insulin to regulate blood sugar levels^{2,3}. Pancreatic amyloid deposits of hIAPP are found in more than 90% of T2D patients^{4,5,6}, and genetic and experimental evidence links them to the pathogenic mechanism of T2D^{7,8,9,10,11}. Accordingly, determining the near-atomic-resolution structures of these disease-related hIAPP fibrils may help to understand the molecular mechanism of pathogenesis of T2D, and can offer essential information for the structure-based design of inhibitors of amyloidosis.

Our previous cryo-EM study revealed the structure of amyloid fibrils formed by full-length recombinant hIAPP fused to a small ubiquitin-like modifier (SUMO)-tag¹², adding to the collection of models and near-atomic-resolution structures reported for fibrils of synthetic and untagged hIAPP peptide^{13,14,15,16}. Whereas these structures were all obtained from hIAPP fibrils formed in vitro, recent studies on tau and α -synuclein suggest that in vitro fibrils may not adopt the same structures as patient-extracted fibrils^{17,18}. These findings raise the question of whether the reported hIAPP fibril structures are disease-related. In this Article, we extract hIAPP fibrils from islet cells of a donor with T2D to seed the fibril formation of synthetic hIAPP, and determine four cryo-EM structures from seeded fibrils, with the expectation that the seeded fibrils may replicate the structure of the seeds (Supplementary Text [1](#)).

Results

T2D patient fibrils seed fibrilization of synthetic hIAPP

We extracted hIAPP fibrils from the islet cells of a donor with T2D, and used these extracted fibrils to seed the fibril formation of synthetic hIAPP peptides. We first tested the amyloid content in islet cells from seven donors with T2D (donors 1 to 7) by Congo red staining. Samples from one donor (donor 6) exhibited strong Congo red staining while samples from two other donors (donors 4 and 5) exhibited medium staining (Fig. [1a](#), Extended Data Fig. [1a](#) and Supplementary Table [1](#)). We selected donor 6 for further study. Genotyping revealed that the hIAPP encoded was wild type. We then performed fibril extraction from islet cells of donor 6 by following the protocol reported previously and combined with immunoprecipitation ([Methods](#)). Our goal was to determine the structure of T2D-relevant hIAPP fibrils. However, patient extracted fibrils are too scarce for cryo-EM study. To amplify the quantity of sample, we used

patient-extracted fibrils to seed synthetic hIAPP, generating abundant, long fibrils for cryo-EM determination (Fig. [1](#) and Extended Data Fig. [1](#); for details see [Methods](#)).

Cryo-electron microscopy structures of seeded hIAPP fibrils

We identified eight different morphologies during two-dimensional (2D) classification: four were clearly twisted (termed TW1–4) and suitable for structure determination and four were not twisted (termed NT1–4; Fig. [2](#) and Extended Data Fig. [2a,b](#)). TW1 is the most abundant species and contains ~30% of all identifiable segments, whereas TW2, TW3 and TW4 contain 15%, 10% and 10%, respectively. NT1–4 together account for the remaining 35% of particles, but their lack of twisting precludes 3D structure determination. We determined the cryo-EM structures of TW1 to TW4 with resolution ranging from 3.8 to 4.1 Å (Fig. [2](#), Extended Data Figs. [2c–e](#) and [3](#) and Table [1](#)). These structures displayed the typical cross- β scaffold of amyloid fibrils, consisting of layers of β -strands stacked with 4.8-Å spacing (Fig. [2](#) and Extended Data Fig. [3](#)).

All four fibril polymorphs are composed of two intertwined chains or protofilaments. The two protofilaments are related by a C_2 symmetry axis in the TW2 fibrils and by a pseudo- $P2_1$ axis in TW3 fibrils. In TW1 and TW4, which we term heterotypic fibrils, the two protofilaments (chains A and B) are conformationally distinct from each other (Fig. [2](#) and Extended Data Fig. [4a](#)).

We note that at the current resolution (3.7 to 4.1 Å), cryo-EM refinement usually suffers from local minima problems¹⁹, which may raise the risk of an incorrect map and/or model. However, we were able to build unambiguous models based on map fitting as well as external information, such as cross-validation via other structures and determining the terminus of the main chain (see [Methods](#) for details).

In the final models, we observed density for the C-terminal residue of hIAPP (Tyr37) in all chains except for TW4 chain B (Extended Data Fig. [5a](#)). The lack of additional density beyond Tyr37 in all polymorphs except TW4 chain B and the lack of space for building additional residues beyond Tyr37 in TW2, TW3 and TW4 chain A supports the registration of the models into the maps (Extended Data Figs. [2c](#) and [5b,c](#)). Furthermore, the finding that the amide moiety of Tyr37 forms a hydrogen bond to stabilize the fibril fold on TW2, TW3 and TW4 chain A is consistent with the previous observation that C-terminal amidation promotes amyloidosis of hIAPP²⁰ (Extended Data Fig. [5c](#)). Density for the N-terminal region was more variable; the first ordered residue differs among polymorphs from Thr6 (in TW2 and TW4 chain A) to Val17 (TW1 chain B; Extended Data Fig. [5a](#)). In TW2–4, we found extra density near the N terminus, suggesting the position of additional N-terminal residues (Fig. [2](#), Extended Data Fig. [5b](#) and [Discussion](#)).

Three seeded polymorphs are previously unreported forms

Three of the four seeded hIAPP fibril structures are distinct from previously reported unseeded ones. We consider fibrils [6Y1A15](#) and [6ZRF16](#) as the unseeded controls of our study, given that, in all three studies, identical starting material (C-terminally amidated hIAPP) and similar fibril growth conditions were used. Those two unseeded control studies revealed fibril structures similar to each other (disregarding the opposite assignment of fibril handedness, Extended Data Fig. [4b](#); see [Methods](#) for details) as the dominant species (90% and 80% of total fibril populations, respectively). In our seeded experiment, we found that one species, TW3, matches the dominant conformation of [6Y1A](#) and [6ZRF](#) (Extended Data Fig. [4b](#)), but it contributes only 10% of the total population. The majority of the seeded fibrils are polymorphs TW1, TW2 and TW4, whose structures differ from the unseeded controls.

Structural alignment at the protofilament level

We selected all unique hIAPP conformations (a single layer in a protofilament in a fibril) among all reported hIAPP fibril structures (Table 2) and performed pairwise structural alignment (Methods and Supplementary Table 2). We found all conformations except TW2 can be categorized into two groups: core fold 1 (CF1) and core fold 2 (CF2; Fig. 3a,b, Supplementary Text 2 and Extended Data Fig. 6). Unlike the other structures, TW2 is composed of unique protofilament folds, as evidenced by the superposition with TW1 chain A and chain B (Extended Data Fig. 4d). In the following, we refer to the fold of TW2 as core fold 3 (CF3).

CF1 and CF2 form different protofilament interfaces

The two core folds (CF1, CF2) we observe in most hIAPP fibril structures can be combined in three different ways to form a fibril. All three have been observed, including the two homotypic types, CF1-CF1 (TW1, 6Y1A and 6ZRF) and CF2-CF2 (6ZRR, 6ZRQ and 6VW2), and the heterotypic type, CF1-CF2 (TW1, TW4 and 6ZRR; Fig. 3b). The interfaces at which the different protofilaments meet are not conserved; for example, when aligned at CF2 (chain B of TW1, TW4 and 6ZRR), we found that the other chains from these three structures are not aligned (Fig. 3c). We examined all substantial interfaces of protofilaments presented in hIAPP fibril structures (Extended Data Fig. 5f), and we found all these interfaces are different (Supplementary Text 3), suggesting that polymorphism arises at the level of protofilament assembly into fibrils.

Comparison of wild-type and S20G hIAPP fibril structures

S20G of hIAPP is the only hereditary mutation found so far in T2D patients, and it has been reported to increase the amyloidosis of hIAPP and cause early onset of T2D^{21,22,23,24}.

Structures of hIAPP fibrils containing the S20G mutation have been determined and deposited in the PDB with codes [6ZRR](#) and [6ZRQ16](#). By contrast, the genotype of the patient sample of our study identifies the seeds as wild-type hIAPP, and the seeded synthetic monomeric hIAPP we used is also wild type. Comparison of the S20G structures with the wild-type structures reveals that S20G structures exhibit the same conserved kernels, CF1 and CF2, as the wild-type structures (Fig. [3b](#)). These observations suggest that the S20G mutation enhances fibrilization not by creating new core folds, but instead by favoring the same folds exhibited by wild-type hIAPP.

To improve our understanding of the mechanism by which S20G stabilizes hIAPP fibril structures, we marked position 20 on hIAPP structures with arrows (Fig. [3b](#)) and found that all are located at or near sharp kinks in the main chain. Because the glycine residue is more permissive of a kinked conformation, we suggest that the S20G hereditary mutation promotes amyloid formation by allowing the peptide chain to more easily adopt a fibril-forming fold.

In addition to favoring a kinked conformation in the main chain, we found another mechanism by which S20G may favor CF2 over CF1 and CF3. We found that among all hIAPP protofilaments that contain CF2, the backbone phi angle at position 20 is positive, whereas the phi angle of position 20 in all other protofilaments is negative (Fig. [3d](#)). Positive phi angles are favored by glycine more than any of the other 19 amino acid residues, because only glycine lacks a side chain larger than hydrogen that would otherwise clash with the main chain in this conformation. This observation suggests that S20G relieves the steric clash among protofilaments that contain CF2, permitting their formation. This hypothesis is supported by the observation that protofilaments with CF2 are more abundant in the S20G fibril dataset than in wild-type datasets, with the exception of [6VW2](#) (Extended Data Fig. [7a](#)). We note in the 6VW2

dataset that the fibrils are formed by SUMO-tagged recombinant hIAPP without C-terminal amidation, differing from the synthetic and untagged hIAPP used in the studies of [6Y1A](#), [6ZRR](#) and [6ZRQ15:16](#). So perhaps the existence of the SUMO-tag and/or the lack of C-terminal amidation drives the fibrils in [6VW2](#) to the S20G favored CF2 conformation.

Structural similarity of hIAPP and amyloid- β fibrils

A clinical link between T2D and Alzheimer's disease has been reported [25:26:27](#), and evidence suggests that this link arises from the cross-seeding between hIAPP and amyloid- β fibrils in patients [28:29:30:31:32](#). Cross-seeding is hypothesized to depend on the structural similarity between fibrils of hIAPP and amyloid- β . Previous structural comparisons between hIAPP and amyloid- β fibrils have supported this hypothesis [12:15:16](#). Here, we perform comparisons while taking special note of the conserved core folds (CF1 and CF2). These cores, CF1 (residues 15–28) and CF2 (residues 20–33), each cover the region of maximal sequence similarity between hIAPP (19–29) and amyloid- β (24–34) (Fig. [3e](#)), raising the question of whether analogous structural cores also exist in amyloid- β . We use TW1 chain A and chain B to represent protofilaments with CF1 and CF2, respectively, and, together with TW2 (CF3), we compare them with all available amyloid- β fibril structures. We confined our superpositions to the residue ranges exhibiting maximal sequence identity as noted above (Supplementary Table [3](#); see [Methods](#) for details). The comparisons reveal that each hIAPP core fold exhibits a root mean square deviation (r.m.s.d.) under 1.8 Å with some amyloid- β fibril structure. The best amyloid- β matches for each of the three groups of hIAPP protofilaments are as follows: [6OIZ](#) with TW1 chain A, [2M4J](#) with TW1 chain B and [5OQV](#) and [6SHS](#) with TW2 (Fig. [3e](#)). These results suggest that hIAPP and amyloid- β may be able to cross-seed each other through these regions of sequence and structural similarity.

Discussion

Do our hIAPP structures recapitulate T2D patient seeds?

In this study, we have determined the cryo-EM structures of four hIAPP polymorphs from fibrils seeded by patient-extracted seeds. Our structure TW3 is similar to [6Y1A15](#) and [6ZRF16](#) from unseeded controls. Structures TW1, TW2 and TW4 have not been reported before. Do these three seeding-specific polymorphs represent the presently unknown conformations of patient-extracted seeds, that is, of pathogenic hIAPP fibrils (Supplementary Text [4](#))?

A warning that seeded structures do not always replicate the structures of the seeds comes from a recent seeding study by Lövestam and colleagues³³. Their work suggests that fibrils of recombinant α -synuclein seeded by patient extracts do not inevitably replicate the structures of the seeds³³. In this α -synuclein study, two parallel seeding attempts using different patient extracted seeds generated seeded fibrils, whose structures were then determined. The structures were composed of two types of protofilament that were either essentially identical or close to that found in unseeded recombinant α -synuclein fibrils. However, a third parallel seeding attempt generated fibrils that largely replicated one out of two protofilaments of the seeds³³. We believe these results suggest that there is a risk in seeding where the seeded protein may form fibrils that adopt the unseeded conformation, even in the presence of seeds. The third attempt by Lövestam et al. demonstrated that the structure of the seeds from the daughter fibrils can be obtained, at least at the level of the core fold of the protofilament. This emphasizes the importance of comparing the seeded fibril structures with unseeded controls at the protofilament level to judge which protofilament may represent the conformation of the seeds.

When we investigate the three new polymorphs found only in seeded hIAPP fibrils (TW1, TW2 and TW4) at the protofilament level, we find that TW2 adopts a unique conformation (CF3). We thus believe that CF3 is the most likely conformation to represent the seeds. By contrast, TW1 and TW4 both contain CF1 and CF2, which also exist in other unseeded hIAPP fibrils. We think this finding does not diminish the possibility that the structures of TW1 and TW4 are inherited from pathogenic hIAPP fibrils. Our reason is that, although TW1 and TW4 both contain a protofilament with CF1 that has been observed in the unseeded controls [6Y1A15](#) and [6ZRF16](#), TW1 and TW4 differ from these unseeded controls in that they also contain a protofilament with CF2. Therefore, the protofilaments that adopt CF2 can be considered as structural elements derived from the pathogenic seeds. We note that CF2 is also found in other unseeded hIAPP fibrils when either S20G (in [6ZRR](#) and [6ZRQ16](#)) or SUMO-tagged (in [6VW212](#)) hIAPP is used as the starting monomer. Therefore, we believe that CF2 can be adopted either through seeding with pathogenic fibrils or by changing the form of hIAPP used for unseeded fibril growth.

Additional factors may mitigate concern about the seeding fidelity in the present study. First, Lövestam et al. suggest that their failure to replicate structures via seeding may be due to the truncation and post-translational modification (PTM) of pathogenic α -synuclein fibrils³³. These PTMs include ubiquitination and phosphorylation and are considered to favor α -synuclein fibril formation in patients, as does C-terminal truncation^{34,35}. Apparently, when recombinant α -synuclein without PTMs or truncation is used for seeding, the daughter fibrils do not adopt the same structure as the seeds. However, PTMs other than the C-terminal amidation, which is present in our hIAPP monomer, or truncation are not required for hIAPP amyloidosis³⁶. Furthermore, in contrast to previously determined, unseeded hIAPP fibril structures that mostly contain the C-terminal portion of hIAPP in the fibril core with the first visible residue around

13–15 (Extended Data Fig. [5a](#)), residues 6–12 are visible in TW2 and chain A of TW4 and adopt similar conformations in both structures (Extended Data Fig. [5b](#), inset). The conformation of residues 1–5 is also indicated by weak densities in TW2 and chain A of TW4 (Extended Data Fig. [5b](#)). The observation of the longer cores in TW2 and TW4 may also support their potential pathological origin, because in vivo fibril formation is much slower than in vitro, which will give the N terminus enough time to occupy a definite conformation. This hypothesis is further supported by the observation that α -synuclein fibrils extracted from patients' brains have larger ordered fibril cores than fibrils formed in vitro¹⁸. In short, the lack of sequence modifications, as well as larger ordered fibril cores may both suggest better fidelity of seeded replication of the hIAPP structure. In addition, hIAPP amyloid is linked with β -cell damage in T2D^{37,38}, and the observation that our seeded hIAPP fibrils are toxic to pancreas cells (Extended Data Fig. [9b](#)) may add to the potential physiological relevance of this study.

We note that if our structures indeed replicate the conformations of pathogenic hIAPP fibrils, these fibrils are extracted from only one T2D patient. Although fibrils formed by the same protein and formed in different patients with the same disease were shown to adopt the same structures in tau and α -synuclein studies^{18,39,40,41}, it is still unknown if this is so for hIAPP fibrils in T2D. Further study with multiple patient samples is needed to address this question.

The origin of heterotypic pairings of hIAPP conformations

A notable finding in this study is that two distinct conformations of protofilaments pair together in two of our fibrils, TW1 and TW4 (Figs. [2](#) and [3b](#), Extended Data Fig. [4a](#), Supplementary Text [5](#) and Extended Data Fig. [8](#)). Amyloid fibril structures with two protofilaments of different conformation have been reported several times (for example, recombinant H50Q α -synuclein fibrils⁴² and brain extracted α -synuclein fibrils¹⁸), but the difference in conformation between

protofilaments has never been as large as we observe in TW1 and TW4 in this study and [6ZRR](#) from synthetic S20G hIAPP¹⁶. We think these unique fibrils combining different conformations are the result of hIAPP's propensity to adopt different core folds, and the ability of these different conformations to interface via different surfaces (Extended Data Fig. [5f](#)).

Conclusion

In summary, in this study we have seeded fibril formation of synthetic hIAPP monomers with fibrils extracted from islet cells of a T2D patient and determined four cryo-EM structures of the seeded fibrils. Analysis of these structures and comparison with previously reported hIAPP fibril structures suggest the structural elements (CF3 and heterotypic assembly of CF1 and CF2) that are related to pathogenic hIAPP fibrils. Comparisons of wild type to S20G hIAPP fibril structures as well as hIAPP to amyloid- β fibril structures reveal the possible mechanism of how the hereditary mutation S20G facilitates hIAPP fibrilization in T2D and the regions that may enable hIAPP and amyloid- β cross-seeding.

Methods

Islet cells from donors with T2D

Islet cells from seven donors with T2D were provided by City of Hope as formalin-fixed paraffin-embedded slices to test the existence of amyloid deposits by Congo red staining. The basic information about each donor is listed in Supplementary Table [1](#). After Congo red staining, we selected the donor with the highest staining signal (donor 6), and frozen islet cells from this donor were further requested from City of Hope and used for genotyping and fibril extractions.

Congo red staining of fixed patient slices

Formalin-fixed paraffin-embedded slices of islet cells (thickness of 5 μm) were baked at 45 $^{\circ}\text{C}$ for 20 min for deparaffinization. The slices were then soaked three times in xylene (Fisher Scientific) for 5 min, then washed with 100%, 95%, 70% and 50% ethanol for 4 min, 4 min, 2 min and 2 min, respectively. The washed slices were soaked in water for 10 min on a shaker. The slices were Congo-red-stained using an amyloid stain, Congo red kit (Sigma-Aldrich), following the kit instructions. After Congo red staining, the slices were dipped twice in hematoxylin (Thermo Scientific) and rinsed in water for counter staining. After staining, the slices were washed three times with 100% ethanol for 5 min, rinsed in xylene, mounted with Permount mounting medium (Fisher Scientific) and allowed to dry for several hours. Stained slices were imaged by an AXIO Observer D1 fluorescence microscope (Zeiss) using bright-field and polarized light.

Genotyping of donor 6 hIAPP

To identify the genotype of the hIAPP from donor 6, we extracted genome DNA from the islet cells using the following protocol. Frozen islet cells were resuspended with 400 μl of SNET buffer (20 mM Tris pH 8.0, 5 mM EDTA, 1% SDS, 400 mM NaCl) supplemented with 0.5 mg ml^{-1} proteinase K, and the solution was incubated at 50 $^{\circ}\text{C}$ with occasional shaking for 1 h. After incubation, the solution was mixed with 400 μl phenol chloroform, vortexed for 15 s, and centrifuged at 14,000g for 30 s. The top aqueous layer of the mixed solution was removed and mixed with 40 μl of 3 M sodium acetate pH 5.4, 0.5 μl GlycoBlue (Thermo Scientific) and 750 μl isopropanol. The mixed solution was incubated at -20°C for 20 min, then centrifuged at 14,000g for 30 s. The supernatant was removed and the pellet was washed with 400 μl ethanol and further centrifuged at 14,000g for 30 s. The supernatant was discarded and the tube was air-dried at room temperature for 20 min. Finally, 50 μl distilled water was added into the tube to

dissolve the pellet, and the solution was used as template for polymerase chain reaction (PCR) using in-house-designed primers to amplify the DNA fragments that contained hIAPP genomic DNA (gDNA). The PCR products were sequenced by Laragen (USA).

The sequencing experiments were performed independently three times using three independent PCR products, and all results show that the genotype of hIAPP from donor 6 is the wild type.

One of the representative sequencing results is given in the following, where the sequence of a part of the genome DNA that contains hIAPP gDNA is shown:

```
...AAAAAAAAATCTCAGCCATCTAGGTGTTTGCAAACAAAACACTGAGTTACTTAT
GTGAAAATTGTTTCTTTGGTTTTTCATCAATACAAGATATTTGATGTCACATGGCTGGA
TCCAGCTAAAATTCTAAGGCTCTAACTTTTCACATTTGTTCCATGTTACCAGTCATCA
GGTGGAAAAGCGGAAATGCAACACTGCCACATGTGCAACGCAGCGCCTGGCAAATT
TTTTAGTTCATTCCAGCAACAACCTTTGGTGCCATTCTCTCATCTACCAACGTGGGATC
CAATACATATGGCAAGAGGAATGCAGTAGAGGTTTTAAAGAGAGAGCCACTGAATT
ACTTGCCCCTTTAGAGGACAAT...
```

Fibrils extraction from islet cells and immunoprecipitation

The fibril extraction was performed according to a previous protocol⁴³. Frozen islet cells were resuspended with 700 µl of saline buffer (0.15 M NaCl), homogenized and centrifuged at 19,000g at 4 °C for 20 min. The supernatant was removed (referred to as the S1 fraction) and the pellet was resuspended with 400 µl saline buffer, homogenized and centrifuged as in the previous step. The supernatant was also removed (referred to as the S2 fraction) and this 400-µl saline wash step was repeated another five times (generating S3–S7 fractions). The pellet was further washed three times with 400 µl H₂O with the same operation (generating S8–S10 fractions). The

final pellet was lyophilized and then redissolved with 100 μ l TBS buffer (20 mM Tris pH 7.6, 150 mM NaCl) and is referred to as the pellet (P). We found that the hIAPP fibrils are enriched mainly in the S1 fraction but not in the pellet, as described previously⁴³ by dot-blot assays and negative-stain EM (Extended Data Fig. [1b,c](#)). We think this discrepancy is a consequence of the different tissues used for fibril extraction. In the previous study, frozen pancreatic tissues were used⁴³, whereas we used purified islet cells instead. In our case, the fibrils may be isolated much more easily, remaining in the supernatant of the first homogenization step. To assess the purity of the S1 fraction, we performed sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and ran a western blot, probing with antibodies that target hIAPP, amyloid- β , tau or α -synuclein (Extended Data Fig. [9a](#)). The results suggest that only hIAPP fibrils were present in the S1 fraction, enabling further purification of the S1 fraction via immunoprecipitation with fibril-binding OC antibody.

Before immunoprecipitation, we treated the S1 fraction with collagenase by mixing 60 μ l of the S1 fraction with 60 μ l TBS buffer supplemented with 4 μ M CaCl₂ and 50 U ml⁻¹ collagenase (Sigma). We then incubated the mixed solution for 1 h at 37 °C. To pull down amyloid fibrils, we first prepared the antibody-coupled magnetic beads by incubating 2 μ g of anti-amyloid fibrils OC antibody (Millipore, lot no. 2345063) with 50 μ l of protein A magnetic beads (Invitrogen) and diluted them in TBST buffer (20 mM Tris pH 7.6, 150 mM NaCl, 0.1% Tween 20) to a final volume of 700 μ l, then the mixed solution was incubated on a tube rotator at room temperature for 1 h. After incubation, we collected the magnetic beads, discarded the supernatant, mixed the beads with 120 μ l of collagenase-treated S1 fraction, and diluted the solution to final volume of 200 μ l with TBS buffer. The final solution was incubated on a tube rotator at room temperature for 2 h, and the beads were then collected (the supernatant was removed and is referred to as the

flowthrough fraction). The collected beads were first eluted with 150 μ l TBST buffer (elute-1 fraction), then eluted twice with 50 μ l of 100 mM glycine pH 2.0, and neutralized with 1 M Tris-HCl pH 9.0 (elute-2 and elute-3 fractions). The eluted beads were resuspended with 50 μ l TBS and are referred to as the beads fraction. We found that the hIAPP fibrils were enriched in the elute-1 fraction, as judged by three criteria: (1) the observation of amyloid fibrils in electron micrographs, (2) the competence of the elute-1 fraction in seeding the aggregation of fresh hIAPP monomers and (3) positive anti-hIAPP and OC signals in dot blots compared to the absence of primary antibody (Fig. [1b,c](#), Extended Data Fig. [1d-f](#) and Supplementary Text [6](#)).

We note that the existence of hIAPP fibrils in the elute-2 fraction cannot be fully ruled out by dot blot, as the elute-2 fraction already contained a certain amount of OC antibody (as shown in the ‘no primary’ line in Extended Data Fig. [1d](#) as a negative control) and the signal might be saturated, whereas in EM and seeding assays the elute fractions did not suggest the presence of hIAPP fibrils compared to the elute-1 fraction (Extended Data Fig. [1d-f](#)). We believe the reason that hIAPP fibrils were not enriched in the elute-2 or elute-3 fractions is that the binding between hIAPP fibrils and OC antibody is not very strong, so that after the hIAPP fibrils were pulled down from the solution they could be easily separated from the antibody when a slight buffer change occurred during the elute-1 step (from half diluted TBS to TBST, see above for details), before a larger buffer change occurred at the elute-2 or elute-3 step.

Dot blot and western blot

In our dot-blot assay, 2 μ l of each sample solution was pipetted onto multiple nitrocellulose membranes (Invitrogen) in parallel and allowed to dry for 15 min. The membranes were probed by amylin polyclonal antibody (also referred to as anti-hIAPP antibody in this study, Invitrogen, lot no. UH2833361, 1:1,000 dilution) or anti-amyloid fibril OC antibody (Millipore, lot no.

2345063, 1:5,000 dilution) respectively, and further probed by anti-rabbit immunoglobulin-G (Sigma, lot no. 023M4757, 1:5,000 dilution) as the secondary antibody, and visualized with an ECL Plus western blotting substrate kit (Thermo Scientific). For an explanation for the high intensity in the ‘no background’ dot blot of elute-2 (Extended Data Fig. [1d](#)), see Supplementary Text [6](#).

In Extended Data Fig. [9a](#), we used western blot to analyze the molecular weight of the S1 fraction. The loading dye contained 2 M urea. Proteins were transferred to nitrocellulose membranes (Invitrogen) and detected by immunoblot analysis with amylin polyclonal antibody (Invitrogen, lot no. UH2833361, 1:1,000 dilution), anti-tau antibody (Agilent Dako, lot no. 20024929, 1:5,000 dilution), anti-amyloid- β antibody 6E10 (Biolegend, lot no. B261546, 1:5,000 dilution) and anti- α -synuclein antibody (MJFR-14-6-4-2, Abcam, lot no. GR3222269-10, 1:5,000 dilution). As a secondary antibody we used goat anti-rabbit horseradish peroxidase (HRP; Invitrogen, lot no. 2116291, 1:5,000 dilution) or goat anti-mouse HRP (Abcam, lot no. GR3271082-2, 1:5,000 dilution). All membranes were developed using Pierce ECL Plus substrate (Thermo Scientific). We used positive controls to validate the antibodies for amyloid- β 1–42 (expected size ~4 kDa, for anti-amyloid- β antibody), tau-K18 (residues 244–372, expected size ~15 kDa, for anti-tau antibody) and α -synuclein 1–140 (expected size ~14 kDa, for anti- α -synuclein antibody).

Negative-stain transmission electron microscopy

Negative-stain transmission EM samples were prepared by applying 5 μ l of solution to 400-mesh carbon-coated formvar support films mounted on copper grids (Ted Pella). The grids were glow-discharged for 30 s before applying the samples. The samples were incubated on the grid for 2 min and then blotted off with filter paper. The grids were stained with 3 μ l of 2% uranyl acetate

for 1 min and washed with an additional 3 μ l of 2% uranyl acetate and allowed to dry for 1 min. The grids were imaged using a T12 (FEI) electron microscope.

Synthetic hIAPP preparation

Synthetic full-length hIAPP peptide was ordered from InnoPep with an amidated C terminus and a disulfide bond between Cys2 and Cys7, and the purity was >95%. The peptide was dissolved in 100% 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) at a concentration of 1 mM, sonicated at 4 $^{\circ}$ C for 1 min, and incubated at room temperature for 5 h. After incubation, HFIP was evaporated by a CentriVap concentrator (Labconco) and the peptide was stored at -20° C. Before use, HFIP-treated peptide was dissolved in 100% DMSO at a concentration of 1 mM or 5 mM, and further diluted 100-fold in phosphate-buffered saline (PBS) and filtered with 0.1 μ m Ultrafree-MC-VV centrifugal filters (Millipore) to form 10 μ M and 50 μ M hIAPP solutions.

Thioflavin T (ThT) seeding assays

Synthetic hIAPP was diluted to 10 μ M or 50 μ M in PBS supplemented with 30 μ M ThT, filtered with 0.1- μ m Ultrafree-MC-VV centrifugal filters (Millipore) and mixed with 1% or 5% (vol/vol) of seeds, respectively. The seeds were sonicated at 4 $^{\circ}$ C for 3 min before use. Mixed solution was pipetted into a Polybase Black 384-well plate with an optical bottom (Thermo Scientific) and incubated at 37 $^{\circ}$ C without shaking. ThT fluorescence was monitored by a FLUOstar Omega plate reader (BMG LABTECH) with excitation and emission wavelengths of 440 nm and 480 nm, respectively. The fluorescence curves were averaged from three independent measured replicates, and error bars show s.d. of replicate readings. To normalize the different ranges of fluorescence readings observed from different experiments (probably due to the different fluorescence gain settings of the plate reader), we normalized the readings to make the minimum

mean value in each panel 0% and the maximum mean value in each panel 100%. To test the seeding ability of different fractions from the immunoprecipitation assay, 10 μ M hIAPP solution was mixed with 1% (vol/vol) of the fractions (elute-1, elute-2 and elute-3) as fibril growing seeds (results are shown in Extended Data Fig. [1f](#)). To monitor the seeding effect at the same concentration as growing fibrils for cryo-EM structure determination, 50 μ M hIAPP solution was mixed with or without 5% (vol/vol) of the elute-1 fraction from immunoprecipitation as the patient extract (shown in Fig. [1c](#)). ThT curves were generated using GraphPad Prism.

Fibril growth and seeding

To obtain the high fibril yields necessary for cryo-EM studies, we raised the concentration of hIAPP monomers from the 10 μ M used in the seeding assays shown in Extended Data Fig. [1f](#) to 50 μ M. We were concerned that the elevated concentration of hIAPP would enhance the growth of unseeded amyloid fibrils and divert monomer from being templated by the seed. However, we found that seeding at this elevated hIAPP concentration still had a strong effect. It notably increased the aggregation rate and ThT signal compared to the monomers incubated in the absence of seeds (Fig. [1c](#)). On imaging samples prepared by 20 h of incubation, we found the monomers incubated with seeds show mature and elongated fibrils, whereas the monomers incubated alone show immature fibrils (Fig. [1d](#)). This observation further suggested the efficacy of seeding.

Freshly diluted and filtered 50 μ M hIAPP solution in PBS was mixed with 5% (vol/vol) patient extract as seeds. The seeds were sonicated at 4 °C for 3 min before use. The solution was then incubated at 37 °C without shaking for 20 h. After incubation, the solution was centrifuged at 4,000g for 1 min, and 95% of the original volume of supernatant was carefully removed from the centrifuge tube to concentrate the fibril solution 20 times. The remaining fibrils were

resuspended by mixing with a pipette and then used for cryo-EM data collection. We tested the cytotoxicity of the seeded fibrils used in cryo-EM by 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT) assays (Extended Data Fig. [9b](#)).

Cryo-electron microscopy data collection and processing

To prepare the grids for cryo-EM data collection, we applied 2.6 μl of fibril solution onto a Quantifoil 1.2/1.3 electron microscope grid, which had been glow-discharged for 2 min before use. The grid was plunge-frozen in liquid ethane with a Vitrobot Mark IV instrument (FEI). Two datasets were collected on the same 300-kV Titan Krios (FEI) microscope with a Gatan K3 camera located at the HHMI Janelia Research Campus, and were collected from grids that were made in parallel from the same batch of fibril samples. The microscope was operated with 300 kV acceleration voltage and a slit width of 20 eV. Videos were collected using super-resolution mode with a nominal physical pixel size of 1.065 \AA per pixel (0.5325 \AA per pixel in super-resolution video frames) with a dose per frame of $\sim 1.5 \text{ e}^-/\text{\AA}^2$. Fifty-seven frames were taken for each video for the first dataset (total dose per image $\sim 85.5 \text{ e}^-/\text{\AA}^2$) and 30 frames were taken for each video for the second dataset (total dose per image of $\sim 45 \text{ e}^-/\text{\AA}^2$). Automated data collection was driven by SerialEM[44](#). Anisotropic magnification distortion estimation, CTF estimation and beam-induced motion correction were performed with mag-distortion-estimate[45](#), CTFFIND 4.1.8[46](#) and Unblur[47](#), respectively. The physical pixel size was corrected to 1.055 \AA per pixel after anisotropic magnification correction with Unblur[47](#).

Particle picking was performed manually using EMAN2 e2heliboxer.py[48](#) for the first dataset, and performed automatically using CrYOLO[49](#) for the second dataset. Particle extraction, 2D classification, helical reconstruction and 3D refinement were performed in RELION[50:51](#).

Particles were extracted using an inter-box distance of 14.1 \AA and a box size of 768 pixels that

downscaled to a box size of 384 pixels (bin factor equal to 2). We performed 2D classification with tau_fudge factor set to 2, and from 2D classes we were able to identify four morphologies with twisting features and another four morphologies without twisting features (TW1–4 and NT1–4, respectively; Fig. 2 and Extended Data Fig. 2). The disruption of each morphology is shown in Fig. 2 (inset). We were only able to pursue the morphologies with twisting features due to the limitation of helical reconstruction. We selected the particles that belonged to 2D classes of TW1–4 and re-extracted them (individually for each morphology) into a box size of 256 pixels (for TW1, TW3 and TW4, which have relatively short pitch) or 384 pixels (for TW2, which has relatively long pitch). No downscaling was used for re-extraction, and the inter-box distance was kept to 14.1 Å. The helical parameters of each twister morphology were estimated from the measured crossover distance and used for further 3D classification. We performed 3D classification with $K = 3$ (or $K = 5$ for TW2) and a Gaussian cylinder initial model as described previously¹². Particles belonging to the best 3D class were selected, and several more rounds of 3D classification ($K = 3$) were performed for each TW morphology to make sure the best particles were selected. The tau_fudge factor was set to 4 at the beginning of each 3D job, and was increased to 64 to push the resolution of the reconstructions. We then used the particles selected from the 3D classification to perform high-resolution gold-standard refinement⁵¹, and the resolution of each reconstruction was estimated using the 0.143 Fourier shell correlation (FSC) resolution cutoff. Before performing gold-standard refinement of TW2, we re-extracted the selected particles from a pixel box size of 384 to a pixel box size of 256 to obtain a higher-quality final map.

Previous structural studies of unseeded synthetic wild-type hIAPP fibrils show structures that are similar to TW3 (PDB [6Y1A15](#) and [6ZRF16](#); Extended Data Fig. 4b), but their two structures

were reconstructed with different handedness (right-handed twist for [6Y1A15](#) and left-handed twist for [6ZRF16](#)). Because, in this study, we were not able to achieve resolutions high enough to identify the handedness of each morphology, here we performed reconstructions by assuming all the TW morphologies are left-handed, as the left-handed twist is most common for amyloid fibrils. We cannot rule out the possibility that one or several of the TW morphologies solved here are right-handed. We believe that handedness should not influence the model we build for each morphology, as in the current resolution the same fold can be generated with either handedness, which is supported by the observation that [6Y1A](#) and [6ZRF](#) adopt a similar structure but have different handedness. We thus believe that determining the handedness should be beyond the scope of this study, as we mostly focus on the structures within each layer of the fibrils.

Atomic model building

The refined maps were sharpened with `phenix.auto_sharpen`[52](#) at the resolution cutoff indicated by half-map FSC. The atomic model of each TW morphology was manually built into the sharpened map using `COOT`[53](#). For TW1, the map clearly suggests that a single layer contains two asymmetric chains. To build each chain, we first located the C-terminal tyrosine residue (Tyr37) at the end of the density tube that does not have extra fuzzy density. The remaining residues were built de novo according to their position in sequence relative to Tyr37. The resulting model fits the map unambiguously for all side chains (except a bump in TW1, Supplementary Text [7](#)). Atomic model building of TW2 to TW4 is also described in Supplementary Text [7](#) (Extended Data Fig. [10](#)).

The initial models were extended to five layers (10 chains) based on the helical symmetry of each TW morphology, and the five-layer models were refined by `phenix.real_space_refine`[54](#) for several rounds. As the last step, the rotamer of each serine, glutamine and asparagine residue was

manually inspected to favor hydrogen bonds, and the final model was validated using MolProbity⁵⁵.

Structural alignment of all reported hIAPP fibril structures

With the four structures reported in this study there are currently nine near-atomic resolution structures of hIAPP fibrils (Table 2). To reduce the amount of structures used for comparison, we used TW3 to represent [6Y1A15](#) and [6ZRF16](#), because these three structures are very similar to each other (Extended Data Fig. 4b). Furthermore, two morphologies from hIAPP S20G fibrils have an almost identical conformation of chain A and B ([6ZRR](#) and [6ZRQ16](#); [6ZRR](#) has an additional chain C in each layer), so we used [6ZRR](#) to represent [6ZRQ](#). Here, we performed structural alignment of these structures at the single-chain level. That is, for those structures that contain two identical monomers in each layer (such as TW3), only chain A was used for structural alignment; for structures that contain two monomers with different structures (such as TW1), both chains A and B were used for structural alignment; for [6ZRR](#), which contains two identical monomers (chains A and B) and one additional monomer with different fold (chain C), chain A and C were used for structure alignment. The alignment was done via PyMOL⁵⁶. Initially, all residues in each chain were used for structural alignment, and after all chains had been categorized into CF1 and CF2, all chains with a CF1 fold were re-aligned with the residue range restricted to 15–28 and those with a CF2 fold were re-aligned with the residue range restricted to 20–33.

We note that our previously reported SUMO-tagged recombinant hIAPP fibril structure contained CF2 when the ‘swapped’ version was used for structural alignment. The swapped version of [6VW2](#) was proposed in our previous structure study of hIAPP fibrils¹², and is formed by swapping the residues from two chains in a single layer at Gly24 that are very close to each

other (Extended Data Fig. [6](#); red dashed line in Fig. [3b](#)); this domain swap is predicted to be chemically and energetically plausible, although not supported by the density map of [6VW212](#).

Note that we only use cryo-EM structures of hIAPP for structural comparison and analysis here, but there are models of hIAPP fibrils reported previously that were generated by different methods [13:14](#). These alternate models, especially the one from the NMR study, may represent the structure of the ribbon morphology of hIAPP fibrils, which cannot be readily determined by cryo-EM. In our seeded fibrils, we also observed four morphologies of the ribbon form of hIAPP fibrils (NT1–4, Extended Data Fig. [2a,b](#)), which together comprised 35% of the total population and were not amenable for cryo-EM structure determination. We cannot rule out the possibility that the structures of NT1–4 may be similar to the NMR model.

Structural alignment of hIAPP and amyloid- β fibril structures

The PDB IDs of the amyloid- β structures used for structural alignment are [6OIZ57](#), [2M4J58](#), [2MVX59](#), [5KK360](#), [5OQV61](#), [2NAO62](#), [2MXU63](#), [2BEG64](#), [2LMN65](#), [2MPZ66](#) and [6SHS67](#). Structural alignment was done via PyMOL [56](#). We aligned each amyloid- β structure with TW1 chain A, chain B and TW2 using residues 24–34 in each amyloid- β structure and residues 19–29 in hIAPP fibril structures. The r.m.s.d. values for each alignment are summarized in Supplementary Table [3](#).

MTT assays

Rin5F cells were purchased from ATCC (cat. no. CRL-2058), then 90 μ l of cells were plated at 60,000 cells per ml, in 96-well plates (cat. no. 3603, Costar, Fisher Scientific) in DMEM medium (Life Technologies, cat. no. 11965092) supplemented with 10% (vol/vol) FBS (Life Technologies, cat. no. A3160401), 1% penicillin/streptomycin (Life Technologies, cat. no.

15140122) and 1% Glutamax (Life Technologies, cat. no. 35050061) at 37 °C, 5% CO₂, in a humidified incubator. To perform MTT assays, patient-fibril-seeded hIAPP fibrils at a concentration of 50 μM were pelleted by centrifugation at 21,000g for 1 h and supernatant was removed. The pellet was suspended in sterile PBS buffer and sonicated for 5 min, then 10 μl of sample solution was added to cells at various concentrations (0, 1, 10 and 50 μM). Experiments were performed in triplicate. After a 24 h incubation of samples with cells, 20 μl of Thiazolyl Blue tetrazolium bromide MTT dye (Sigma) was added to each well and incubated for 3.5 h at 37 °C under sterile conditions. The MTT dye stock was prepared by dissolving 5 mg ml⁻¹ in sterile PBS buffer. The MTT assay was stopped by carefully aspirating off the culture medium and adding 100 μl of 100% DMSO to each well. Absorbance was measured at 570 nm and the background reading was recorded at 700 nm and subsequently subtracted from the value at 570 nm.

Statistics and reproducibility

The western blot, ThT and cytotoxicity experiments were replicated at least three times with similar results. We did not replicate the Congo red staining. One negative-stain EM grid was prepared and imaged for each sample, whereas multiple (more than three) micrographs with similar results were observed for each grid.

Data availability

Structural data have been deposited into the Worldwide Protein Data Bank (wwPDB) and the Electron Microscopy Data Bank (EMDB) with the following accession codes:

PDB [7M61](#), [EMD-23686](#) (TW1); PDB [7M62](#), [EMD-23687](#) (TW2); PDB [7M64](#), [EMD-23688](#) (TW3); PDB [7M65](#), [EMD-23689](#) (TW4). PDB accession codes for previously reported

coordinates used for structural analysis in this study are: [6Y1A](#), [6ZRR](#), [6ZRQ](#), [6ZRF](#), [6VW2](#) for hIAPP fibrils

and [6OIZ](#), [2M4J](#), [2MVX](#), [5KK3](#), [5OQV](#), [2NAO](#), [2MXU](#), [2BEG](#), [2LMN](#), [2MPZ](#), [6SHS](#) for amyloid- β fibrils. All data are available in the paper or the [Supplementary](#) Information.

Code availability

Energetic calculations were performed using custom written software. The code is available at the MBI website

(https://people.mbi.ucla.edu/sawaya/amyloidatlas/accessibleurfacearea_v07.2d.f).

Acknowledgements

We thank the Southern California Islet Cell Resources Center for providing human islets for this study. We thank X. Zhao at the HHMI Janelia Cryo-EM Facility for help with microscope operation and data collection. We acknowledge support from NIH AG 054022, NIH AG061847 and DOE DE-FC02-02ER63421. D.R.B. was supported by the National Science Foundation Graduate Research Fellowship Program.

Author contributions

Q.C. designed experiments and performed data analysis. F.K. prepared islet cells from donors. Q.C. and L.S. performed Congo red staining of islet cells. Q.C., L.S. and B.A.N. performed fibril extraction from islet cells. Q.C. and R.A. performed immunoprecipitation in fibril extraction. R.A. and J.L. performed western blot and MTT assays. K.A.M. helped with western blots. Q.C. prepared hIAPP fibrils and cryo-EM grids. Q.C. and D.R.B. collected cryo-EM data. Q.C. performed cryo-EM data processing and model building, J.L. assisted with particle picking. Q.C.

and M.R.S. performed solvation energy calculations. All authors analyzed the results and wrote the manuscript. D.S.E. supervised and guided the project.

Competing interests

D.S.E. is an advisor and equity shareholder in ADRx, Inc. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41594-021-00646-x>.

Supplementary information The online version contains supplementary material

available at <https://doi.org/10.1038/s41594-021-00646-x>.

Correspondence and requests for materials should be addressed to David S. Eisenberg.

Peer review information *Nature Structural & Molecular Biology* thanks Sjors Scheres and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Figures

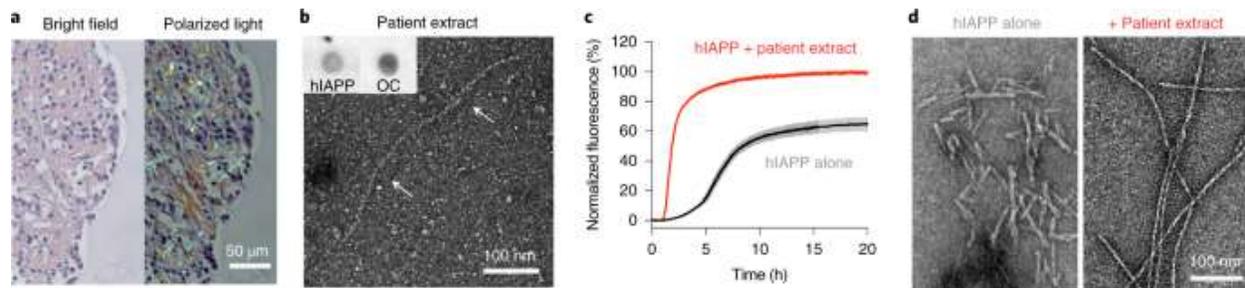


Fig. 1: Extraction of patient islet cells and seeding of fibril growth.

a, Congo red staining of a slice of islet cells from a donor with T2D (donor 6). **b**, Negatively stained EM image of the elute-1 fraction in the immunoprecipitation assay. Amyloid fibrils are indicated by arrows. Insets: dot blots of the same fraction probed by anti-hIAPP (amylin polyclonal antibody) and anti-amyloid fibril OC antibodies. **c**, ThT aggregation curves of freshly prepared synthetic hIAPP peptide incubated with (red) and without (black) the patient extracted fibrils as fibril growth seeds. Data are shown as mean \pm s.d., $n = 3$ independent experiments. **d**, Negatively stained EM images of the same samples as in **c** after 20 h of incubation.

Table 1 | Cryo-EM data collection, refinement and validation statistics

	TW1 (EMD-23686, PDB 7M61)	TW2 (EMD-23687, PDB 7M62)	TW3 (EMD-23688, PDB 7M64)	TW4 (EMD-23689, PDB 7M65)
Data collection and processing				
Magnification	×64,000	×64,000	×64,000	×64,000
Voltage (kV)	300	300	300	300
Electron exposure (e ⁻ /Å ²)	45–85	45–85	45–85	45–85
Defocus range (μm)	0.7–2.7	0.7–2.7	0.7–2.7	0.7–2.7
Pixel size (Å)	1.055	1.055	1.055	1.055
Symmetry imposed	C ₁	C ₂	C ₁	C ₁
Helical rise (Å)	4.81	4.80	2.40	4.81
Helical twist (°)	–2.94	178.42	178.33	–2.94
Initial particle images (no.)	339,132	953,276	285,645	170,541
Final particle images (no.)	26,608	23,957	17,853	20,365
Map resolution (Å)	3.8	3.9	4.0	4.1
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	200–3.8	200–3.9	200–4.0	200–4.1
Refinement				
Initial model used (PDB code)	De novo	De novo	6Y1A	De novo
Model resolution (Å)	3.7	4.1	4.0	4.0
FSC threshold	0.5	0.5	0.5	0.5
Model resolution range (Å)	200–3.7	200–4.1	200–4.0	200–4.0
Map sharpening β factor (Å ²)	83	98	155	100
Model composition				
Nonhydrogen atoms	1,690	2,380	1,850	2,025
Protein residues	230	320	250	275
Ligands	0	0	0	0
B factors (Å²)				
Protein	45.8	150.3	67.7	92.4
Ligand	-	-	-	-
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.008	0.009
Bond angles (°)	1.166	1.126	1.150	1.506
Validation				
MolProbity score	2.51	2.78	2.30	2.70
Clashscore	23.6	32.8	19.5	31.1
Poor rotamers (%)	0	0	0	0
Ramachandran plot				
Favored (%)	85.71	76.67	91.3	82.35
Allowed (%)	14.29	23.33	8.7	17.65
Disallowed (%)	0	0	0	0

Table 2 | Comparative solvation energy calculations

Fibril structure	Fibril information	Energy per layer (kcal mol ⁻¹)	Energy per residue (kcal mol ⁻¹)
TW1 (PDB 7M61)	Synthetic wild-type hIAPP peptide seeded by patient extract	-24.4	-0.53 (chain A, -0.58; chain B, -0.48)
TW2 (PDB 7M62)	Synthetic wild-type hIAPP peptide seeded by patient extract	-31.5	-0.49
TW3 (PDB 7M64)	Synthetic wild-type hIAPP peptide seeded by patient extract	-27.0	-0.54
TW4 (PDB 7M65)	Synthetic wild-type hIAPP peptide seeded by patient extract	-26.7	-0.49 (chain A, -0.49; chain B, -0.48)
6Y1A	Synthetic wild-type hIAPP peptide aggregated in vitro	-23.2	-0.46
6VW2	SUMO-tagged recombinant hIAPP aggregated in vitro	-21.6	-0.45
6ZRF	Synthetic wild-type hIAPP peptide aggregated in vitro	-24.0	-0.48
6ZRQ	Synthetic hIAPP S20G peptide aggregated in vitro	-25.1	-0.55
6ZRR	Synthetic hIAPP S20G peptide aggregated in vitro	-37.1	-0.54 (chain A, -0.53; chain B, -0.62; chain C, -0.46)

References

1. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
2. Roberts, A. N. et al. Molecular and functional characterization of amylin, a peptide associated with type 2 diabetes mellitus. *Proc. Natl Acad. Sci. USA* **86**, 9662–9666 (1989).
3. Westermark, P. Amyloid in the islets of Langerhans: thoughts and some historical aspects. *Ups. J. Med. Sci.* **116**, 81–89 (2011).
4. Westermark, P. et al. Amyloid fibrils in human insulinoma and islets of Langerhans of the diabetic cat are derived from a neuropeptide-like protein also present in normal islet cells. *Proc. Natl Acad. Sci. USA* **84**, 3881–3885 (1987).
5. Cooper, G. J. et al. Amylin found in amyloid deposits in human type 2 diabetes mellitus may be a hormone that regulates glycogen metabolism in skeletal muscle. *Proc. Natl Acad. Sci. USA* **85**, 7763–7766 (1988).
6. Höppener, J. W., Ahrén, B. & Lips, C. J. Islet amyloid and type 2 diabetes mellitus. *N. Engl. J. Med.* **343**, 411–419 (2000).
7. Westermark, P., Wernstedt, C., Wilander, E. & Sletten, K. A novel peptide in the calcitonin gene related peptide family as an amyloid fibril protein in the endocrine pancreas. *Biochem. Biophys. Res. Commun.* **140**, 827–831 (1986).
8. Westermark, G. T., Gebre-Medhin, S., Steiner, D. F. & Westermark, P. Islet amyloid development in a mouse strain lacking endogenous islet amyloid polypeptide (IAPP) but expressing human IAPP. *Mol. Med.* **6**, 998–1007 (2000).

9. Fernández-Alvarez, J. et al. Stable and functional regeneration of pancreatic β -cell population in nSTZ-rats treated with tungstate. *Diabetologia* **47**, 470–477 (2004).
10. Pilkington, E. H. et al. Pancreatic β -cell membrane fluidity and toxicity induced by human islet amyloid polypeptide species. *Sci. Rep.* **6**, 21274 (2016).
11. Mukherjee, A. et al. Induction of IAPP amyloid deposition and associated diabetic abnormalities by a prion-like mechanism. *J. Exp. Med.* **214**, 2591–2610 (2017).
12. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nat. Struct. Mol. Biol.* **27**, 653–659 (2020).
13. Luca, S., Yau, W.-M., Leapman, R. & Tycko, R. Peptide conformation and supramolecular organization in amylin fibrils: constraints from solid-state NMR. *Biochemistry* **46**, 13505–13522 (2007).
14. Bedrood, S. et al. Fibril structure of human islet amyloid polypeptide. *J. Biol. Chem.* **287**, 5235–5241 (2012).
15. Röder, C. et al. Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid- β fibrils. *Nat. Struct. Mol. Biol.* **27**, 660–667 (2020).
16. Gallardo, R. et al. Fibril structures of diabetes-related amylin variants reveal a basis for surface-templated assembly. *Nat. Struct. Mol. Biol.* **27**, 1048–1056 (2020).
17. Zhang, W. et al. Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *eLife* **8**, e43584 (2019).
18. Schweighauser, M. et al. Structures of α -synuclein filaments from multiple system atrophy. *Nature* **585**, 464–469 (2020).

19. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr. D Struct. Biol.* **76**, 94–101 (2020).
20. Chen, M.-S. et al. Characterizing the assembly behaviors of human amylin: a perspective derived from C-terminal variants. *Chem. Commun.* **49**, 1799–1801 (2013).
21. Sakagashira, S. et al. Missense mutation of amylin gene (S20G) in Japanese NIDDM patients. *Diabetes* **45**, 1279–1281 (1996).
22. Meier, D. T. et al. The S20G substitution in hIAPP is more amyloidogenic and cytotoxic than wild-type hIAPP in mouse islets. *Diabetologia* **59**, 2166–2171 (2016).
23. Sakagashira, S. et al. S20G mutant amylin exhibits increased in vitro amyloidogenicity and increased intracellular cytotoxicity compared to wild-type amylin. *Am. J. Pathol.* **157**, 2101–2109 (2000).
24. Cao, P. et al. Sensitivity of amyloid formation by human islet amyloid polypeptide to mutations at residue 20. *J. Mol. Biol.* **421**, 282–295 (2012).
25. Janson, J. et al. Increased risk of type 2 diabetes in Alzheimer disease. *Diabetes* **53**, 474–481 (2004).
26. Miklossy, J. et al. Beta amyloid and hyperphosphorylated tau deposits in the pancreas in type 2 diabetes. *Neurobiol. Aging* **31**, 1503–1515 (2010).
27. Peila, R., Rodriguez, B. L., Launer, L. J. & Honolulu-Asia Aging Study. Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: The Honolulu-Asia Aging Study. *Diabetes* **51**, 1256–1262 (2002).

28. Oskarsson, M. E. et al. In vivo seeding and cross-seeding of localized amyloidosis: a molecular link between type 2 diabetes and Alzheimer disease. *Am. J. Pathol.* **185**, 834–846 (2015).
29. Moreno-Gonzalez, I. et al. Molecular interaction between type 2 diabetes and Alzheimer's disease through cross-seeding of protein misfolding. *Mol. Psychiatry* **22**, 1327–1334 (2017).
30. O'Nuallain, B., Williams, A. D., Westermarck, P. & Wetzel, R. Seeding specificity in amyloid growth induced by heterologous fibrils. *J. Biol. Chem.* **279**, 17490–17499 (2004).
31. Krotee, P. et al. Common fibrillar spines of amyloid- β and human islet amyloid polypeptide revealed by microelectron diffraction and structure-based inhibitors. *J. Biol. Chem.* **293**, 2888–2902 (2018).
32. Andretto, E. et al. A hot-segment-based approach for the design of cross-amyloid interaction surface mimics as inhibitors of amyloid self-assembly. *Angew. Chem. Int. Ed.* **54**, 13095–13100 (2015).
33. Lövestam, S. et al. Seeded assembly in vitro does not replicate the structures of α -synuclein filaments from multiple system atrophy. *FEBS Open Bio* **11**, 999–1013 (2021).
34. Fujiwara, H. et al. α -Synuclein is phosphorylated in synucleinopathy lesions. *Nat. Cell Biol.* **4**, 160–164 (2002).
35. Sorrentino, Z. A. & Giasson, B. I. The emerging role of α -synuclein truncation in aggregation and disease. *J. Biol. Chem.* **295**, 10224–10244 (2020).

36. Asthana, S., Mallick, B., Alexandrescu, A. T. & Jha, S. IAPP in type II diabetes: basic research on structure, molecular interactions and disease mechanisms suggests potential intervention strategies. *Biochim. Biophys. Acta Biomembr.* **1860**, 1765–1782 (2018).
37. Jurgens, C. A. et al. β -cell loss and β -cell apoptosis in human type 2 diabetes are related to islet amyloid deposition. *Am. J. Pathol.* **178**, 2632–2640 (2011).
38. Krotee, P. et al. Atomic structures of fibrillar segments of hIAPP suggest tightly mated β -sheets are important for cytotoxicity. *eLife* **6**, e19273 (2017).
39. Falcon, B. et al. Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol.* **136**, 699–708 (2018).
40. Falcon, B. et al. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
41. Zhang, W. et al. Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
42. Boyer, D. R. et al. Structures of fibrils formed by α -synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nat. Struct. Mol. Biol.* **26**, 1044–1052 (2019).

Chapter 6

Low complexity domains of the nucleocapsid protein of SARS-CoV-2 form amyloid fibrils

Abstract

The self-assembly of the Nucleocapsid protein (NCAP) of SARS-CoV-2 is crucial for its function. Computational analysis of the amino acid sequence of NCAP reveals low-complexity domains (LCDs) akin to LCDs in other proteins known to self-assemble as phase separation droplets and amyloid fibrils. Previous reports have described NCAP's propensity to phase-separate. Here we show that the central LCD of NCAP is capable of both, phase separation and amyloid formation. Within this central LCD we identified three adhesive segments and determined the atomic structure of the fibrils formed by each. Those structures guided the design of G12, a peptide that interferes with the self-assembly of NCAP and demonstrates antiviral activity in SARS-CoV-2 infected cells. Our work, therefore, demonstrates the amyloid form of the central LCD of NCAP and suggests that amyloidogenic segments of NCAP could be targeted for drug development.

Introduction

The Nucleocapsid protein (NCAP) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an RNA-binding protein that functions in viral replication by packaging the genomic viral RNA (vRNA) and aiding virion assembly^{1,2,3,4,5,6,7,8,9}. During its function, NCAP engages in multivalent RNA–protein and protein–protein interactions and self-associates via several interfaces¹⁰. Increasing replication efficiency, NCAP forms concentrated protein–RNA compartments through a process of phase separation (PS)^{1,2,4,5,6,7,8,10,11,12}. NCAP PS is enhanced in low salt buffers^{4,5} and in the presence of zinc ions², and these PS droplets may exist in a liquid or solid-like state^{1,2,4,8,11}. The liquid state of the droplets is favored by NCAP phosphorylation and is presumed to enable vRNA processing in the early stages of infection^{4,8}.

In contrast, non-phosphorylated NCAP oligomerizes and forms solid-like droplets, possibly to facilitate RNA packaging and nucleocapsid assembly in later stages^{4,8}.

The sequence of NCAP encompasses both RNA-binding and low-complexity domains. Low-complexity domains (LCDs) are protein segments comprised of a restricted subset of amino acid residues such as glycine, arginine, lysine, and serine^{13,14,15}. Long mysterious in function, LCDs have recently been established to drive PS and form unbranched, twisted protein fibrils known as amyloid-like fibrils. Such behavior was observed in LCD-containing human RNA-binding proteins such as FUS, TDP-43, and hnRNPA2. By PS and amyloid formation, LCDs non-covalently link their parent proteins, and in some cases RNAs, into larger assemblies^{13,16,17,18}. These larger assemblies are associated with the formation of subcellular bodies known variously as hydrogels^{13,19}, condensates²⁰, and membrane-less organelles¹⁹. In short, the self-association of several RNA-binding proteins has been shown to be driven at least in part by amyloid-like fibrils formed by their LCDs and to be a regulatory element of RNA metabolism in cells¹⁹.

Motivating our study is a medical experience that even efficient vaccines rarely eradicate viral diseases and their legacies of morbidity and mortality²¹, so COVID-19 therapies are needed. Along with others¹⁰ we hold that NCAP of SARS-CoV-2 is a worthy drug target and that a better understanding of the structure and mechanism of action of NCAP may aid in drug development. NCAP is abundant in SARS-CoV-2-infected cells and its function is crucial for viral replication and assembly¹⁰. NCAP is also evolutionarily conserved in the coronavirus genus¹⁰, which may render it as an effective target not only for COVID-19 treatments but possibly also for future coronavirus pandemics.

Here we show that NCAP possesses two fibril-forming LCDs, one central and one C-terminal.

The central LCD forms Thioflavin-S (ThS)-positive PS droplets and amyloid fibrils that exhibit a

characteristic diffraction pattern. At least three adhesive segments in this central LCD are capable of mediating amyloid typical interactions, and we elucidated the atomic structure of the fibrils formed by each. Guided by these structures, we designed a peptide that shifts NCAP to a less ordered mode of aggregation and investigated the peptide's effect on the infection of human cells by SARS-CoV-2.

Results

NCAP contains central and C-terminal LCDs

Using the SEG algorithm²² we analyzed the sequence of NCAP and identified a 75-residue LCD (residues 175–249) within NCAP's central intrinsically disordered region, as well as a second, lysine-rich LCD of 19 residues (residues 361–379) within its C-terminal tail (CTT) (Fig. [1a, b](#)). SEG is a widely used algorithm that identifies segments in a sliding window as either high or low complexity by statistically analyzing the amino acid distribution as a measure of sequence complexity²². While not all LCDs identified this way are capable of PS and amyloid formation, LCDs that do phase separate are readily identified by SEG^{23,24}. In NCAP, those central and C-terminal LCDs, along with an N-terminal disordered region, flank the structured RNA-binding and dimerization domains of the protein (Fig. [1a](#)).

NCAP's LCDs participate in fibril formation

To assess possible amyloid formation of NCAP's LCDs and to identify adhesive segments that drive it, we expressed and purified NCAP and its LCD-containing segments in *E. coli*. Those segments included residues comprising NCAP's central LCD and surrounding residues (construct named LCD, residues 171–263) and a segment that includes the C-terminal LCD with the C-terminal tail and dimerization domain (construct DD-C_{term}, residues 257–419) (Fig. [1c](#)).

Only RNA-free protein fractions were combined at the last step of protein purification for use in subsequent experiments. We then verified that our purified full-length NCAP protein is capable of PS by mixing it with a 211-nucleotide 5'-genomic vRNA segment named hairpin-Site2 (S2hp; Supplementary Fig. [1](#), Supplementary Table [1](#)) in the presence and absence of the PS enhancing $ZnCl_2$ (Supplementary Fig. [2a](#) and Supplementary text). The S2 vRNA sequence was previously suggested to be a strong NCAP cross-linking site⁷, and we extended it by including the adjacent hairpin regions that improve binding to NCAP²⁵.

Using our recombinant protein system we found that NCAP's LCDs are capable of binding the amyloid-dye Thioflavin-T (ThT). In a ThT amyloid-formation kinetic assay performed over ~35 h of measurement (Fig. [1d, e](#)), S2hp vRNA mixtures (in 4:1 protein: vRNA molar ratio) of the central LCD and the DD-C_{term} segments of NCAP produced amyloid formation curves. Whereas the DD-C_{term} + vRNA curve plateau after ~3 h of incubation, LCD + vRNA plateaus ~10 h after the start of measurements while producing a significantly higher fluorescence signal than that of DD-C_{term}. The full-length NCAP also exhibited increased ThT fluorescence over 5 h of measurement when mixed with S2hp vRNA, followed by a slight decrease in signal, possibly because of spontaneous disaggregation (Supplementary Fig. [2e](#)). However, neither NCAP nor the DD-C_{term} segment demonstrated a clear lag phase in their ThT curves. Also, in the absence of S2hp vRNA, we did not detect an increase in ThT fluorescence in any of the samples within 35 h of measurements. This suggests that vRNA promotes the formation of ThT-positive aggregates from those LCD-containing protein constructs, at least in the first 1.5 days of incubation.

Visualization of fibrils by electron microscopy (EM) confirmed the propensity of the LCD-containing constructs to adopt fibrillar morphologies (Fig. [1f](#) and Supplementary Fig. [2c](#)). To observe fibrils of NCAP and its LCD-containing segments by EM we increased protein

concentration and incubated each protein separately for ~1–2 weeks with and without S2hp vRNA. Of note, under the conditions used for the kinetic ThT experiment (Fig. [1d,e](#) and Supplementary Fig. [2e](#)) we did not detect fibrils by EM, suggesting that the ThT experiment is more sensitive for the detection of amyloid-like aggregates or that ThT interacts with pre-fibrillar assemblies of the proteins. Other explanations, such as poor adherence of the protein fibrils to the EM grid and fibril reversibility are also reasonable. Nevertheless, with increased protein concentration and incubation time, fibrils were detected by EM both in the presence and absence of the vRNA. Fibrils of the DD-C_{term} segment with vRNA are morphologically different than those grown in its absence, however, the central LCD segment produces amyloid-looking fibrils under both conditions. Indeed, concentrated LCD-only samples exhibit increased ThT fluorescence signal upon 6 and 11 days of incubation, but with large sample-to-sample variability (Fig. [1g](#)). vRNA is, therefore, not essential for fibril formation and ThT binding, but may promote these processes.

The full-length NCAP also forms fibrillar morphologies in samples containing higher protein-to-vRNA ratio (40:1 protein:vRNA molar ratio), as well as when incubated with zinc ions in PBS (Supplementary Fig. [2b](#)), and particularly in a low ionic strength buffer (Supplementary Fig. [2d](#)) upon 3–6 days of incubation (as indicated in Supplementary Fig. [2](#)). NCAP and also the DD-C_{term} fibrils are much sparser in EM images compared to the central LCD, and their morphologies differ from those of the central LCD or canonical amyloid fibrils. Together, those observations suggest that NCAP and its LCD-containing segments, particularly the central LCD segment, are capable of forming aggregates of fibrillar morphologies as well as ThT-positive species.

The central LCD forms amyloid typical fibrils

To examine the amyloid property of fibrils formed by NCAP and its LCD segment we used X-ray fiber diffraction. The X-ray fiber diffraction patterns of the central LCD showed a sharp reflection at 4.7 Å spacing and a diffuse reflection at 10 Å typical of amyloid fibrils. This is true for fibrils formed by the LCD alone (no RNA), and by LCD with S2hp vRNA or a non-specific RNA segment (Fig. [2a](#)). This capacity of the central LCD segment of NCAP to stack into amyloid fibrils associates it with LCDs of other RNA-binding proteins that are involved in functional amyloid-formation and amyloid pathologies [13](#)[16](#)[18](#). We were unable, however, to obtain a clear diffraction pattern from the full-length NCAP. This may be a result of low fibril concentration, as evident by EM (Supplementary Fig. [2c](#)), and/or from fibril decomposition during washing steps meant to eliminate salts from the sample.

The central LCD segment of NCAP also readily forms unbranched fibrils in the presence of short, unstructured vRNA types such as the Site1 (S1), Site1.5 (S1.5) and S2 segments, as well as with a non-specific RNA segment of a similar length (Fig. [2b](#); Supplementary Table [1](#)), and even with no RNA (Fig. [1f](#)). When the LCD segment is incubated for one day with either S1 or S2 vRNA segments, the LCD produces heavily-stained clusters with fibrils protruding from their edges, but these clusters disperse after 4 days of incubation. Such behavior is not observed with S1.5 vRNA or the non-specific RNA segment (Fig. [2b](#)). This may suggest that the LCD fibril growth process may be altered by the RNA sequence. Overall, the amyloid formation of the central LCD offers that this region could potentially promote ordered self-assembly of NCAP under the appropriate conditions.

NCAP's central LCD forms PS droplets and solid particles

Next, we examined the capacity of the central LCD to form PS droplets with different S2hp vRNA concentrations and followed the behavior and character of the droplets over time in the

presence of the amyloid dye Thioflavin-S (ThS) using light and fluorescence microscopy. In samples of 4:1 and 40:1 LCD: S2hp vRNA molar ratios we visualized PS droplets that gradually transition into rough, less circular, seemingly solid particles (Fig. [3a, b](#)). In the 40:1 LCD: S2hp sample, PS droplets form and begin to fuse within 30 min of incubation, and ThS partitions into the droplets and produces rather bright fluorescence (Fig. [3a, c](#)). Upon 2 h of incubation, larger asymmetric droplets appear, and after 6 h, filamentous structures decorate the droplets. Within 4 days of incubation, the droplets transform into what appear as solid-like filamentous particles. At a higher S2hp concentration (4:1 LCD: S2hp molar ratio), small PS droplets appear after ~30 min, but those droplets show almost no ThS fluorescence (Fig. [3a, c](#)). Additional PS droplets form after 2 h of incubation and a weak ThS signal is detected. However, after 6 h incubation, and even more predominantly after 4 days, most droplets convert into brightly fluorescent particles (Fig. [3a, c](#)).

An analysis of LCD assemblies (droplets and solid-like particles) from a series of light microscope images taken at different time points of incubation shows that the mean area of the 40:1 LCD:vRNA assemblies somewhat increases upon the transition from liquid droplets to the fibrous looking particles. The median value of the mean circularity of the assemblies (weighted by the size of the droplet/particle) drops by ~60% between the first (day 1) and last (day 4) measurements (Fig. [3b](#), left). A similar analysis of the 4:1 LCD: vRNA sample revealed a greater increase in the mean area of the assemblies upon 4 days of incubation, and a greater decrease of ~80% in the median value of the mean circularity (Fig. [3b](#), right), suggesting a massive transition of circular liquid droplet into large, amorphous, solid-like particles. Quantification of the mean ThS fluorescence from images taken at 0.5 h and 4 days of incubation of both samples show a ~4-fold increase in ThS fluorescence in the 40:1 LCD:S2hp sample, and ~58-fold increase in

fluorescence intensity in the 4:1 LCD:S2hp ratio (Fig. 3c). Here too, no fibrils could be detected by EM at the concentration and incubation times used for the PS assay.

In a separate experiment, we also followed the aggregation of the central LCD segment when incubated alone or with S2hp vRNA (in 4:1 respective ratio) by measuring turbidity (Fig. 3d). We detected elevated turbidity of the LCD + vRNA sample at the beginning of the measurement, as opposed to the LCD only sample that was not turbid. This offers that the central LCD immediately aggregates upon mixing with vRNA. The LCD + vRNA sample shows biphasic behavior, with a decrease in turbidity between 0 and 5.5 h of incubation, followed by a renewed increase. This biphasic behavior of the 4:1 LCD:S2hp vRNA sample may be related to the transition from liquid droplets to solid particles visualized in this sample between 2 and 6 h of incubation (Fig. 3a). Overall, our results indicate that the central LCD of NCAP forms ThS-positive PS droplets that transition from circular liquid droplets to fibrous or amorphous solid-like particles, and that the RNA concentration governs the kinetics of this process and the morphology of the assemblies.

Structures of LCD-derived steric-zipper-forming segments

To interfere with the self-assembly of the LCD segment, and thereby possibly of NCAP, we seek structural information of specific amyloid-like LCD sequences. Amyloid fibrils are stabilized by pairs of tightly mating β -sheets, with zipper-like interfaces termed steric zippers that can be predicted by a computer algorithm²⁶ [<https://services.mbi.ucla.edu/zipperdb/>]. Within the central LCD, we identified (Supplementary Fig. 3a) and crystallized three such steric zipper-forming segments: ¹⁷⁹GSQASS₁₈₄, ²¹⁷AALALL₂₂₂, and ²⁴³GQTVTK₂₄₈. X-ray structures confirmed that each segment forms amyloid-like fibrils composed of pairs of β -sheets stabilized by steric zipper interfaces (Fig. 4, and Supplementary Figs. 4 and 5; Table 1). GSQASS and GQTVTK segments

both form parallel, in-register β -sheets, whereas the AALALL segment is crystalized in two forms, both with antiparallel β -sheets²⁷. The weaker zipper interface of the second form incorporates polyethylene glycol (Supplementary Fig. 4), and we do not consider it further. Solvation-free energy calculations based on our crystal structures (Supplementary Table 2) suggest that the AALALL steric-zipper is the most stable of the three, consistent with its predominance of hydrophobic residues. GSQASS and GQTVTK, on the contrary, contain mostly polar residues (Fig. 4c). The AALALL segment also overlaps with a region predicted to participate in context-dependent interactions of NCAP (Supplementary Fig. 3b, residues 216–221), namely interactions that change between disordered and ordered modes as a function of cellular environment and protein interactors and are likely to be responsible for the formation of amyloid fibrils within liquid droplets²⁸. For drug design, we pursued AALALL and GQTVTK as targets but excluded GSQASS because it resembles LCDs found in the human proteome²⁹.

A structure-based disruptor of NCAP's PS exhibits antiviral activity

To modulate NCAP's self-assembly we exploited the propensity of NCAP's LCD to form steric-zipper structures. Guided by our amyloid-spine structures we screened an array of peptides, each designed to interact with a specific steric-zipper forming segment. We have found such peptides to inhibit the aggregation and prion-like seeding of other amyloid-forming proteins (e.g. refs. ³⁰³¹³²³³). To design the steric-zipper targeting disruptors of NCAP self-assembly we implemented two approaches: sequence/structure-based design and Rosetta-based modeling³⁴. Both approaches produce sequences that bind strongly to our steric zipper structure targets and contain bulky residues that block the interactions of additional NCAP molecules via this interface (Fig. 5a).

Screening of a panel of our designed peptides in vitro revealed that a peptide we named G12 disrupts NCAP's PS. G12 is a D-amino acid peptide with the sequence d-(rrffmvlm), designed against the AALALL steric zipper-forming segment (Fig. [5a](#); Supplementary Table [3](#)).

Increasing concentrations of G12 disrupt the formation of circular NCAP PS droplets and instead promote the formation of large network-like aggregates as judged by light microscopy (Fig. [5b](#), [c](#) and Supplementary Fig. [6](#)).

We then proceeded to test G12's antiviral activity in HEK293 cells that express the human ACE2 receptor (HEK293-ACE2 cells). First, we verified that HEK293-ACE2 cells transfected with FITC-labeled G12 show that G12 remains soluble and diffuse in the cytoplasm for at least 24 h (Supplementary Fig. [7](#)). Next, we used quantitative immunofluorescence labeling to detect the percentage of SARS-CoV-2 infection in cells transfected with increasing concentrations of G12 or a vehicle only control. The percentage of infected cells in each G12-treated culture was normalized to the infected vehicle-only control (Fig. [5d](#) and Supplementary Fig. [8](#)). Cytotoxicity was tested with the same cells and G12 concentrations using the LDH toxicity assay (Fig. [5d](#), red curve). Whereas G12 concentrations lower than 6 μM slightly increase the relative percent infectivity of treated cells, in the range of $\sim 6\text{--}16\ \mu\text{M}$, G12 exhibits dose-dependent antiviral activity while reducing the amount of virus detected in the culture by up to $\sim 50\%$ without inflicting cytotoxicity (Fig. [5d](#) and Supplementary Fig. [8](#)). Since G12 is dissolved in DMSO we could not test higher G12 concentrations in this cell-based assay to obtain a complete dose-response curve, however fitting a non-linear regression model to our data allowed a rough IC_{50} estimation of $7\text{--}11\ \mu\text{M}$ (Fig. [5d](#)). We, therefore, suggest that G12 serves as a proof-of-concept showing that by targeting amyloidogenic segments within the central LCD of NCAP we interfere with NCAP's self-assembly and thereby the viral life cycle.

Discussion

The NCAP protein of SARS-CoV-2 belongs to the subclass of fibril-forming proteins that contains both an RNA-binding domain and LCDs (Fig. [1a, b](#)). NCAP undergoes PS [1·2·3·4·5·6·7·8](#) (Supplementary Fig. [2a](#)), and as we show here, its central LCD forms amyloid-like fibrils (Figs. [1](#) and [2](#)).

The central LCD of NCAP forms fibrils with the long and structured S2hp vRNA segment (Fig. [1](#) and Supplementary Fig. [1](#); Supplementary Table [1](#)), with various short, single-stranded RNA sequences (Fig. [2](#); Supplementary Table [1](#)), and also with no RNA (Fig. [1f](#)). This suggests that specific LCD–RNA interactions are not required for LCD-amyloid formation. Nevertheless, the LCD does bind to at least S2hp vRNA [25](#), and LCD fibril maturation is influenced by the RNA sequence and length (Fig. [2b](#)), so LCD–RNA interactions play a role. The LCD segment is highly positively charged (Fig. [1b](#)), especially in its non-phosphorylated form. Therefore we expect it to engage in non-specific polar interactions with the negatively charged RNA, which in turn may promote the accumulation of LCD molecules, including through PS formation (Fig. [3](#)), and their amyloid-like assembly (Fig. [2a](#)).

The amyloid-like characteristics of the central LCD of NCAP are similar to those of the LCDs of FUS [35·36](#), hnRNPA2 [37](#), TDP-43 [38](#), and other RNA-binding proteins that are involved in RNA metabolism in eukaryotic cells [13·17](#), and under certain circumstances, also in amyloid-associated pathologies [19](#). This equivalent ability of the LCD of NCAP to PS and stack into amyloid-like structures in the presence of RNA proposes its potential function in the yet elusive mechanism of NCAP self-assembly.

Full-length NCAP is capable of only sparse fibril formation in the presence and absence of S2hp vRNA and with ZnCl₂ (Supplementary Fig. [2](#)). Whereas fibrils formed in the presence of S2hp do not exhibit amyloid-typical morphology (Supplementary Fig. [2b, c](#)), the NCAP + S2hp aggregates produce a ThT amyloid formation curve, but it lacks a lag phase (Supplementary Fig. [2e](#)). Short or absent lag phase in ThT curves may result from the existence of pre-formed amyloid seeds in the tested sample³⁹, or from a fast pickup of the ThT signal prior to starting the measurements. The latter may be reasonable given that NCAP rapidly aggregates and becomes turbid in the presence of S2hp (Supplementary Fig. [2f](#)). In a parallel study, we show that the structured regions of S2hp are essential for strong binding to NCAP, whereas S2 and other short, single-stranded RNA segments bind to it weakly²⁵. Here, we detected fibrils of NCAP with S2hp (Supplementary Fig. [2b, c](#)), but its LCD segment is also capable of forming fibrils in the presence of the short, unstructured vRNA segments S1, S1.5, and S2 (Fig. [2b](#); Supplementary Table [1](#)). We, therefore, speculate that robust amyloid formation of full-length NCAP requires strong interactions with specific vRNA sequences and/or co-factors that we are yet to identify.

The amyloid formation of the central LCD of NCAP is attributed to at least three adhesive peptide sequences: ¹⁷⁹GSQASS₁₈₄, ²¹⁷AALALL₂₂₂, and ²⁴³GQTVTK₂₄₈ (Fig. [4](#) and Supplementary Fig. [3a](#)). ¹⁷⁹GSQASS₁₈₄ and ²⁴³GQTVTK₂₄₈, are predominantly polar (Fig. [4c](#)), similar to the highly polar reversible amyloid fibrils formed by the LCDs of FUS and hnRNPA2⁴⁰. The segment ¹⁷⁹GSQASS₁₈₄ is part of a conserved serine/arginine (SR)-rich region (residues 176–206)⁴ and it includes the two phosphorylation sites S180 and S184¹¹. Phosphorylation of the SR-rich region facilitates the transformation of NCAP's PS droplets from a solid to a liquid-like state during viral genome processing. The non-phosphorylated protein, however, is associated with solid PS droplets and nucleocapsid

assembly⁸. Both S180 and S184 face the dry, tight interface formed between the β -sheets in the structure of $_{179}\text{GSQASS}_{184}$ (Fig. [4a](#)). Phosphorylation of those residues is indeed likely to reverse the solid, amyloid-like packing of this segment. Of note, all results in this paper showing the ordered, solid-like mode of aggregation were obtained with non-phosphorylated proteins and peptides.

The second adhesive segment, $_{217}\text{AALALL}_{222}$, is highly hydrophobic and produces the most stable steric-zipper structure (Supplementary Table [2](#)). $_{217}\text{AALALL}_{222}$ is also predicted to help switch between disordered and ordered modes of protein aggregation as a factor of cellular environment and protein interactors (Supplementary Fig. [3c](#), residues 216–221)²⁸. Those properties of $_{217}\text{AALALL}_{222}$ render it an important target for the disruption of NCAP's self-assembly. The $_{243}\text{GQTVTK}_{248}$ segment, however, resembles sequences in LCDs found in the human proteome²⁹, and is therefore a poor target for drug design.

The self-assembly of NCAP is crucial for RNA packaging and SARS-CoV-2 replication¹⁰. The amyloid formation of NCAP's LCD is a form of NCAP self-assembly, but it is yet unclear whether NCAP forms and functions as amyloid in the viral life cycle. Nevertheless, PS-mediated self-assembly of NCAP was shown to occur in NCAP-transfected and SARS-CoV-2 infected cells^{4,7,11,41}. By targeting the amyloidogenic segment $_{217}\text{AALALL}_{222}$ (Fig. [4](#) and Supplementary Figs. [4 and 5](#)) with G12, we inhibited the PS formation of NCAP in vitro (Fig. [5b, c](#)). G12 is a peptide designed to interact and block the $_{217}\text{AALALL}_{222}$ interface by exploiting the tendency of this segment to form steric-zipper structures (Fig. [5a](#)). G12 is, however, incapable of complete disruption of NCAP self-assembly, perhaps because assembly is guided by several proteins interfaces¹⁰. Evaluation of G12 in SARS-CoV-2-infected cells revealed dose-dependent antiviral activity in concentrations of 6–16 μM without inflicting

cytotoxicity (Fig. [5d](#)). G12 concentrations lower than 6 μM , however, led to increased viral infection in treated cells. We speculate that when administered in subeffective concentrations, G12 partitions into NCAP droplets and increases NCAP's effective concentration which possibly promotes self-assembly and formation of new virions. When administered in proper concentrations, we anticipate that the antiviral activity of G12 results from its interference with the self-assembly of NCAP, as designed, leading to poor RNA packaging and viral particle assembly.

The three steric-zipper-forming segments we identified in this work are conserved between the NCAPs of SARS-CoV-2 and SARS-CoV. The only exception is alanine in position 217 in the sequence of SARS-CoV-2 which is replaced by threonine in the NCAP of SARS-CoV (Supplementary Fig. [9a](#)). A ZipperDB²⁶ [<https://services.mbi.ucla.edu/zipperdb/>] calculation on the LCD of the NCAP of SARS-CoV revealed that this threonine shifts the steric-zipper forming segment to the hydrophobic ALALLL sequence (with Rosetta free energy score of -24.700) that is aligned and conserved with residues 218–223 in the NCAP of SARS-CoV-2. This suggests that the LCD in the NCAP of SARS-CoV may also form amyloids, and that future SARS coronaviruses might share this targetable property. A SEG analysis²² performed on the sequence of the NCAPs of a number of α , β and γ coronaviruses from various species showed that many of these viruses contain LCDs that could potentially participate in amyloid formation (Supplementary Fig. [9b](#)). This suggests that amyloid formation of NCAP LCDs is a general mechanism of action and a common targetable trait in coronaviruses.

Despite the high conservation of NCAP¹⁰, some mutations have been identified in strains that emerged since the initial SARS-CoV-2 outbreak in Wuhan, China. To date, no NCAP mutations were detected within our amyloid steric-zipper spine segments: ¹⁷⁹GSQASS₁₈₄, ²¹⁷AALALL₂₂₂,

and ²⁴³GQTVTK₂₄₈. Nevertheless, some mutations were detected within the central LCD, including the prevalent R203K/M, G204R/M, and T205I substitutions^{42:43:44}. The R203K/G204R mutants exhibit higher PS propensity compared to the Wuhan variant⁴¹, and the R204M mutation promotes RNA packaging and viral replication in the delta variant⁴⁵. Also interesting are the G214C (Lambda variant) and G215C (Delta variant) substitutions^{42:43:44} that are adjacent to the ²¹⁷AALALL₂₂₂ steric-zipper segment. The Delta variant spread faster and caused more infection compared to its predecessors^{46:47:48:49}. The Delta variant also carries a D377Y mutation in the C-terminal LCD segment of NCAP. It is possible that mutations in NCAP's LCD enhance amyloid formation, similarly to mutations in other RNA-binding proteins^{35:50:51:52}. This is important to explore since amyloid fibrils are associated with numerous dementias and movement disorders^{53:54}. Amyloid cross-talk and hetero-amyloid aggregation, including between microbial and human amyloid proteins (e.g. refs. ^{55:56:57:58}), is a well-known phenomenon that is postulated to exacerbate amyloid pathology⁵⁹.

The possible connection of amyloid formation of NCAP to neurodegeneration was already recently suggested. NCAP was shown to interact and accelerate the amyloid formation of the Parkinson's disease-related protein, α -synuclein, which may explain the correlation between Parkinsonism and SARS-CoV-2 infection⁶⁰. NCAP was also shown to partition into PS droplets⁵ and accelerate amyloid formation⁶¹ of FUS, TDP-43, hnRNPA1, and hnRNPA2. In certain forms, those proteins are associated with neurodegenerative and movement disorders¹⁹. In SARS-CoV-2 infected cells, NCAP impairs the disassembly of stress granules into which it partitions, and in cells expressing an ALS-associated mutant of FUS, NCAP enhances FUS aggregation into amyloid-containing puncta⁶¹. Those observations, together with the capacity of

NCAP's central LCD to form amyloid, call for further investigation of the possible NCAP-amyloid formation and regulation in SARS-CoV-2-infected cells, and of the possible involvement of NCAP in amyloid cross-talk and human neurodegeneration.

Our study of the amyloid formation of NCAP expands an emerging class of known amyloid-forming viral proteins. In the Influenza A virus, the full-length and N-terminal segment of the PB1-F2 protein form cytotoxic amyloid fibrils when mixed with liposomes, and the C-terminal segment forms cytotoxic amyloid oligomers⁶². A 111-residue segment from the V protein of Hendra virus, a respiratory virus that may progress in humans to severe encephalitis, was shown to undergo a liquid-to-hydrogel transition of its PS droplets and to produce amyloid-like fibrils⁶³. The RIP-homotypic interaction motif containing segments of the herpes simplex virus 1 (HSV-1) protein ICP⁶⁴, the murine cytomegalovirus protein M45⁶⁵ and the varicella-zoster virus protein ORF20⁶⁶ are capable of forming heteromeric amyloid complexes with host proteins. Other examples of amyloid-forming peptide segments include avibirnavirus viral protease that contributes to protease self-assembly⁶⁷, peptides from the fiber protein of adenovirus^{68:69}, and a nine-residue peptide from the C-terminus of the SARS-CoV envelope protein⁷⁰. Recent studies also showed the amyloidogenic properties of various segments of the spike protein⁷¹, and other regions in the proteome⁷² of SARS-CoV-2. None of these previously studied viral amyloids, however, was associated with NCAPs. Nevertheless, LCDs and prion-like sequences, such as those that exist in NCAP⁵ were identified in over two million eukaryotic viruses⁷³. Therefore, our finding of the amyloid formation of this viral RNA-binding protein may foreshadow a much wider field for investigation.

In summary, this work extends knowledge of amyloidogenic viral proteins and their LCD segments, associates NCAP with known amyloid-forming RNA-binding proteins, and may

inspire future investigation of NCAP amyloid formation in SARS-CoV-2 infection. Finally, we also suggest an approach for the development of SARS-CoV-2 therapeutics via disruption of NCAP self-assembly by targeting and capping amyloid-driving steric-zipper segments of NCAP.

Methods

Molecular biology reagents

Phusion HF DNA polymerase, Quick Ligase, and restriction enzymes were purchased from New England BioLabs. Custom DNA oligonucleotides were synthesized by IDT (Coralville, IA). RNA oligonucleotides, S1, S1.5, S2, and the non-specific RNA (siDGCR8-1, antisense strand) were synthesized by Horizon Discovery Biosciences.

Computational predictions and sequence alignment

Prediction of low-complexity sequences in the NCAP of SARS-CoV-2

The amino acid sequence of the Nucleocapsid protein of SARS-CoV-2 (NCAP; UniProtKB⁷⁴ accession number: P0DTC9 [<https://covid-19.uniprot.org/uniprotkb/P0DTC9#Sequence>]) was evaluated using SEG²² with default settings: window length = 12, trigger complexity 2.2, extension complexity 2.5. LCDs were defined by strings of at least 10 low-complexity residues. Long LCDs, such as the central NCAP-LCD, were allowed no more than five interrupting non-low-complexity residues between strings of 10 or more low-complexity residues²².

Prediction of LCDs in the NCAPs of various coronaviruses

A list of coronavirus Nucleocapsid proteins was downloaded from the European Nucleotide Archive (ENA; [<https://www.ebi.ac.uk/genomes/virus.html>]), and protein sequences were

retrieved from Uniprot [<https://www.uniprot.org/>]. Low complexity residues were identified using the SEG algorithm²² with default parameters (see above). Redundant low-complexity region sequences from strains of individual viruses were removed. Low-complexity region sequences were aligned in BioEdit using the ClustalW algorithm with gap penalties set to 100 in order to avoid the insertion of gaps in the aligned sequences. Gaps consisting of hyphens in between amino acid stretches in an individual sequence represent an interrupting, non-low-complexity segment of at least 20% the length of the longest LCD in the protein rather than defined gaps in the alignment. Some of these gaps were manually made larger or smaller to achieve a more accurate alignment. Supplementary Fig. [9b](#) is the representation of this alignment in Jalview.

Prediction of steric-zipper forming segments

This was done on the Nucleocapsid proteins of SARS-CoV-2 (UniProtKB⁷⁴ accession number: [P0DTC9](#)) and SARS-CoV (UniProtKB⁷⁴ accession number: [P59595](#)) using the ZipperDB algorithm²⁶ [<https://services.mbi.ucla.edu/zipperdb/>].

Prediction of PS forming regions and context-dependent interactions

Was performed using the Fuzdrop algorithm²⁸ [<https://fuzdrop.bio.unipd.it/predictor>] on the Nucleocapsid protein of SARS-CoV-2 (UniProtKB⁷⁴ accession number: [P0DTC9](#) [<https://covid-19.uniprot.org/uniprotkb/P0DTC9#Sequence>]).

Sequence conservation

Sequence conservation analysis was performed on the LCDs of the NCAPs of SARS-CoV and SARS-CoV-2 (UniProtKB⁷⁴ accession numbers: [P59595](#) and [P0DTC9](#), respectively). The sequences were aligned and colored according to conservation in Jalview.

Construct design

Full-length SARS-CoV-2 Nucleocapsid protein gene and its fragments were PCR amplified from 2019-nCoV Control Plasmid (IDT Inc., cat. no. 10006625) and spliced with N-terminal 6xHis-SUMO tag⁷⁵ using splicing by overlap extension (SOE) technique⁷⁶. 5' KpnI and 3' SacI restriction sites introduced with the flanking primers were used to ligate the resulting fragments into pET28a vector. When needed, an additional round of SOE was performed to generate internal Nucleocapsid protein deletion mutants. Construct sequences were confirmed by Sanger sequencing (Laragen, Culver City, CA). Primers used for cloning are given in Supplementary Table 4, DNA sequences and alignment of translated amino acid sequences from Sanger sequencing are given in Supplementary Figs. 10 and 11, respectively.

Protein expression and purification

NCAP segments and full-length protein were expressed as fusions to 6xHis-SUMO (6xHis-SUMO-NCAP). Plasmids were transformed into *Escherichia coli* Rosetta2 (DE3) strain (MilliporeSigma cat. no 71-397-4) and small-scale cultures were grown at 37 °C overnight in LB with 35 µg/mL kanamycin and 25 µg/mL chloramphenicol. TB with 35 µg/µL kanamycin was inoculated with overnight starter culture at a 1:100 ratio and large-scale cultures were grown at 37 °C with 225 rpm shaking until the OD₆₀₀ reached ~0.6. Protein expression was induced with 1 mM IPTG and cultures were further incubated with shaking at 28 °C overnight, then harvested at 5000×g at 4 °C for 15 min. Bacterial pellets were either used right away or stored at -20 °C. Pellets from 2–4 L of culture were re-suspended in ~200 mL chilled Buffer A (20 mM Tris pH 8.0, 1 M NaCl) supplemented with Halt Protease Inhibitor Cocktail (ThermoScientific cat. no. 87785) and sonicated on the ice at 80% amplitude for a total sonication time of 15 min, with pauses at regular intervals so the sample does not exceed 15 °C. Cell debris was removed via

centrifugation at 24,000×g at 4 °C for 30- 60 min, filtered twice through 0.45 µm high particulate syringe filters (MilliporeSigma cat. no. SLCRM25NS), and imidazole added to 5 mM. Filtered clarified lysate was loaded onto HisTrap HP columns (GE Healthcare) and proteins were eluted over a step-gradient with Buffer B (20 mM Tris pH 8.0, 1 M NaCl, 500 mM imidazole), with extensive low-imidazole (<20%) washes to improve purity. NCAP proteins were generally eluted in 20–50% Buffer B. Fractions were analyzed by SDS–PAGE, pooled and dialyzed against 20 mM Tris pH 8.0, 250 mM NaCl at 4 °C overnight. Following dialysis, the sample was concentrated using Amicon Ultra-Centrifugal filters (MilliporeSigma) and urea was added up to 1 M final concentration if protein precipitation was observed. Ulp1 protease (homemade) was added at a 1:100–1:200 w/w ratio to purified proteins, along with 1 mM DTT, and the sample was incubated at 30 °C with 195 rpm shaking for 1–2 h. After cleavage, NaCl was added to 1 M final concentration to reduce aggregation, and the sample was incubated with HisPur Ni-NTA resin (Thermo Scientific cat. no. PI88222) equilibrated in Buffer A at 25 °C with 140 rpm shaking for 30 min. Cleaved NCAP proteins were eluted from the resin via gravity flow chromatography, then the resin was washed twice with Buffer A, twice with Buffer A + 5 mM imidazole, and finally with Buffer B. The flow-through and appropriate washes were concentrated and flash-frozen for storage or further purified by gel filtration. Directly prior to gel filtration, the sample was centrifuged at 21,000×g for 30 min at 4 °C to remove large aggregates. Soluble protein was injected on a HiLoad Sephadex 16/600 S200 (for proteins larger than ~25 kDa) or S75 (for proteins smaller than ~25 kDa) (GE Healthcare) equilibrated in SEC buffer (20 mM Tris pH 8.0, 300 mM NaCl) and run at a flow rate of 1 mL/min. Elution fractions were assessed by SDS–PAGE for purity, and confirmed to have low RNA contamination as assessed by 260/280 nm absorbance ratio. Pooled fractions were concentrated and 0.2-µm filtered. Protein

concentration was measured by A280 absorbance using a NanoDrop One (ThermoScientific) and calculated by the sequence-specific extinction coefficient, and aliquots were flash-frozen and stored at -80°C . Of note, the first N-terminal residue in all purified proteins (residue #1) is a threonine remaining from cleavage of the 6xHis-SUMO tag during protein purification.

Rosetta-based peptide inhibitor design

Crystal structures of LCD segments GSQASS and AALALL (form 1) were used as templates for the design of peptide inhibitors in Rosetta3 software³⁴. 5 layers of the steric zipper structure were generated. A 6-residue peptide chain was placed at the top or bottom of the fibril-like structure. Rosetta Design was used to sample all amino acids and their rotamers on the sidechains of the fixed peptide backbone. The lowest energy conformations of the sidechains were determined by minimizing an energy function containing terms for Lennard–Jones potential, orientation-dependent hydrogen bond potential, solvation energy, amino acid-dependent reference energies, and statistical torsional potential dependent on the backbone and sidechain dihedral angles. Buried surface area and shape complementarity were scored by AREAIMOL⁷⁷ and Sc⁷⁸, respectively, from the CCP4 suite of crystallographic programs⁷⁹.

Solvation-free energy estimates were calculated using software available here:

[<https://doi.org/10.5281/zenodo/6321286>]. Design candidates were selected based on their calculated binding energy to the top or bottom of the fibril-like structure, shape complementarity, and propensity for self-aggregation. The binding energy for an additional strand of the native sequence (i.e., AALALL) was computed for comparison with peptide inhibitor designs. The structural model of each candidate peptide was manually inspected in PyMOL⁸⁰. Many computational designs produced sequences with high hydrophobic content, thus two arginine residues were added onto the N-terminal end to increase peptide solubility. Candidate G12 was

the most effective inhibitor in preliminary screens and therefore was chosen for further evaluation.

Peptide synthesis and purification

The NCAP steric zipper segments $_{179}\text{GSQASS}_{184}$ and $_{243}\text{GQTVTK}_{248}$ were synthesized by LifeTein. The inhibitor candidate G12 was synthesized by LifeTein and GenScript. All peptides were synthesized at over 98% purity. The NCAP segment $_{217}\text{AALALL}_{222}$ was synthesized and purified in-house as H-AALALL-OH. Peptide synthesis was carried out at a 0.1 mmol scale. A 2-chlorotrityl chloride resin (Advanced Chemtech) was selected as the solid support with a nominal loading of 1.0 mmol/g. Each loading of the first amino acid was executed by adding 0.1 mmol of Fmoc-Leu-OH (Advanced Chemtech FL2350/32771) and 0.4 mmol of diisopropylethylamine (DIPEA), dissolved in 10 mL of dichloromethane (DCM), to 0.5 g of resin. This mixture was gently agitated by bubbling with air. After 30 min, the supernatant was drained, and the resin was rinsed twice with 15 mL aliquots of the capping solution, consisting of 17:2:1 DCM/MeOH/DIPEA. With the first amino acid loaded, the elongation of each polypeptide was completed in a CEM Liberty Blue™ Microwave Peptide Synthesizer. A 1.0 M solution of N,N'-diisopropylcarbodiimide (DIC) in DMF was used as the primary activator, and a 1.0 M solution of ethyl cyanohydroxyiminoacetate (oxyma) in DMF, buffered by 0.1 M of DIPEA was used as a coupling additive. The Fmoc-L-Ala-OH used was also purchased from Advanced Chemtech (FA2100/32786). The microwave synthesizer utilizes 0.2 M solutions of each amino acid. For the deprotection of N-termini, Fmoc protecting groups, a 9% w/v solution of piperazine in 9:1 N-Methyl-2-Pyrrolidone to EtOH buffered with 0.1 M of oxyma was used. For 0.1 mmol deprotection reactions, 4 mL of the above deprotection solution was added to the resin. The mixture was then heated to 90 °C for 2 min while bubbled with nitrogen gas. The

solution was drained, and the resin was washed 4 times with 4 mL aliquots of DMF. For 0.1 mmol couplings, 2.5 mL of 0.2 M amino acid solution (0.5 mmol) was added to the resin along with 1 mL of the DIC solution (1.0 mmol) and 0.5 mL of oxyma solution (0.5 mmol). This mixture was agitated by bubbling for 2 min at 25 °C, then heated to 50 °C followed by 8 min of bubbling. After the last deprotection, the resin was washed with methanol, diethyl ether, dried over the vacuum, and introduced to a cleavage cocktail consisting of 20 mL of trifluoroacetic acid (TFA), 0.50 mL of water, 0.50 mL of triisopropylsilane (TIS). After 2 h of vigorous stirring, the mixture was filtered, and the filtrate was concentrated in vacuo. The residue was triturated with cold diethyl ether, and precipitated, the crude peptide was collected by filtration. The crude peptide was then purified by RP-HPLC, using an Interchim puriFlash® 4125 Preparative Liquid Chromatography System equipped with a Luna (Phenomenex, C18(2), 5 µm, 100 Å, 30 × 100 mm) column. For purification, two buffer systems were utilized. Initial purifications and salt exchanges were executed with a 13 mM aqueous solution of trifluoroacetic acid (TFA; [A]) and a 2:3 water to acetonitrile solution, buffered by 13 mM of TFA ([B]). For the better resolution of diastereomers and other impurities, ultrapure water, buffered by 14 mM of HClO₄, and a 2:3 water to acetonitrile solution, buffered by 5.6 mM of HClO₄, were selected as mobile phases A and B, respectively. The purity of the purified fractions was analyzed by RP-HPLC, using an Agilent 1100 Liquid Chromatography System equipped with a Kinetex (Phenomenex, C18, 5 µm, 100 Å, 4.6 × 250 mm) column. Ultrapure water with 0.1% TFA, and a 1:9 water to acetonitrile solution with 0.095% TFA were selected as mobile phases [A] and [B], respectively. The flow rate was set at 1.0 mL/min and the gradient used is detailed in Supplementary Table 5. The UV absorption at 214 nm was monitored. The resulting chromatogram is shown in Supplementary Fig. 12.

RNA in vitro transcription and purification

The nucleic acid sequence corresponding to S2hp (Supplementary Table [1](#)) was cloned from a gBlock (IDT) of the first 1000 nucleotides of the 5'-end of the SARS-CoV-2 genome into pUC19 vectors using the restriction sites EcoRI and KpnI. Forward primer P2627 (5'-TAATACGACTCACTATAGGCTGTGTGGCTGTCCTCG-3') containing the T7 promoter sequence was added at a low concentration of 0.5 nM in addition to forward primer P1471 (5'-GCGAATTCTAATACGACTCACTATAGG-3') containing the EcoRI restriction sequence and T7 promoter sequence at the normal concentration of 500 nM. Reverse primer P2644 (5'-CGGGGTACCTCGTTGAAACCAGGGACAAG-3') containing the KpnI restriction sequence was added at 500 nM. The clone was sequence-confirmed and the miniprep was used as a template for PCR. The forward primer for PCR containing the T7 promoter sequence was biotinylated on the 5' end for removal of PCR template after transcription. The PCR product was purified by HiTrap column. The running buffer solutions (0.2- μ m filtered) contained 2 M NaCl, 10 mM HEPES pH 7.0 (buffer A), and 10 mM NaCl, 10 mM HEPES pH 7.0 (buffer B). The purified PCR products were concentrated using Amicon Ultra centrifugal filter units (Millipore) and buffer-exchanged against 10 mM Na/HEPES pH 7.0. Transcription reactions ranging from 5 to 100 μ L were set up. The transcription reaction was incubated at 37 °C with gentle shaking for one hour. After transcription, streptavidin beads (ThermoFisher) were added to the transcription and set on a rotator at room temperature for an additional 15 min. The transcription reaction was centrifuged at 500 \times *g* for 10 min at 4 °C. The supernatant was decanted and the pellet containing any PCR template remaining was discarded. The transcription reaction was then purified by 5 mL HiTrap Q HP column in several rounds, loading ~5 μ L into the column each round. The purified RNA was concentrated using Amicon Ultra-15 (Millipore) and the buffer was

exchanged into 10 mM HEPES pH 7.0. The purity of the RNA was confirmed using denaturing polyacrylamide gels. The concentration was calculated by measuring OD₂₆₀ and a conversion factor of 40 µg/mL/OD₂₆₀.

PS assays

All solutions were prepared using DNase/Rnase-free water (ultrapure water) and were filtered twice using a 0.22-µm syringe filter. Preparations were done under sterile conditions and using sterile filter pipette tips to prevent RNA degradation.

PS of NCAP with ThS staining (Supplementary Fig. [2a](#))

Experiments were carried out in 96-well black/clear glass-bottom plates (Cellvis glass-bottom plates cat. no. P96-1.5H-N). S2hp RNA, stored at -20 °C was thawed, then annealed by heating at 95 °C for 3 min and transferring quickly on the ice. The RNA was diluted by its original buffer of 10 mM HEPES pH 7.0 to 750 µM and 75 µM working solutions. 1 mM ZnCl₂ was prepared in ultrapure water and filtered twice with a 0.22-µm syringe filter. Fresh Thioflavin S (ThS) solution was prepared from powder (MP Biomedicals) in ultrapure water at 0.002% w/v and filtered. Purified NCAP stock solution was centrifuged at 15,000×g for 15 min at 4 °C to remove large aggregates. NCAP, S2hp vRNA, and ZnCl₂ were mixed in PBS at final concentrations of 30 µM NCAP with 0 or 0.75 µM S2hp vRNA, and 0 or 20 µM ZnCl₂ as indicated in the figure. ThS was diluted into the wells to a final concentration of 0.0002% w/v. Blank solutions containing everything but NCAP were prepared as controls. After dispensing the samples the plates were immediately covered with optical film (Corning Sealing Tape Universal Optical) and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C with 700 rpm shaking. The plates were imaged at indicated time points of incubation. All samples were imaged with

ZEISS Axio Observer D1 fluorescence microscope with ZEN 2 software, equipped with a 100x oil objective lens, using the 1,4-Diphenylbutadiene fluorescence channel with a DAPI filter for ThS, as well as a DIC filter. Images were processed and rendered with FIJI (imageJ)[81](#).

PS of the LCD segment with ThS staining (Fig. [3](#))

S2hp and ThS solutions were prepared as above. Purified LCD protein solution was centrifuged at $15,000\times g$ for 15 min at 4 °C to remove large aggregates. The Protein, RNA, and ThS were then mixed in wells of 96-well black/clear glass-bottom plate (Cellvis glass-bottom plates cat. no. P96-1.5H-N) at 40:1 and 4:1 LCD: S2hp vRNA molar ratios in triplicates. ThS was added to 0.0002% w/v final concentration. This experiment was repeated with both 30 and 10 μM final LCD concentrations showing similar results. Respective protein and RNA blank solutions were prepared as controls. The plate was immediately covered with an optical film (Corning Sealing Tape Universal Optical) and incubated at 37 °C with 700 rpm shaking in a plate reader (BMG LABTECH FLUOstar Omega). Images were obtained at indicated time points and processed as above.

PS of NCAP with G12 (Fig. [5](#))

Directly prior to assay setup, purified NCAP protein was centrifuged at $15,000\times g$ for 15 min at 4 °C to remove large aggregates and the supernatant was used for the experiment. S2hp RNA was briefly annealed by heating at 95 °C for 3 min and transferring quickly on the ice. G12 stock solutions were prepared in DMSO in 1 mM concentration from lyophilized peptide powder and serially diluted in PBS buffer complemented with 10 % DMSO and was added to wells of 384-well black/clear glass-bottom plate containing 10 μM NCAP protein and 0.25 μM S2hp RNA (40:1 molar ratio) in PBS buffer. NCAP: G12 (or buffer control) molar ratios are indicated in the

figure. The final DMSO concentration in all wells was 1%. The plate was covered with optical film (Corning Sealing Tape Universal Optical) and incubated for ~4 h at room temperature without shaking prior to imaging. Images were acquired using an Axio Observer D1 microscope (Zeiss) with ZEN 2 software, equipped with a $\times 100$ oil objective lens using a DIC filter. Images were processed and rendered with FIJI (imageJ)[81](#). Mean area and mean circularity (weighted by particle size) of particles and droplets were calculated using MATLAB as described in the Brightfield Image Segmentation and Shape Analysis section.

PS with FITC labeled G12 (Supplementary Fig. 6)

FITC-labeled G12 stock solution (made in DMSO) was added to non-labeled stocks at a 1:9 labeled:non-labeled ratio. The mixture was then added to a 96-well plate with glass bottom at a final concentration of 10 μM NCAP, 0.25 μM S2hp RNA (40:1 molar ratio), and 0 or 10 μM G12 in 20 mM Tris pH 8, 50 mM NaCl, and 20 μM ZnCl_2 . The final DMSO concentration in all wells was 0.5 %. The plate was covered with optical film (Corning Sealing Tape Universal Optical) and incubated at 37 °C without shaking for 24 h prior to imaging with ZEISS Axio Observer D1 fluorescence microscope with ZEN 2 software, equipped with a $\times 100$ oil objective lens, using the FITC fluorescence channel with a GFP filter and a DIC filter. Images were processed and rendered with FIJI (imageJ)[81](#).

Measurements of ThS fluorescence in LCD PS droplets (Fig. 3c)

The PS experiment of the LCD segment with ThS staining was performed as described above. To evaluate the change in ThS fluorescence upon incubation of the PS droplets we combined for each experimental condition and time point 5 fluorescence images per well from triplicate wells and 3 biological repeats ($n = 45$ images per condition per time point). Background fluorescence

was subtracted individually from each image using FIJI after measuring the mean gray value and STD of a region containing no features of interest and calculating it with Eq. (1):

Background fluorescence signal = $3 \times \text{STD} + \text{mean gray value}$

Then the mean fluorescence (gray value) of the entire background subtracted image was measured and averaged across all images from the same condition and time point. The plot was rendered in Prism software and error bars represent standard error of the mean. Two-tailed *t*-test with Welch's correction was performed in Prism to evaluate statistical significance of the change in ThS fluorescence between time points of each condition. Mean area and mean circularity (weighted by particle size) of particles and droplets were calculated using MATLAB as described below in the Brightfield Image Segmentation and Shape Analysis section.

Image segmentation and Shape analysis (Figs. 3 and 5)

Brightfield microscopy images were imported into MATLAB 9.13.0 (R2022b) where all subsequent processing and image analysis were carried out. Image segmentation was carried out by initial Gaussian filtering of each image to achieve local smoothing of the image data. Each image was filtered using a Gaussian kernel with a standard deviation of 5 pixels (px). The Laplacian of the Gaussian-filtered image was then found to highlight areas of rapid change in intensity to facilitate edge detection. Edge detection was performed on each image by finding points of maximum local gradients, using the Sobel approximation to derivatives that are implemented using the MATLAB Image Processing Toolbox. The detected edges on the resulting binary image were then dilated and holes, defined by the connectivity of edges and corners, filled. Regions with an area $< 100 \text{ px}^2$ were removed to reduce segmentation errors. For small regions, defined by an area less than $10,000 \text{ px}^2$, refined segmentation was then carried out

in which each region was extracted from the unprocessed image data using a padded square extraction box with a side length of 1.5 times the maximal length of the region on the xy -plane. Image segmentation was carried out on each extracted small region individually as described above, with the difference of using the Canny algorithm, implemented using the MATLAB Image Processing Toolbox, for edge detection⁸². Each detected region of the segmented image then represented an area of interest for which shape analysis was carried out. For each region, its area was found from the total number of pixels and the circularity of the area was calculated as shown in Eq. (2).

$$\text{circularity} = 4\pi \times \text{area} \times \text{perimeter}^{-2}$$

For the representation of LCD assemblies with S2hp vRNA (Fig. 3), calculated circularity measures were then weighted by the area of each corresponding region in the analysis of the sample means and standard errors of the means. Statistical analysis of the area and weighted circularity of segmented regions from a total of 45 images combined from 5 individual images collected from each technical triplicate of 3 biological repeats, was finally performed and visualized as boxplots showing the 25th percentile, median, and 75th percentile of the mean values for triplicate experiments. The whiskers of the plots extend to the most extreme data points. Observations beyond the whisker length (shown as circles in the figure) are values more than 1.5 times the interquartile range beyond the bottom or top of the box. For the representation of NCAP particles with G12 (Fig. 5), calculated circularity measures were weighted by the area of each corresponding region in the analysis of the sample means. Mean area and mean weighted circularity was calculated across regions of 15 images per experimental condition, obtained by combining 5 images for each technical triplicate. Every biological repeat was analyzed separately. A representative boxplot is shown in the figure, in which the central mark indicates

the median of the experimental triplicate means, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively.

Thioflavin-T assays

All solutions in these experiments were prepared using DNase/RNase-free water (ultrapure water) and were filtered twice using a 0.22- μ m syringe filter. Preparations were done under sterile conditions and using sterile filter pipette tips to ensure RNA preservation. Thioflavin T (ThT) stock solution was freshly prepared from powder (Sigma, CAS ID: 2390-54-7) at a concentration of 20 mM in DNase/RNase ultrapure water, followed by 0.22- μ m filtration.

Thioflavin-T fibrillation kinetic assays

Purified NCAP protein and its segments were separately diluted into 20 mM Tris pH 8.0, 300 mM NaCl buffer at 235 μ M concentration. S2hp RNA was diluted by 10 mM HEPES pH 7.0 buffer to 75 μ M concentration. The proteins, RNA and ThT were then mixed to final concentrations of 300 μ M ThT, 30 μ M protein, and 0 or 7.5 μ M RNA (as indicated in Fig. [1](#) and Supplementary Fig. [2](#)), in 1X PBS pH 7.4. Blank samples containing everything but the protein were prepared. The reaction was carried out in a black 384-well clear-bottom plate (NUNC 384) covered with optical film (Corning Sealing Tape Universal Optical) and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C, with 700 rpm double orbital shaking for 30 s before each measurement. ThT fluorescence was measured with excitation and emission wavelengths of 430 and 485 nm, respectively. Measurements were made with technical triplicates for each sample. All triplicate values were averaged, and blank readings from samples without proteins were averaged and subtracted from the values of corresponding protein

mixtures. The results were plotted against time. The experiment was repeated at least three times on different days.

Thioflavin-T endpoint assay

Purified LCD protein segment was diluted into 20 mM Tris pH 8.0, 300 mM NaCl buffer at 235 μ M concentration. The proteins and ThT were then mixed to final concentrations of 300 μ M ThT and 100 μ M protein, in 1X PBS pH 7.4. A blank sample containing everything but the protein was prepared and measured as a buffer control. Fibril formation was carried out in parafilm-covered PCR tubes, incubated in a floor shaker (Torrey Pines Scientific Inc, Orbital mixing chilling/heating plate) at 37 °C, with fast mixing speed for 11 days. 30 μ L of the samples were taken out of the tubes at days 1, 6, and 11 of incubation and put in a black 384-well clear-bottom plate (NUNC 384) covered with optical film (Corning Sealing Tape Universal Optical) and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C, with 700 rpm double orbital shaking for 30 s before the measurement. ThT fluorescence was measured with excitation and emission wavelengths of 430 and 485 nm, respectively.

Turbidity assay

All solutions were prepared using DNase/RNase-free water (ultrapure water) and were filtered twice using a 0.22- μ m syringe filter. Preparations were done under sterile conditions and using sterile filter pipette tips to ensure RNA preservation. Protein and RNA working solutions were prepared as described above for the ThT experiment of NCAP and its segments. Each reaction sample contained 30 μ M protein and 0 or 7.5 μ M RNA in 1X PBS pH 7.4. Blank samples contained everything but the protein. The reaction was carried out in a black 384-well clear-bottom plate (NUNC 384) covered with optical film (Corning Sealing Tape Universal Optical)

and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C, with mixing before and between measurements. Turbidity was measured with absorbance (OD) at 600 nm. Measurements were made with technical triplicates for each sample. Triplicate values were averaged, and appropriate blank readings (samples without the protein) were averaged and subtracted from the corresponding readings. The results were plotted against time. The experiment was repeated at least three times on different days.

Negative stain transmission electron microscopy (TEM)

Samples for negative staining TEM were prepared as described below. All solutions in these experiments were prepared using DNase/RNase-free water (ultrapure water) and were filtered twice using a 0.22- μ m syringe filter. Preparations were done under sterile conditions and using sterile filter pipette tips to ensure RNA preservation. For grid preparation and screening, 4 μ L of each sample was applied directly onto 400-mesh copper TEM grids with Formvar/Carbon support films (Ted Pella), glow discharged (PELCO easiGlowxs) for 45 s at 15 mA immediately before use. Grids were incubated with the samples for 2 min, then the samples were blotted off using filter paper. The grids were washed three times with water and once with 2% uranyl acetate solution with blotting after each wash. The grids were then incubated with 6 μ L of uranyl acetate solution for 30–45 s before blotting. Micrographs were imaged using an FEI Tecnai T12 microscope at room temperature with an accelerating voltage of 120 kV. Images were recorded digitally with a Gatan US 1000 CCD camera, using the Digital-Micrograph® Suite software, and processed in the ImageJ⁸³ software.

NCAP fibrils from PS droplets formed in PBS

NCAP samples with and without 0.75 μM S2hp and 20 μM ZnCl_2 (Supplementary Fig. [2b](#)) were prepared in PBS as described in the PS method section. Samples were vigorously scrapped from the bottom of the wells after 6 days of incubation using a 100 μl pipette tip and used for TEM grid preparation. A blank control containing 0.75 μM S2hp, 20 μM ZnCl_2 and 0.0002% w/v ThS in PBS was imaged as well.

NCAP fibrils in 2 mM Tris pH 8.0, 30 mM NaCl (Supplementary Fig. [2d](#))

Purified NCAP was diluted to 50 μM final concentration from its stock solution (made in 20 mM Tris pH 8.0, 300 mM NaCl buffer) into ultrapure water supplemented with ZnCl_2 in 0 (water only) or 20 μM final concentration. Samples were incubated for 3 days with acoustic resonance mixing at 37 $^\circ\text{C}$ using a custom-built 96-well plate shaker set to 40 Hz. The samples were then recovered and applied to the EM grid as described above.

Fibrils of NCAP and its segments in PBS

NCAP (Supplementary Fig. [2c](#)) and its segments (Fig. [1f](#)) were separately diluted to 235 μM concentration by 20 mM Tris pH 8.0, 300 mM NaCl. The S2hp RNA was diluted to 250 μM by 10 mM HEPES, pH 7.0 buffer. The proteins and RNA were further diluted in 1X PBS pH 7.4 such that each reaction sample contained 100 μM protein and 0/ 25 μM RNA. Fibril formation was carried out in parafilm-covered PCR tubes, incubated in a floor shaker (Torrey Pines Scientific Inc, Orbital mixing chilling/heating plate) at 37 $^\circ\text{C}$, with fast mixing speed for 6 (LCD and DD-C_{term}) to 14 (NCAP) days.

LCD fibril formation with short RNA segments (Fig. [2b](#))

RNA stock solutions were thawed, then annealed by heating at 95 $^\circ\text{C}$ for 3 min and transferring quickly on the ice. The RNAs were diluted to 1 mM concentration by their original buffer of

10 mM HEPES pH 7.0. LCD protein stock was freshly thawed and added together with the appropriate RNA solution into 1X PBS to reach a 1:2 protein:RNA molar ratio at 50 μ M final concentration of LCD, in 50 μ L final volume in a black 384-well clear-bottom plate (NUNC 384). The plate was covered with optical film (Corning Sealing Tape Universal Optical) and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C with shaking. Samples were taken for TEM screening after 12 h (day 1) and 4 days of incubation.

X-ray fiber diffraction

LCD with and without S2hp vRNA

1.27 mM purified LCD stock solution was thawed and dialyzed in a dialysis cassette with a 3.5 kDa cutoff (Thermo Scientific cat. no. 87724) for 4 h at RT in 20 mM Tris pH 7.4, 50 mM NaCl buffer with or without the addition of S2hp vRNA in 4:1 LCD:S2hp molar ratio (955 μ M protein and 236 μ M RNA). After dialysis, the samples were added to a black 384-well clear-bottom plate (NUNC 384), covered with optical film (Corning Sealing Tape Universal Optical), and incubated in a plate reader (BMG LABTECH FLUOstar Omega) at 37 °C, with 30 s of 700 rpm double orbital shaking every 5 min for 3 weeks. The fibrils were pelleted and washed three times in water by centrifugation at 13,000 \times *g* for 10 min at RT, then pelleted again and resuspended in 5 μ L of deionized water. Fibrils were aligned by pipetting 2 μ L of the fibril resuspension in a 3 mm gap between two fire-polished glass rods, positioned end-to-end. After 1 h of drying at room temperature, another 2 μ L of the fibril suspension was applied, thickening the sample. After another hour of drying, the aligned fibril sample was transferred to the exterior of a standard crystal mounting loop. To glue the sample to the loop, the loop was wetted with 50 % v/v ethylene glycol solution, then touched to the surface of the sample and immediately plunged in liquid nitrogen. The samples were shipped to the Advanced Photon Source, beamline

24-ID-E at Argonne National Laboratory for remote data collection. The sample was kept at 100 K using a nitrogen cryo-stream. Diffraction patterns were collected on a Dectris Eiger 16M pixel detector using a 2 s exposure at 100% transmission and 1-degree rotation. The X-ray beam wavelength was 0.9792 Å and impinged on the sample only, avoiding the loop and ethylene glycol, so these later materials do not contribute to the diffraction pattern. The detector was placed 350 mm from the sample. Diffraction images were displayed with the ADXV program (Scripps).

LCD with non-specific RNA (antisense siDGCR8-1 RNA)

LCD stock solution was concentrated to 2.2 mM and the buffer was exchanged to 20 mM Tris pH 8.0, 150 mM NaCl in a centrifugal filter with 3 kDa cutoff (Millipore Sigma Amicon Ultra cat. no. C82301). Antisense siDGCR8-1 RNA stock solution, stored at -20 °C, was thawed and combined with the LCD solution in 1:3 LCD:RNA molar ratio. The solution was titrated to reach a final pH of ~5 as confirmed with pH paper. The final protein concentration was 283 µM and RNA concentration was 849 µM. The reaction mixture was incubated in a floor shaker (Torrey Pines Scientific Inc, Orbital mixing chilling/heating plate) at 37 °C, with rapid mixing speed for 7 days. The fibrils were prepared and mounted as described above except that the fibrils were aligned with a single application of 5 µL of the fibril suspension, rather than two smaller applications. Diffraction was measured at beamline 24-ID-C, rather than 24-ID-E. Diffraction patterns were collected on a Dectris Eiger2 16M pixel detector using a 1 s exposure at 90% transmission and 0.5-degree sample rotation. The X-ray beam wavelength was 0.9791 Å and impinged on the sample only, avoiding the loop and ethylene glycol, so these later materials do not contribute to the diffraction pattern. Exposures were collected at sample-to-detector distances of 200 and 500 mm. Diffraction images were displayed with the ADXV program (Scripps).

Crystallization of NCAP peptide segments

The NCAP segment ²¹⁷AALALL²²² crystallized in batch just before the purification by RP-HPLC. The peptide had been deprotected and cleaved from the resin, triturated with cold diethyl ether, and precipitated. Most of the product had been collected via filtration, but some residual peptide remained in the round bottom flask and we intended to use this residual peptide to check the peptide purity by analytical HPLC. We dissolved the residual peptide with water, acetonitrile, and TFA in a volume ratio of approximately 45:45:10 and transferred it to a 1 mL glass vial for HPLC injection. The solution was left in the sample holder and needle-like crystals formed after a week. Some of these crystals were retained for crystal structure determination. The bulk of the peptide was further purified, as described above. Later, we showed we could reproduce the crystals by dissolving 0.75 mg of AALALL in 50 μ L of TFA and then diluting with 225 μ L of acetonitrile and 225 μ L of water. This was left to sit in an HPLC vial which had its septum top poked open once with the HPLC injection needle, and the same crystal form appeared in 3 months.

We screened for additional AALALL crystals using 96-well kits and purified peptide dissolved at 10 mg/mL concentration in 19.6 mM LiOH. Crystals were grown by the hanging drop vapor diffusion method. The UCLA Crystallization Facility set up crystallization trays with a Mosquito robot dispensing 200 nL drops. Needle-shaped crystals of ²¹⁷AALALL²²² grew at 20 °C in a reservoir solution composed of 30% w/v polyethylene glycol (PEG) 3000 and 0.1 M n-cyclohexyl-2-aminoethanesulfonic acid (CHES), pH 9.5. The purified NCAP segment ¹⁷⁹GSQASS¹⁸⁴ was dissolved in water at 100 mg/mL concentration. Hanging drop crystallization trays were set using 200 nL drops. Needle-shaped crystals grew at 20 °C using a reservoir solution composed of 1.0 M Na, K tartrate, 0.2 M Li₂SO₄, and Tris pH 7.0. Needle-

shaped crystals appeared immediately after setting up the tray. The purified NCAP segment $_{243}\text{GQTVTK}_{248}$ was dissolved in water at 68 mg/mL concentration. Hanging drop crystallization trays were set using 200 nL drops. Needle-shaped crystals appeared within 1 day at 20 °C using a reservoir solution composed of 2.0 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M sodium HEPES, pH 7.5, and 2% v/v PEG 400.

Structure determination of NCAP peptide segments

Microfocus X-ray beam optics were required to measure crystal diffraction intensities from our crystals since they were needle-shaped and less than 5 microns thick. We used microfocus beamline 24-ID-E of the Advanced Photon Source located at Argonne National Laboratory. Crystals were cooled to a temperature of 100 K. Diffraction data were indexed, integrated, scaled, and merged using the programs XDS and XSCALE⁸⁴. Data collection statistics are reported in Table 1. Initial phases for AALALL and GSQASS were obtained by molecular replacement with the program Phaser⁸⁵ using a search model consisting of an ideal β -strand with sequence AAAAAA. Phases for GQTVTK were obtained by direct methods using the program ShelxD⁸⁶. Simulated annealing composite omits maps⁵⁷ were calculated using Phenix⁸⁷. Refinement was performed using the program Refmac⁸⁸. Model building was performed using the graphics program Coot⁸⁹. Structure illustrations were created using PyMOL⁸⁰. Residue hydrophobicity of the steric zipper segments was assigned and colored according to the Kyte and Doolittle hydrophobicity scale embedded in UCSF Chimera⁹⁰.

G12 evaluation in HEK293-ACE2 cells infected with SARS-CoV-2

Lyophilized G12 peptide powder was dissolved in 100 % DMSO (Sigma cat. no. D2650) to approximately 10 mM, centrifuged at $21,000\times g$ for 30 min to remove large aggregates, then

aliquoted and stored at $-20\text{ }^{\circ}\text{C}$ until use. To determine peptide concentrations accurately, the stock was diluted in UltraPure distilled water (ThermoFisher cat. no. 10977015), and the concentration was measured using the Pierce Quantitative Fluorometric Peptide Assay (ThermoFisher cat. no. 23290). HEK293-ACE2 cells (ATCC, cat. no. CRL-3216, authenticated and quality tested by ATCC [<https://www.atcc.org/products/crl-3216>]) stably over-expressing the human ACE2 receptor⁹¹ were cultured in DMEM (Gibco cat no. 11995-065) supplemented with 10% FBS (Gibco cat no. 26140-079), 1% penicillin-streptomycin (Gibco cat no. 15140-122), 10 mM HEPES pH 7 (Gibco cat no. 15630106), 50 μM 2-mercaptoethanol (Sigma cat no. M3148), and 1 $\mu\text{g}/\text{mL}$ puromycin (Gibco cat no. A1113803) for selection, at $37\text{ }^{\circ}\text{C}$, 5% CO_2 . Cells were confirmed negative for mycoplasma by PCR using a Universal Mycoplasma Detection Kit (ATCC cat. no. 30-1012K). The HEK293-ACE2 cells were plated in 96-well black/clear plates (Greiner Bio-One cat. no. 655090) at 2×10^4 cells per well. The cells were incubated for 1–2 days at $37\text{ }^{\circ}\text{C}$, 5% CO_2 , then exchanged into antibiotic-free media and incubated for an additional day. Cells were then transfected with the peptide-based inhibitors, either unlabeled (Fig. 5d and Supplementary Fig. 8; Final peptide concentrations are detailed in the figures), or with $\sim 15\text{ }\mu\text{M}$ of FITC-labeled G12 (Supplementary Fig. 7) by diluting stock solutions (made in 5% DMSO) into cell culture medium to a 10X concentration, and serially diluting from there for dose–response assays while maintaining similar DMSO concentration in all peptide dosages (Fig. 5d and Supplementary Fig. 8). 10 μL of 10X peptide diluted in culture medium was added to 90 μL media in each well, for a final DMSO concentration of 0.5% in all wells. Finally, Endo-Porter (PEG-formulation) transfection reagent (GeneTools LLC, Philomath, OR) was added to each well at a final concentration of 6 μM . Plates were incubated for 2- 4 h, then the cells were infected with SARS-CoV-2 (Isolate USA-WA1/2020) (BEI Resources) in the

UCLA BSL3 High-Containment Facility⁹¹ by adding the virus in 200 μ l final volume at an MOI of 0.05 for evaluation of dose dependence antiviral activity with the inhibitor G12 (Fig. [5d](#) and Supplementary Fig. [8](#)). The uninfected control received only the base media used for diluting the virus. The plates were incubated for an additional 24 h at 37 °C, 5% CO₂, and fixed with 100% methanol for immunofluorescence assay. Fixed cells were washed 3 times with PBS pH 7.4 (Gibco cat. no. 10010-023) and incubated with blocking buffer (2% BSA, 0.3% Triton X-100, 5% goat serum, 5% donkey serum, 0.01% NaN₃ in PBS) for 2 h at room temperature. Anti-Spike protein primary antibody was diluted into blocking buffer and incubated overnight at 4 °C. Either of these primary anti-Spike protein antibodies was used (depending on availability): BEI Resources, NIAID, NIH rabbit monoclonal Anti-SARS-Related Coronavirus 2 Spike Glycoprotein S1 Domain (produced in vitro), cat. no. NR-53788, clone no. 007, Lot: HA14AP3001 (purchased from SinoBiological, cat. no. 40150-R007), at a 1:100 dilution ratio, or BEI Resources, NIAID, NIH: Mouse Monoclonal Anti-SARS-CoV S Protein (Similar to 240C), cat. no. NR-616, Lot: 102204 (purchased from ATCC), at a 1:300 dilution ratio. Following overnight incubation, cells were washed with PBS and incubated for one hour at room temperature with AlexaFluor-555 conjugated secondary goat anti-mouse (Abcam cat. no. ab150114, Lot: GR299321-5), or goat anti-rabbit (Abcam cat. no. ab150078, Lot: GR302355-2) antibody, diluted at 1:1000. All antibodies used in this section were validated by their respective vendors. Following incubation with the secondary antibody, the cells were stained with 10 μ g/mL DAPI (ThermoFisher cat. no. D1306) for 10 min, and stored in PBS for imaging. Plates were imaged using an ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices, San Jose, CA) in widefield mode at 10X magnification. 9 sites per well were imaged, and the percentage of infected cells was quantified using the MetaXpress

multiwavelength cell scoring module. We considered spike protein-expressing cells as infected and calculated their percentage from the total number of cells in the well. Raw values were exported into Microsoft Excel, and percent-infected cells were then normalized to an infected culture that was treated with vehicle only. Statistical analysis was performed using one-way ANOVA in GraphPad Prism, and IC₅₀ values were estimated (Fig. [5d](#)) using a four-parameter non-linear fit dose–response curve.

Cytotoxicity assay in HEK293-ACE2 cells (Fig. [5d](#))

HEK293-ACE2 cells were plated and transfected with peptides following the same protocol as used for the viral assays, but following transfection were incubated at 37 °C and 5% CO₂ for 24 h. Peptide cytotoxicity was then assessed using the CyQUANT LDH Cytotoxicity Assay (ThermoFisher cat no. C20300) following the manufacturer protocol. Absorbance was measured at 490 and 680 nm (background subtraction) using a SpectraMax M5 (Molecular Devices) with SoftMaxPro v5.3 software.

Statistics and reproducibility

All turbidity and ThT fibrillation kinetic experiments were repeated three independent times with technical triplicates. Technical triplicates were averaged and blank subtracted. Representative curves are presented in the figures. Endpoint ThT measurements of the LCD-only segment were done using three samples. Each sample was measured once per every time point. X-ray diffractions of LCD only and LCD+ S2hp fibrils were each collected three times on different days, using different diffractometers and x-ray sources while showing similar results. Diffraction of LCD+ non-specific RNA fibrils was collected twice from different regions of the same loop, showing similar results. EM micrographs of LCD-only fibrils were captured at least five

independent times. LCD fibrils with the different vRNA segments were visualized by EM at least two independent times per vRNA type, once of which with different time points. NCAP with and without S2hp vRNA in PBS was imaged by EM from two independent samples. Other EM images were taken from a single sample. PS of the LCD-only segment with ThS and PS of NCAP with and without different concentrations of G12 were each performed three independent times with technical triplicates showing similar results. PS of NCAP with ThS was repeated twice (2nd repeat incubated for 3 days only) showing ThS partitioning into NCAP's PS droplets. FITC-labeled G12 was tested on NCAP PS droplets in vitro once. Antiviral activity of G12 in cells was tested three independent times with G12 concentrations of over 10 μ M showing inhibition of ~40–60% in viral infectivity. Full dose response of G12 and its cytotoxicity in cells was tested in triplicated wells. Distribution of FITC labeled G12 in HEK293-ACE2 cells was tested two independent times with duplicated wells.

Data availability

Atomic coordinates that support the findings of this study are available in the RCSB Protein Data Bank (PDB) under accession numbers: 7LV2 [<https://doi.org/10.2210/pdb7LV2/pdb>], 7LTU [<https://doi.org/10.2210/pdb7LTU/pdb>] (form 1), 7LUX [<https://doi.org/10.2210/pdb7LUX/pdb>] (form 2), and 7LUZ [<https://doi.org/10.2210/pdb7LUZ/pdb>]. The amino acid sequences of the Nucleocapsid proteins of SARS-CoV-2 and SARS-CoV analyzed in this study are available on UniProtKB, accession numbers: [P0DTC9](#), and [P59595](#) respectively. Amino acid sequences of other coronavirus Nucleocapsid proteins were accessed from the European Nucleotide Archive [ENA; <https://www.ebi.ac.uk/genomes/virus.html>]. Raw EM images, light and fluorescence microscopy images and fiber diffraction source files generated in this study have been deposited in the Figshare respiratory at

[\[https://figshare.com/projects/Low_Complexity_Domains_of_the_Nucleocapsid_Protein_of_SA_RS-CoV-2_Form_Amyloid_Fibrils/162391\]](https://figshare.com/projects/Low_Complexity_Domains_of_the_Nucleocapsid_Protein_of_SA_RS-CoV-2_Form_Amyloid_Fibrils/162391). Data for all plots presented in this manuscript are provided with this paper in the Source Data file. [Source data](#) are provided with this paper.

Acknowledgements

We thank Megan Bentzel, Jose Rodriguez, Meytal Landau, and Mark Arbing for the discussions. We thank the staff at the Northeastern Collaborative Access Team, which is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beamline is funded by an NIH-ORIP HEI grant (S10OD021527). The Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Some of this work was also performed at the Stanford-SLAC Cryo-EM Center (S2C2), which is supported by the National Institutes of Health Common Fund Transformative High-Resolution Cryo-Electron Microscopy program (U24 GM129541). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors also acknowledge the use of instruments at the Electron Imaging Center for NanoMachines supported by NIH (1S10RR23057 to ZHZ) and CNSI at UCLA. Mass spectrometry data were collected on instrumentation maintained and made available through the support of the UCLA Molecular Instrumentation Center—Mass Spectrometry Facility in the Department of Chemistry. This material is based upon work supported by the National Science Foundation under Grant No. (MCB 1616265), NIH/NIA R01 Grant AG048120, the U.S. Department of Energy (DOE) Contract No. DOE-DE-FC02-02ER63421, and by UCLA David Geffen School of Medicine—Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Award Program,

Broad Stem Cell Research Center (BSCRC) COVID 19 Research Award (OCRC #20-73). This study is also supported by the UCLA W.M. Keck Foundation COVID-19 Research Award and National Institute of Health awards 1R01EY032149-01, 5U19AI125357-08, 5R01AI163216-02 and 1R01DK132735-01 to V.A. The Human Frontiers Science Project Organization (HFSPO) (LT000623/2018-L) supported E.T-F. NIH NIGMS GM123126 grant supported Luk.S. C.-T.Z. was funded by the UCLA Dissertation Year Fellowship.

Author Contributions

Constructs design and cloning: P.M.S., Luk.S. Protein preparation and experimentation: E.T-F, J.T.B., S.L.G., X.C., R.A., J.L., Y.X.J. RNA preparation and experimentation: C.E.T., Y.L., Luk.S. Peptide preparation: C.-T.Z. X-ray crystallography: M.R.S., C.-T.Z, J.L., K.H., G.F., D.C. Fluorescence and electron Microscopy: E.T.-F., J.T.B., X.C., D.R.B., R.A., Y.X.J., H.P., G.M.R., J.L. Computational analysis and peptide self-assembly modulators design: G.M.R., P.M.S., Y.X.J., E.T-F., Lor.S., K.A.M. Brightfield image segmentation and shape analysis in MATLAB: L.L. In-cell assays: J.T.B., G.G. Jr. Writing and figure preparation: E.T.-F., M.R.S., J.T.B., F.G., D.S.E. Technical support: D.H.A., Project management: E.T.-F., M.R.S., R.D., V.A., F.G., and D.S.E.

Competing interests

D.S.E. is an advisor and equity shareholder in ADRx, Inc. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s41467-023-37865-3>.

Correspondence and requests for materials should be addressed to David S. Eisenberg.

Peer review information Nature Communications thanks Nicholas Reynolds, Dan Li and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Figures

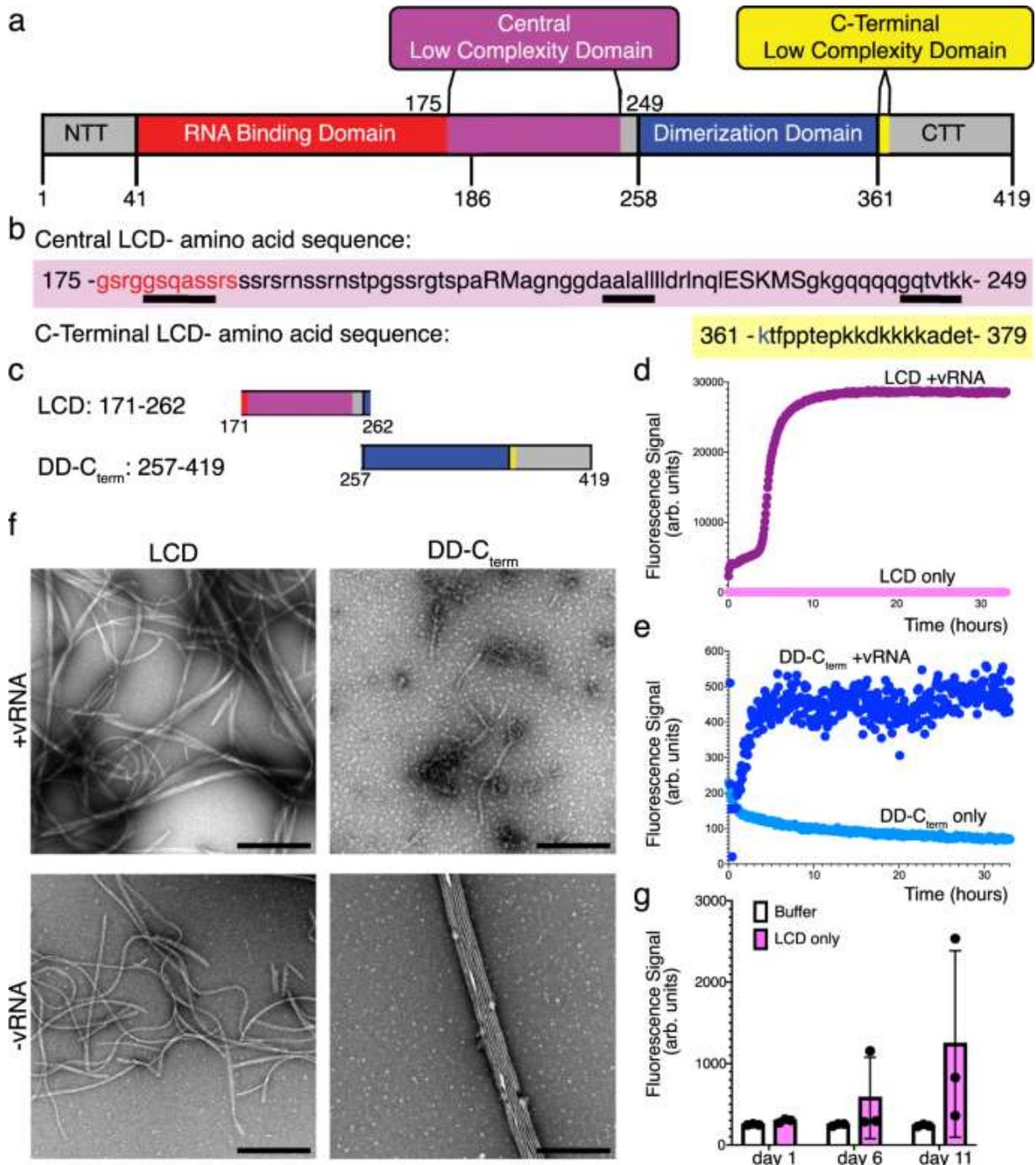


Fig. 1: NCAP's LCDs form fibrils and ThT-positive species.

a NCAP's domain organization. Domain definitions: N-terminal tail (NTT, gray), RNA-binding domain (red); Central low complexity domain (LCD, purple; residues 175–249), Dimerization domain (blue); C-terminal tail (CTT, gray). The C-terminal LCD is highlighted in yellow (residues 361–379). **b** Amino acid sequence of the central and C-terminal LCDs highlighted and colored according to the color scheme in (a). Lowercase letters represent residues of low

complexity while capital letters represent non-low-complexity residues. No more than five interrupting non-low-complexity residues between strings of 10 or more low-complexity residues were allowed. Steric-zipper forming sequences that are discussed below are underlined in the sequence of the central LCD. **c** Protein segments used in this study are abbreviated as LCD, consisting of the central LCD and surrounding residues, and as DD-C_{term}, consisting of the dimerization domain (DD) and the C-terminal tail, including the C-terminal LCD. The LCD and DD-C_{term} segments are colored according to the color scheme in **(a)**. **d** and **e** ThT fibril formation kinetic assays of the LCD (**d**) and DD-C_{term} (**e**) segments incubated with (purple/navy, respectively) and without (pink/light blue, respectively) hairpin-Site2 (S2hp) viral RNA (vRNA). **f** Fibril formation from concentrated LCD and DD-C_{term} samples observed by negative stain EM after 6 days of incubation with and without S2hp vRNA. Scale bar = 500 nm. **g** Endpoint ThT fluorescence measurements of concentrated LCD-only samples (pink) and buffer-only controls (white) at days 1, 6, and 11 of incubation. Dots indicate individual data points and bars represent mean values \pm SD. $n = 3$ samples. Source data for panels **d**, **e**, and **g** are provided as a Source Data file.

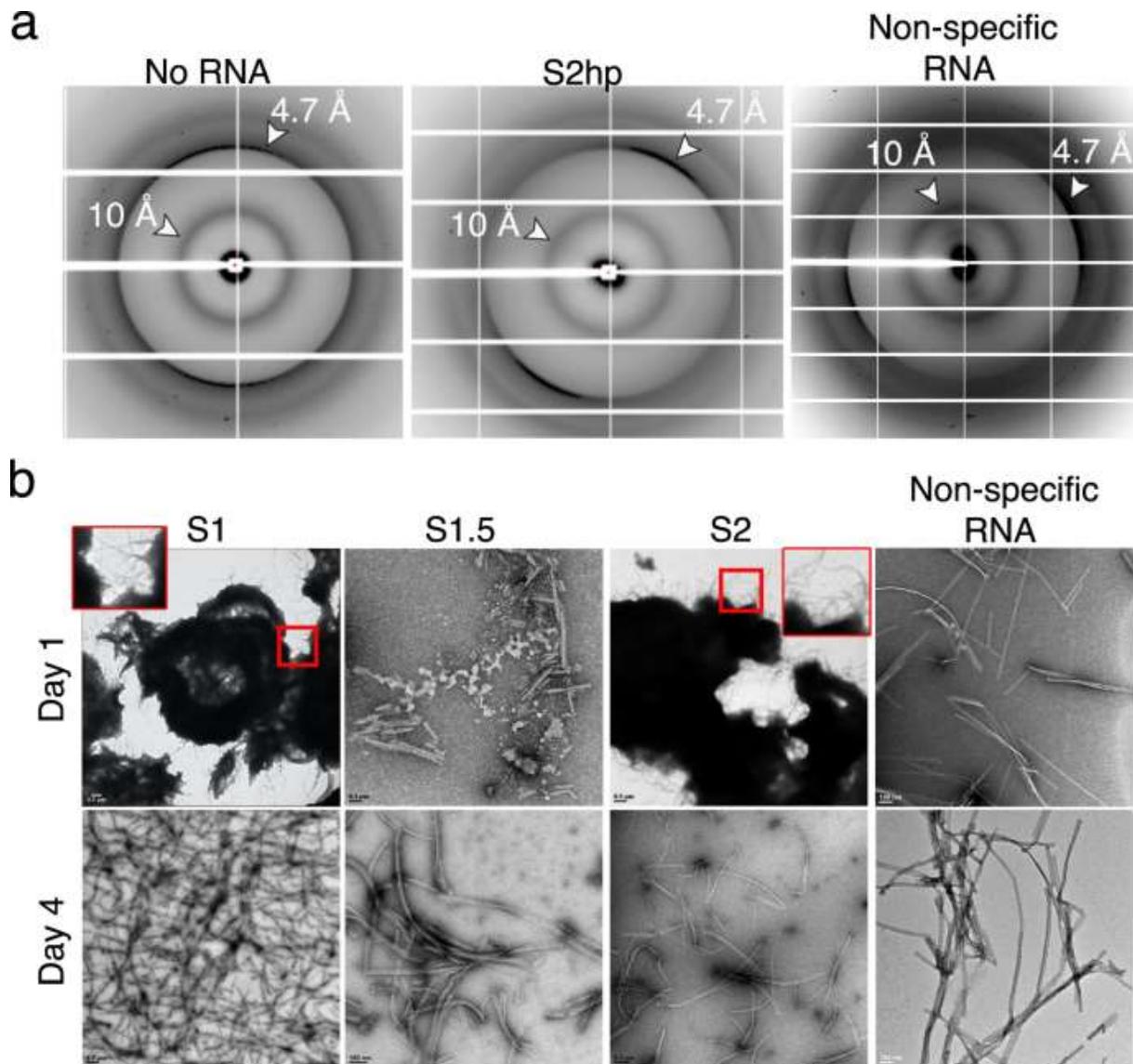


Fig. 2: The central LCD segment of NCAP demonstrates amyloid-like characteristics.

a X-ray diffractions of LCD-only fibrils (No RNA), and LCD fibrils grown with hairpin-Site2 (S2hp) vRNA or non-specific RNA (antisense siDGCR8-1), show amyloid-characteristic 4.7 and 10 Å reflections labeled by white arrows. **b** Negative stain EM micrographs of LCD fibrils grown in the presence of the short vRNA segments Site1 (S1; 11 nucleotides), Site 1.5 (S1.5; 22 nucleotides) and Site2 (S2; 22 nucleotides), as well as with a non-specific RNA (antisense siDGCR8-1). All RNA sequences are given in Supplementary Table 1. This figure shows that the central LCD produces amyloid-typical fibrils in the absence and presence of different RNA segments and that the RNA sequence may influence the morphology of the LCD assemblies over time.

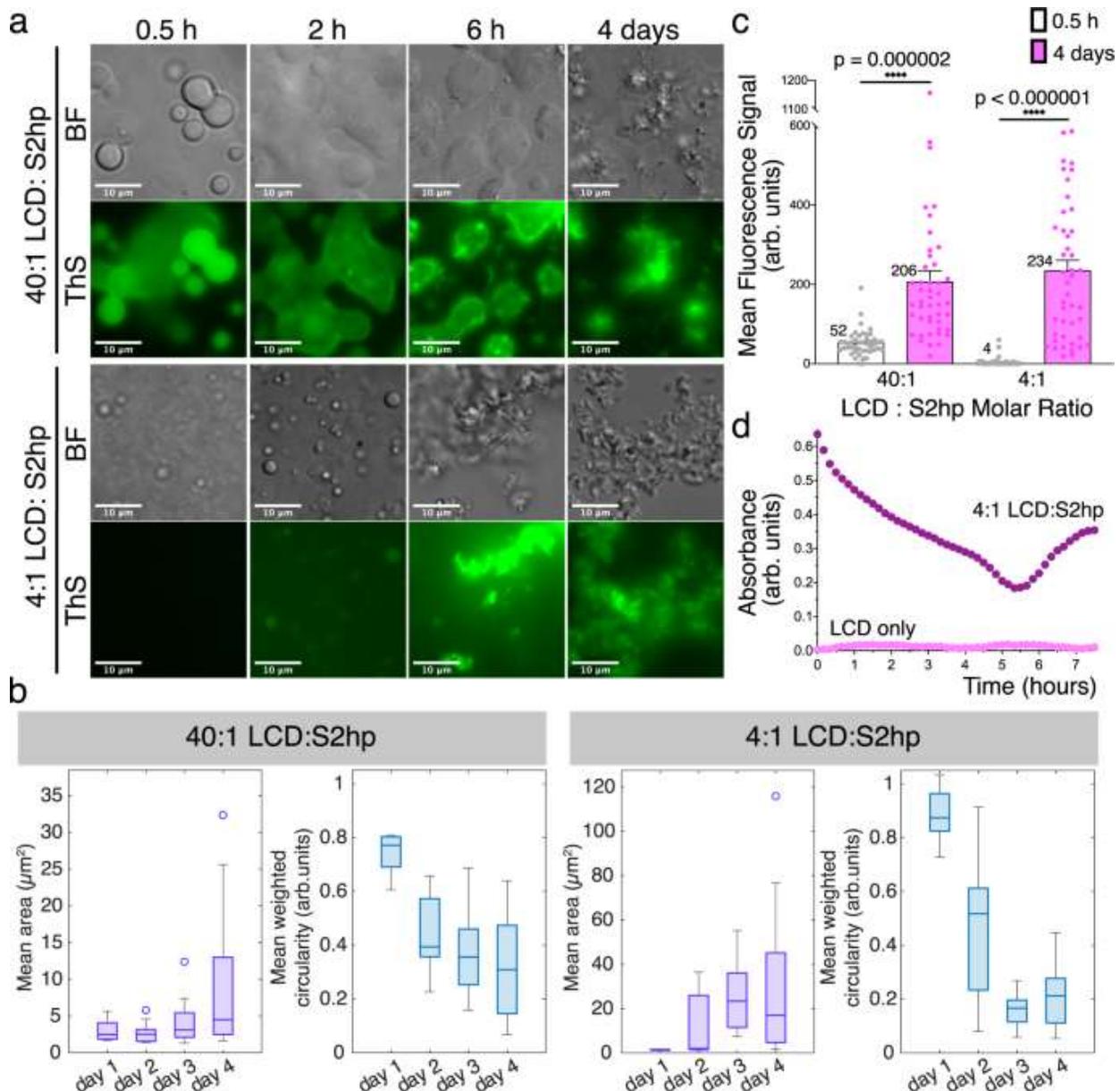


Fig. 3: The central LCD segment of NCAP forms ThS fluorescent PS droplet that transition into amorphous and fibrous solid-like particles.

a Brightfield (BF) and Thioflavin-S (ThS) fluorescence (green) microscopy images of 40:1 and 4:1 LCD: hairpin-Site2 (S2hp) vRNA molar ratio mixtures incubated for ~0.5, 2, 6 h and 4 days. **b** Mean area (purple) and mean circularity (blue; normalized to particle size) of droplets and particles quantified from a series of light microscopy images of 40:1 and 4:1 LCD: S2hp mixtures imaged at day 1–4 of incubation. The experiment was performed in three biological repeats, each with technical triplicates. Five images were collected for every technical replicate. Boxplots show the 25th percentile, median, and 75th percentile of the mean values for triplicate experiments. The whiskers extend to the most extreme data points. Observations beyond the whisker length, shown as circles, are values more than 1.5 times the interquartile range beyond the bottom or top of the box ($n = 9$ replicates). **c** Mean ThS fluorescence signal measured from

background-subtracted fluorescence microscopy images taken from 40:1 and 4:1 LCD: S2hp mixtures at 0.5 h (white) and 4 days (pink) of incubation. The experiment was performed in three biological repeats, each with technical triplicates. Five images were collected for every technical replicate. Data from all repeats were combined for the quantification. The dots are of individual data points and the bars represent mean values \pm SEM ($n = 45$ images). Statistical significance was calculated in Prism using an unpaired two-tailed t-test with Welch's correction. The p values are indicated with numbers and stars—**** $p < 0.0001$. Welch's corrected $t = 5.377/ 8.597$ and $df = 46.59/44.33$ for 40:1 and 4:1 LCD: S2hp samples, respectively. **d** Time-dependent shift in turbidity of LCD only (pink) and 4:1 LCD: S2hp (purple) solutions evaluated by measuring absorbance at 600 nm. Source data for panels **b–d** are provided as a Source Data file.

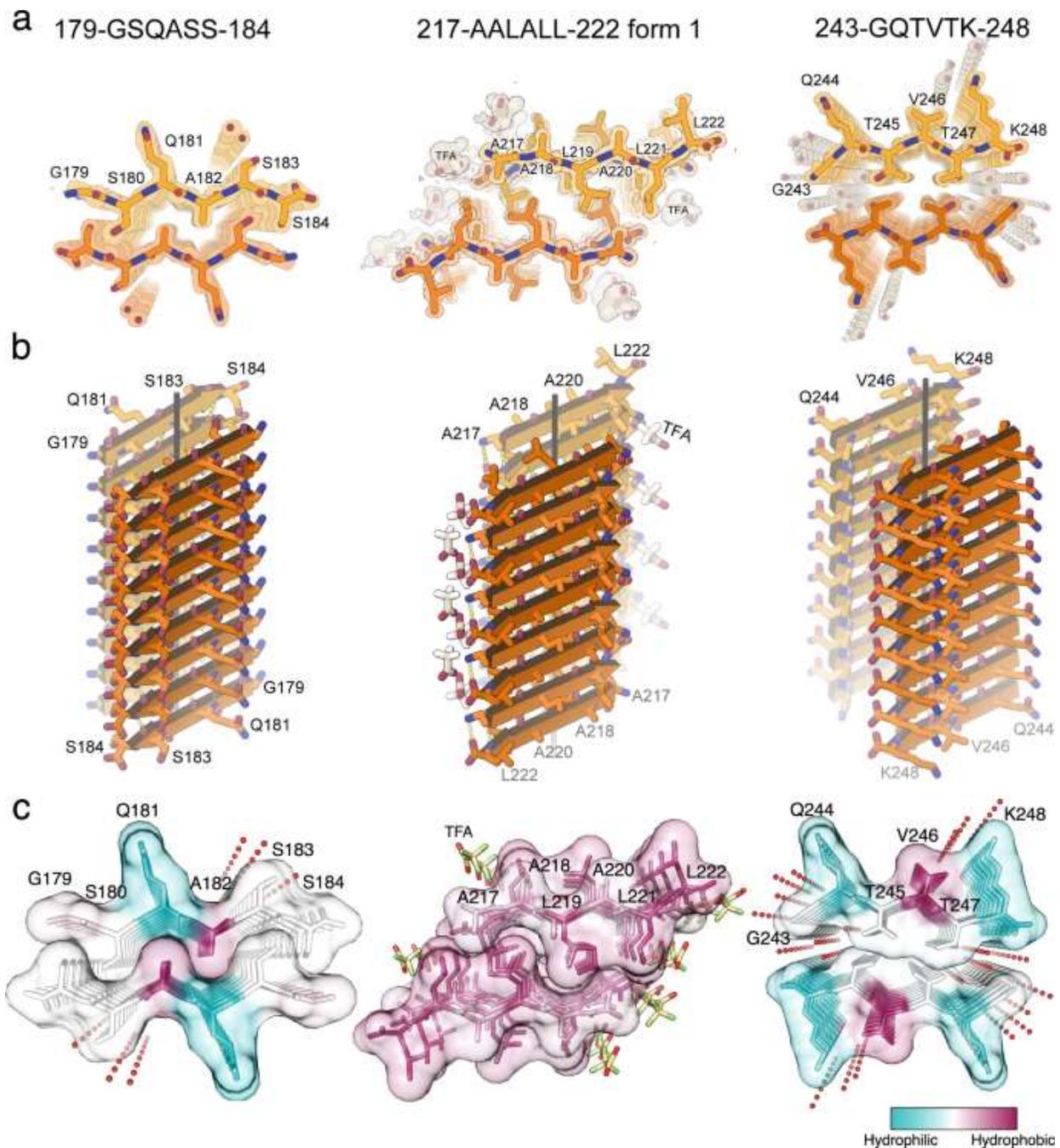


Fig. 4: Atomic structures of amyloid-like association of NCAP segments revealed by crystallography.

a Quality of the fit of each atomic model to its corresponding simulated annealing composite omit maps⁹². The maps are contoured at the 1.0 sigma level. All structural features are well defined by the density. The view is down the fibril axis. Each chain shown here corresponds to one strand in a β -sheet. Thousands of identical strands stack above and below the plane of the page making ~100 micron-long β -sheets. The face of each β -sheet of AALALL (PDB 7LTU) [<https://doi.org/10.2210/pdb7LTU/pdb>] (form 1) is symmetric with its back. However, GSQASS

(PDB 7LV2) [<https://doi.org/10.2210/pdb7LV2/pdb>] and GQTVTK (PDB 7LUZ) [<https://doi.org/10.2210/pdb7LUZ/pdb>] each reveal two distinct sheet–sheet interfaces: face-to-face and back-to-back. The tighter associated pair of sheets is shown in this figure. **b** 18 strands from each of the steric zippers at a view nearly perpendicular to the fibril axis. GSQASS and GQTVTK are parallel, in-register sheets, mated with Class 1 zipper symmetry. The AALALL zippers are antiparallel, in register sheets, mated with Class 7 zipper symmetry. Trifluoroacetic acid (TFA) appears in the AALALL-form 1 steric zipper, and polyethylene glycol (PEG) binds form 2 (PDB 7LUX [<https://doi.org/10.2210/pdb7LUX/pdb>] (form 2); Supplementary Fig. 4). Carbon atoms in **a** and **b** are shown in orange and heteroatoms are colored by atom type. Water molecules are shown as red dots. **c** Steric zipper structures (same order as in **a**) viewed down the fibril axis with residues colored according to the Kyte and Doolittle hydrophobicity scale (bottom right) shown with semi-transparent surface representation to emphasize the shape complementarity and tight fit between the β -sheets. Images in **c** were rendered with UCSF Chimera⁹⁰. A stereo view of all structures is given in Supplementary Fig. 5.

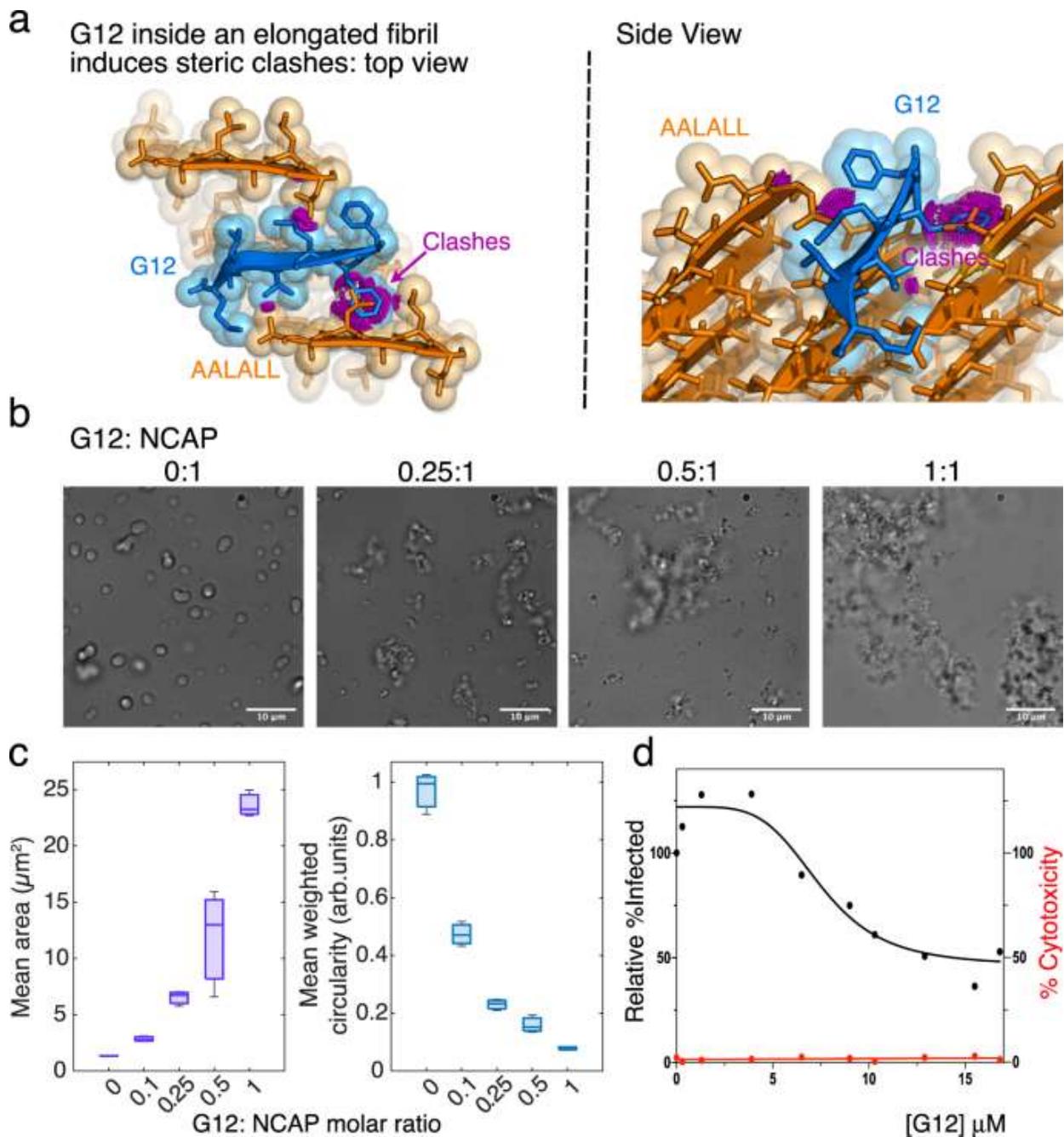


Fig. 5: Design and evaluation of NCAP's self-assembly disruptor, G12.

a The Rosetta-based design of G12 templated by the AALALL X-ray crystal structure form 1 (Fig. 4; Table 1). Model of the G12 (blue) capping an AALALL fibril (orange). The top view is down the fibril axis and the side view is tilted from the axis. Additional AALALL strands are shown adjacent to the bound G12 to illustrate their steric clashes (magenta). **b** Differential interference contrast (DIC) images of NCAP + S2hp mixtures incubated in the absence (0:1) and presence of increasing concentrations of G12 revealing the PS disrupting activity of G12. **c** Mean area (purple) and mean circularity (blue; normalized to particle size) of droplets and particles

quantified from a series of light microscopy images of NCAP + S2hp mixtures with increasing concentrations of G12. The experiment was performed in three biological repeats, each with technical triplicates. Five images were collected for every technical replicate. A representative plot is presented. In boxplots the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points ($n = 3$ replicates). **d** Dose–response analysis of HEK293-ACE2 cells treated with 10 different concentrations of G12 and fitted with a nonlinear regression model (black line). The 95% confidence interval of the IC₅₀ for G12 was estimated to be between 7 and 11 μ M. Cytotoxicity testing of G12 at each concentration (red line) indicates that G12 is non-toxic. Each dot represents the mean value of three technical replicates. Source data for panels **c** and **d** are provided as a Source Data file.

Tables

Table 1 | Crystallographic data collection and refinement statistics from SARS-CoV-2 NCAP segments

Segment	¹⁹ GSQA55 ¹⁹	²⁰ AALALL ²⁰ Form 1	²⁰ AALALL ²⁰ Form 2	²⁴ GQTVTK ²⁴
Data collection				
Beamline	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E
Space group	P2 ₁ 2 ₁ 2 ₁	P1	P2 ₁ 2 ₁ 2	P2 ₁
Resolution (Å)	1.30 (1.39–1.30) ^a	1.12 (1.18–1.12)	1.30 (1.36–1.30)	1.10 (1.17–1.10)
Unit cell dimensions: a,b,c (Å)	4.77, 13.60, 42.44	9.45, 11.34, 20.27	44.46, 9.54, 10.95	19.57, 4.78, 22.03
Unit cell angles: α,β,γ (°)	90.0, 90.0, 90.0	74.9, 79.1, 67.8	90.0, 90.0, 90.0	90.0, 94.0, 90.0
Measured reflections	1833 (338)	5371 (323)	4666 (550)	4677 (344)
Unique reflections	809 (139)	2270 (136)	1234 (139)	1726 (170)
Overall completeness (%)	93.2 (95.9)	78.4 (31.1)	93.0 (84.8)	87.1 (50.9)
Overall redundancy	2.3 (2.4)	2.4 (2.4)	3.8 (4.0)	2.7 (2.0)
Overall R_{merge}	0.126 (1.04)	0.084 (0.397)	0.105 (0.808)	0.065 (0.446)
CC _{1/2}	99.7 (56.7)	98.5 (89.2)	99.7 (54.4)	99.5 (84.3)
Overall I/σ	3.5 (0.7)	5.9 (2.0)	5.9 (1.8)	6.0 (1.4)
Refinement				
$R_{\text{work}}/R_{\text{free}}$	0.259/0.253	0.158/0.197	0.217/0.248	0.133/0.177
RMSD bond length (Å)	0.015	0.009	0.010	0.009
RMSD angle (°)	1.4	1.3	1.6	1.5
Number of segment atoms	40	180 ^b	40	93 ^b
Number of water atoms	2	1	1	12
Number of other solvent atoms	0	21	14	0
Average B-factor of peptide (Å ²)	12.3	12.3	14.2	8.2
Average B-factor of water (Å ²)	19.9	12.8	26.6	24.7
Average B-factor other solvent (Å ²)	N/A	20.8	27.3	N/A
PDB ID code	7LV2 [https://doi.org/10.2210/pdb7LV2/pdb]	7LTU [https://doi.org/10.2210/pdb7LTU/pdb] (form 1)	7LUX [https://doi.org/10.2210/pdb7LUX/pdb] (form 2)	7LUZ [https://doi.org/10.2210/pdb7LUZ/pdb]

^aNumbers in parentheses report statistics in the highest resolution shell.

^bCount includes hydrogen atoms.

References

1. Savastano, A., Ibáñez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* 11, 1–10 (2020).
2. Chen, H. et al. Liquid–liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* 30, 1143–1145 (2020).
3. Cubuk, J. et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* 12, 1936 (2021).
4. Lu, S. et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* 12, 502 (2021).
5. Perdikari, T. M. et al. SARS-CoV-2 nucleocapsid protein phase separates with RNA and with human hnRNPs. *EMBO J.* 39, e106478 (2020).
6. Cascarina, S. M. & Ross, E. D. A proposed role for the SARS-CoV-2 nucleocapsid protein in the formation and regulation of biomolecular condensates. *FASEB J.* 34, 9832–9842 (2020).
7. Iserman, C. et al. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell* 80, 1078–1091.e6 (2020).
8. Carlson, C. R. et al. Phospho-regulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol. Cell* 80, 1092–1103.e4 (2020).

9. Ye, Q., West, A.M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 29, 1890–1901 (2020).
10. Bai, Z., Cao, Y., Liu, W. & Li, J. The SARS-CoV-2 nucleocapsid protein and its role in viral structure, biological functions, and a potential target for drug or vaccine mitigation. *Viruses* 13, 1115 (2021).
11. Jack, A. et al. SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol.* 19, e3001425 (2021).
12. Wu, Y. et al. RNA-induced liquid phase separation of SARS-CoV-2 nucleocapsid protein facilitates NF- κ B hyper-activation and inflammation. *Signal Transduct. Target. Ther.* 6, 1–13 (2021).
13. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767 (2012).
14. Romero, P. et al. Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* 42, 38–48 (2001).
15. Cascarina, S. M., Elder, M. R. & Ross, E. D. Atypical structural tendencies among low-complexity domains in the Protein Data Bank proteome. *PLoS Comput. Biol.* 16, e1007487 (2020).
16. Xiang, S. et al. The LC domain of hnRNPA2 adopts similar conformations in hydrogel polymers, liquid-like droplets, and nuclei. *Cell* 163, 829–839 (2015).

17. Hughes, M. P. et al. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* 359, 698–701 (2018).
18. Zhou, X. et al. Mutations linked to neurological disease enhance self-association of low-complexity protein sequences. *Science* 377, eabn5582 (2022).
19. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438 (2017).
20. Kato, M., Zhou, X. & McKnight, S. L. How do protein domains of low sequence complexity work? *RNA* 28, 3–15 (2022).
21. Roush, S. W. et al. Historical comparisons of morbidity and mortality for vaccine-preventable diseases in the United States. *J. Am. Med. Assoc.* 298, 2155–2163 (2007).
22. Wootton, J. C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* 266, 554–571 (1996).
23. Rosenberg, G. M. et al. Bioinformatic identification of previously unrecognized amyloidogenic proteins. *J. Biol. Chem.* 298, 101920 (2022).
24. Martin, E. W. & Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* 57, 2478–2487 (2018).
25. Tai, C. E. et al. The SARS-CoV-2 nucleocapsid protein preferentially binds long and structured RNAs. Preprint at bioRxiv <https://doi.org/10.1101/2021.12.25.474155> (2021).
26. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* 107, 3487–3492 (2010).

27. Sawaya, M. R. et al. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* 447, 453–457 (2007).
28. Hatos, A., Tosatto, S. C. E., Vendruscolo, M. & Fuxreiter, M. FuzDrop on AlphaFold: visualizing the sequence-dependent propensity of liquid-liquid phase separation and aggregation of proteins. *Nucleic Acids Res.* 50, W337–W344 (2022).
29. Gruca, A. et al. Common low complexity regions for SARS-CoV-2 and human proteomes as potential multidirectional risk factor in vaccine development. *BMC Bioinforma.* 22, 1–18 (2021).
30. Sievers, S. A. et al. Structure-based design of non-natural amino acid inhibitors of amyloid fibril formation. *Nature* 475, 96–100 (2011).
31. Seidler, P.M. et al. Structure-based inhibitors halt prion-like seeding by Alzheimer's disease- and tauopathy-derived brain tissue samples. *J. Biol. Chem.* 294, 16451–16464 (2019).
32. Saelices, L. et al. Amyloid seeding of transthyretin by ex vivo cardiac fibrils and its inhibition. *Proc. Natl Acad. Sci.* 115, E6741–E6750 (2018).
33. Griner, S. L. et al. Structure-based inhibitors of amyloid beta core suggest a common interface with tau. *Elife* 8, e46924 (2019).
34. Leaver-Fay, A. et al. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487, 545–574 (2011).
35. Luo, F. et al. Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nat. Struct. Mol. Biol.* 25, 341–346 (2018).

36. Murray, D. T. et al. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* 171, 615–627.e16 (2017).
37. Lu, J. et al. CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nat. Commun.* 11, 4090 (2020).
38. Guenther, E. L. et al. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat. Struct. Mol. Biol.* 25, 463–471 (2018).
39. Holubová, M., Štěpánek, P. & Hrubý, M. Polymer materials as promoters/ inhibitors of amyloid fibril formation. *Colloid Polym. Sci.* 299, 343–362 (2021).
40. Cao, Y., Adamcik, J., Diener, M., Kumita, J. R. & Mezzenga, R. Different folding states from the same protein sequence determine reversible vs. irreversible amyloid fate. *J. Am. Chem. Soc.* 143, 11473–11481 (2021).
41. Zhao, M. et al. GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase condensation of its nucleocapsid protein. *Nat. Commun.* 12, 1–14 (2021).
42. Julia, M. L. et al. Laura & the Center for Viral Systems Biology outbreak.info. <https://outbreak.info/> (2020).
43. Shu, Y. & McCauley, J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* 22, 30494 (2017).
44. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* 1, 33–46 (2017).

45. Syed, A. M. et al. Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles. *Science* 374, 1626–1632 (2021).
46. Ong, S. W. X. et al. Clinical and virological features of SARS-CoV-2 variants of concern: a retrospective cohort study comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta). *SSRN Electron. J.* 75, e1128–e1136 (2021).
47. Sheikh, A., McMenamin, J., Taylor, B. & Robertson, C. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *Lancet* 397, 2461–2462 (2021).
48. Dagpunar, J. Interim estimates of increased transmissibility, growth rate, and reproduction number of the Covid-19 B.1.617.2 variant of concern in the United Kingdom. Preprint at medRxiv <https://doi.org/10.1101/2021.06.03.21258293> (2021).
49. Fisman, D. N. & Tuite, A. R. Evaluation of the relative virulence of novel SARS-CoV-2 variants: a retrospective cohort study in Ontario, Canada. *CMAJ* 193, E1619–E1625 (2021).
50. Furukawa, Y. & Nukina, N. Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases. *Biochim. Biophys. Acta—Mol. Basis Dis.* 1832, 1271–1278 (2013).
51. Ramaswami, M., Taylor, J. P. & Parker, R. Altered ribostasis: RNA–protein granules in degenerative disorders. *Cell* 154, 727–736 (2013).
52. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163, 123–133 (2015).

53. Goedert, M., Eisenberg, D. S. & Crowther, R. A. Propagation of Tau aggregates and neurodegeneration. *Annu. Rev. Neurosci.* 40, 189–210 (2017).
54. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* 148, 1188–1203 (2012).
55. Haikal, C. et al. The bacterial amyloids phenol soluble modulins from *Staphylococcus aureus* catalyze alpha-synuclein aggregation. *Int. J. Mol. Sci.* 22, 11594 (2021).
56. Javed, I. et al. Accelerated amyloid beta pathogenesis by bacterial amyloid FapC. *Adv. Sci.* 7, 2001299 (2020).
57. Sampson, T. R. et al. A gut bacterial amyloid promotes α -synuclein aggregation and motor impairment in mice. *Elife* 9, e53111 (2020).
58. Friedland, R. P. & Chapman, M. R. The role of microbial amyloid in neurodegeneration. *PLoS Pathog.* 13, e1006654 (2017).
59. Ren, B. et al. Fundamentals of cross-seeding of amyloid proteins: an introduction. *J. Mater. Chem. B* 7, 7267–7282 (2019).
60. Semerdzhiev, S. A., Fakhree, M. A. A., Segers-Nolten, I. & Blum, C., Claessens, M. M. A. E. Interactions between SARS-CoV-2 N-protein and α -synuclein accelerate amyloid formation. *ACS Chem. Neurosci.* 13, 143–150 (2022).
61. Li, Y. et al. SARS-CoV-2 impairs the disassembly of stress granules and promotes ALS-associated amyloid aggregation. *Protein Cell* 13, 602–614 (2022).

62. Vidic, J. et al. Amyloid assemblies of influenza A virus PB1-F2 protein damage membrane and induce cytotoxicity. *J. Biol. Chem.* 291, 739–751 (2016).
63. Salladini, E. et al. Identification of a region in the common amino terminal domain of hendra virus P, V, and W proteins responsible for phase transition and amyloid formation. *Biomolecules* 11, 1324 (2021).
64. Shanmugam, N. et al. Herpes simplex virus encoded ICP6 protein forms functional amyloid assemblies with necroptosis-associated host proteins. *Biophys. Chem.* 269, 106524 (2021).
65. Pham, C. L. et al. Viral M45 and necroptosis-associated proteins form heteromeric amyloid assemblies. *EMBO Rep.* 20, e46518 (2019).
66. Steain, M. et al. Varicella zoster virus encodes a viral decoy RHIM to inhibit cell death. *PLoS Pathog.* 16, e1008473 (2020).
67. Zheng, X. et al. The C-terminal amyloidogenic peptide contributes to self-assembly of Avibirnavirus viral protease. *Sci. Rep.* 5, 14794 (2015).
68. Luckey, M. et al. A peptide from the adenovirus fiber shaft forms amyloid-type fibrils. *FEBS Lett.* 468, 23–27 (2000).
69. Papanikolopoulou, K. et al. Amyloid fibril formation from sequences of a natural β -structured fibrous protein, the adenovirus fiber. *J. Biol. Chem.* 280, 2481–2490 (2005).
70. Ghosh, A. et al. Self-assembly of a nine-residue amyloid-forming peptide fragment of SARS corona virus E-protein: mechanism of self aggregation and amyloid-inhibition of hIAPP. *Biochemistry* 54, 2249–2261 (2015).

71. Nyström, S. & Hammarström, P. Amyloidogenesis of SARS-CoV-2 spike protein. *J. Am. Chem. Soc.* 144, 8945–8950 (2022).
72. Charnley, M. et al. Neurotoxic amyloidogenic peptides in the proteome of SARS-COV2: potential implications for neurological symptoms in COVID-19. *Nat. Commun.* 13, 1–11 (2022).
2022 131.
73. Tetz, G. & Tetz, V. Prion-like domains in eukaryotic viruses. *Sci. Rep.* 8, 1–10 (2018).
74. Consortium, T. U. et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489 (2021).
75. Seidler, P. M. *Structural and Biochemical Studies of the hsp90 Chaperone Proteins* (State University of New York at Buffalo, 2014).
76. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* 2, 924–932 (2007).
77. Saff, E. B. & Kuijlaars, A. B. J. J. Distributing many points on a sphere. *Math. Intell.* 19, 5–11 (1997).
78. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946–950 (1993).
79. Winn, M. D. et al. Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 235–242 (2011).
80. Schrödinger, L. *The PyMOL Molecular Graphics System, Version 1.2r3pre.*

81. Schindelin, J. et al. Fiji: an open-source platform for biological image analysis. *Nat. Methods* 9, 676–682 (2012).
82. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-8, 679–698 (1986).
83. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675 (2012).
84. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 125–132 (2010).
85. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674 (2007).
86. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 479–485 (2010).
87. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 352–367 (2012).
88. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 355–367 (2011).
89. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 486–501 (2010).
90. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).

91. Garcia, G. et al. Antiviral drug screen identifies DNA-damage response inhibitor as potent blocker of SARS-CoV-2 replication. *Cell Rep.* 35, 108940 (2021).
92. Terwilliger, T. C. et al. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 64, 515–524 (2008).