Title

Structural and biochemical studies of patient-extracted amyloid fibrils

Author

Jiang, Yi Xiao

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structural and biochemical studies of patient-extracted amyloid fibrils

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Yi Xiao Jiang

2023

ABSTRACT OF THE DISSERTATION

Structural and biochemical studies of patient-extracted amyloid fibrils

by

Yi Xiao Jiang

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2023

Professor David S. Eisenberg, Chair

Human neurodegenerative diseases cause the gradual loss of neurons and progressive decline in memory, cognitive ability, movement and behavior. These devastating conditions afflict millions of patients worldwide and their families, however the molecular mechanisms of pathogenesis are not well understood, resulting in the lack of effective treatments and early diagnostic tools. Gross analyses of post-mortem brains using immunohistochemistry has identified abnormal deposits of amyloid proteins as defining features of neurodegenerative diseases. Amyloid proteins, like prions, can template aggregation of monomers into highly ordered, stable fibrils composed of tightly interdigitating β-sheets. Recent advances in cryogenic-electron microscopy (cryo-EM) have enabled the structural determination of amyloid fibrils to near-atomic resolution. These structures provide insight into the formation and propagation of amyloid fibrils, and can guide the design of therapies that prevent, delay, or reverse the aggregation of proteins in

neurodegenerative diseases. Moreover, previous work has shown that the structures of brain-extracted fibrils differ from the structures of fibrils assembled *in vitro*, underscoring the importance of studying patient-derived samples. In my thesis research, I extracted amyloid fibrils from autopsied brains of patients with frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP) and determined their structures using cryo-EM. Surprisingly, the fibrils examined were not formed by TDP-43 but by a fragment of transmembrane protein 106B (TMEM106B), revealing a previously unsuspected amyloid protein. In addition, I used a biosensor cell line, which detects proteopathic TDP-43 aggregates from sarkosyl-insoluble brains extracts, to study cohorts of TDP-43 proteinopathy cases including amyotrophic lateral sclerosis (ALS), FTLD-TDP and Alzheimer's disease (AD). My work expands the insight into the molecular basis of amyloid disorders, and inspires future research to address newly raised questions about TMEM106B and TDP-43 in neurodegenerative diseases.

The dissertation of Yi Xiao Jiang is approved.

Feng Guo

Lin Jiang

Jose A. Rodriguez

David S. Eisenberg, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter 1

ACKNOWLEDGEMENTS

I am eternally grateful to my mentor, David Eisenberg, who gave me an opportunity as a young scientist. David established the laboratory as a place for ambitious, rigorous and collaborative research. Within this environment, he gave me freedom to explore my scientific curiosities, which were often unsuccessful hypotheses. I appreciate the times when I show him confusing data or come to him with muddled ideas, and he can parse through the red herrings and point me toward the right direction. As a PI, David is exceedingly astute, always encouraging and fiercely supportive. After telling David that I was rejected for my first fellowship application, he wrote to me in response:

Do not be concerned. Life is long and you will receive lots of awards, and if you are like the rest of us, lots of turn downs. In “Middlemarch”, George Eliot says "We mortals, both men and women, devour many a disappointment between breakfast and supper." I mutter this to myself on all such occasions.

I want to thank my thesis committee members, Feng Guo, Lin Jiang and Jose Rodriguez for their scientific guidance, especially in the early days when my projects were not working. Feng is my Home Area Director who oversaw my entrance into the Biochemistry, Biophysics & Structural Biology program. He’s always made time to hear out my issues and offer advice, and to write the many letters of recommendation on my behalf. I also want to thank Anne Coleman, who started on my committee when I was pursuing the glaucoma-related project, and later withdrew from the committee when my research focus shifted.

I could not have asked for a better mentor than Michael Sawaya. He taught me all I know about structural biology, from crystallization to data processing to model building. He did so with

patience, often explaining the same thing multiple times for me. Michael is my role model for the scientist I want to be: passionate, detail-oriented and humble. He taught me importance of simple and clear science communication, and the joys of data visualization. As brilliant and accomplished Michael is as a scientist, he is an even better person and friend. I'm equally indebted to Duilio Cascio for his mentorship and training. He instilled a high standard for scientific quality and work ethic, but he also cared about students on a personal level. I estimate that my PhD would have taken double the time if not for Michael and Duilio's constant teaching, troubleshooting and encouragement in my five years here.

I am incredibly grateful for the scientific support and daily comradery of my colleagues in the laboratory. Qin Cao was my co-first author in our exciting discovery of a new amyloid in the brain. I learned so much from him in my first successful and high-profile project. Peng Ge and David Boyer were always present to answer cryo-EM questions, and provide timely refreshers on Bayesian statistics. Kevin Murray trained me to use Rosetta to perform *in silico* screening of structure-based peptide inhibitors of amyloid fibrils. My fellow graduate students, Cindy Cheng, Hope Pan, Jiahui Lu and Gregory Rosenberg, kept each other's morale high, mostly with mahjong and Chinese food. I benefitted from the wisdom of post-docs including Einav Tayeb-Fligelman, Romany Abskharon, Ke Hou, who also allowed me to contribute to their work. I had the fortune of working with diligent undergraduate students, Hillary Hernandez and Filipe Melo, who tremendously accelerated our research progress. I'm grateful for the staff who ensured everything ran smoothly from ordering chemical regents to maintaining computer clusters to navigating departmental logistics: Daniel Anderson, Melinda Balbernie, Alex Lisker, Michael Collazo, Genesis Falcon and Cindy Chau. It's an honor to pursue scientific research at this high level, but it's a privilege to work alongside genuine and good-hearted people.

In my thesis work, I had the pleasure of collaborating with renowned scientists, including Dennis Dickson and Michael DeTure at Mayo Clinic in Florida. They shared precious brain tissues, as well as their expertise as two of the most experienced neuropathologists in the world. Joseph Loo, Rachel Ogorzalek Loo and Janine Fu were crucial in quickly collecting mass spectrometry data for us to publish the TMEM106B paper. I greatly appreciate John Ravits, Anne Hiniker and Anita Hermann at UCSD for their wonderful collaboration and sharing brain tissues from the ALS and Alzheimer's Disease Research Centers.

Access to cryogenic-electron microscopy is a rare resource, so I want to recognize the facilities which have helped image our samples: Electron Imaging Center for Nanomachines at UCLA and the staff including Ivo Atanasov, Wong Hoi Hui, David Strugatsky; New York Structural Biology Center and the staff including Ed Eng, Charlie Dubbeldam, Jing Wang; Janelia Cryo-EM and the staff including Zhiheng Yu, Momoko Shiozaki, Rui Yan, Shixin Yang; Stanford-SLAC Cryo-EM Center and the staff including Lisa Dunn, Patrick Mitchell.

I want to acknowledge internal funding support from UCLA and Molecular Biology Institute, including the Whitcome Pre-Doctoral Fellowship, Audree V. Fowler Fellowships in Protein Science, and Dissertation Year Fellowship. I am truly honored to be a recipient of the Brain Diseases Award from the University of Zurich Foundation.

Graduate school is long and will inevitably present trying times. I am grateful to have friends and a support system along the journey: Yuki Hebner, Kaiser Atai, Kaleab Tessema, Justin Miller, Kat Skilton, Devin Gibbs, Sylwia Lipor, Barathan Jeevaretanam, Roger Castells-Graells, and Lenka Milojevic. Finally, I deepest gratitude goes to my parents, Hong Jiang and Xiuxuan (Nina) Huang, and my sister Ruth Jiang, for their unconditional support.

VITA

Education

McGill University, B.Sc. Biochemistry with First Class Honours 2018

Research Experience

Graduate Student Researcher, David S. Eisenberg lab, UCLA 2019–2023

Graduate Student Rotations, Tamir Gonen lab; Juli Feigon lab, UCLA 2018–2019

Undergraduate Honours Research, T. Martin Schmeing lab, McGill University 2016–2018

Research Assistant, William J. Muller lab, McGill University 2016

Awards and Honors

Molecular Biology Institute Dissertation Year Award, UCLA 2023

Dissertation Year Fellowship, UCLA 2023

Brain Diseases Award, University of Zurich Foundation 2023

Life Science Dean's Excellence Award, UCLA 2022

Travel Award, Biophysical Society 2022

Whitcome Pre-Doctoral Fellowship in Molecular Biology, UCLA 2022, 2021

Audree V. Fowler Fellowships in Protein Science, UCLA 2021

Keyfitz Scholar, McGill University 2015-2018

Undergraduate Student Research Award, NSERC 2017

NSERC Bursary Supplement, Fonds de Recherche du Québec Nature et Technologies 2017

Publications

Miller, J.E. †, Castells-Graells, R. †, Arbing, M.A. †, Munoz, A., **Jiang, Y.**, Espinoza, C.T., Nguyen, B., Moroz, P., Yeates, T.O. Design of beta-2 microglobulin adsorbent protein nanoparticles. *Biomolecules* 13, 1122 (2023). PMID: 37509158.

Wang, Y., He, Y., Wang, Y., Yang, Y., Singh, M., Eichhorn, C.D., Cheng, X., **Jiang, Y.**, Zhou, Z.H., Feigon, J. Structure of LARP7 protein p65–telomerase RNA complex in telomerase revealed by cryo-EM and NMR. *Journal of Molecular Biology* 435, 168044 (2023). PMID: 37330293.

Tayeb-Fligelman, E., Bowler, J.T., Tai, C.E., Sawaya, M.R., **Jiang, Y.**, Garcia, G., Griner, S., Cheng, X., Salwinski, L., Lutter, L., Seidler, P.M., Lu, J., Rosenberg, G.M., Hou, K., Abskharon, R., Pan, H., Zee, C., Boyer, D.R., Li, Y., Anderson, D.H., Murray, K.A., Falcon, G., Cascio, D., Saelices, L., Damoiseaux, R., Arumugaswami, V., Guo, F., Eisenberg, D.S. Low complexity domains of the nucleocapsid protein of SARS-CoV-2 form amyloid fibrils. *Nature Communications* 14, 2379 (2023). PMID: 37185252.

Jiang, Y. †, Cao, Q. †, Sawaya, M.R., Abskharon, R., Ge, P., DeTure, M., Dickson, D.W., Fu, J.Y., Ogorzalek Loo, R.R., Loo, J.A., Eisenberg, D.S. Amyloid fibrils in disease FTLD-TDP are composed of TMEM106B not TDP-43. *Nature* 605, 304–309 (2022). PMID: 35344984.

Tai, C.E., Tayeb-Fligelman, E., Griner, S., Salwinski, L., Bowler, J.T., Abskharon, R., Cheng, X., Seidler, P.M., **Jiang, Y.**, Eisenberg, D.S., Guo, F. The SARS-CoV-2 nucleocapsid protein preferentially binds long and structured RNAs. *BioRxiv* (2021). doi: https://doi.org/10.1101/2021.12.25.474155.

Rosenberg, G.M., Murray, K.A., Sawaya, M.R., **Jiang, Y.**, Eisenberg, D.S. Genetic and structural origins of amyloid diseases. (in preparation)

Hou, K., Ge, P., Sawaya, M.R., Dolinsky, D., **Jiang, Y.**, Yang, Y., Lutter, L., Boyer, D.R., Cheng, X., Pi, J., Zhang, J., Lu, J., Eisenberg, D.S. Fragmentation of Alzheimer's brain-derived tau fibrils by D-peptide fibrils of opposing twist. (in preparation)

OVERVIEW

Chapter 1 is the main work of my thesis, a reprint of a published manuscript that describes the discovery of TMEM106B amyloid fibrils in FTLD-TDP patients using cryo-EM. At the start of 2021, Qin Cao and I worked together to extract amyloid fibrils from FTLD-TDP brain tissues provided by Dennis Dickson and Michael DeTure from The Mayo Clinic in Florida. The fibrils samples were relatively sparse on the cryo-EM grids; Peng Ge played a crucial role during data collection to maximize the number of particles imaged. Qin taught me how to process and refine cryo-EM data for the helical reconstruction of amyloids. Together, we determined structures of three fibril polymorphs which share a common protofilament fold. We assumed these were TDP-43 fibrils and struggled to build in the sequence of TDP-43 into the map, until Michael Sawaya identified the fibril-forming protein by unbiased model building and a BLAST search. These fibrils were composed of a C-terminal fragment of a lysosomal transmembrane protein, TMEM106B, which none of us were aware of. We extracted and determined structures of the three identical fibril polymorphs from four FTLD-TDP patients, corresponding to subtypes A through D. We concluded that the amyloid protein in FTLD-TDP is TMEM106B, rather than TDP-43. Our article published in Nature was accompanied by two other reports of TMEM106B fibrils found in FTLD-TDP, but also in a variety of other neurodegenerative disease and even healthy cases. Goedert and Scheres from Cambridge proposed that TMEM106B fibrils are age-related. I will follow the neurodegenerative research community's progress with eager anticipation, as we seek to understand the mechanism and potential disease involvement of TMEM106B aggregation. I am grateful that my first first-authored project was such a surprising

and impactful discovery. I am the first to acknowledge that it is the result of good fortune, being at right place at the right time, and working with an exceptionally talented team of scientists.

Chapter 2 is a draft of a future manuscript that describes the use of a biosensor cell line to specifically and efficiently detect proteopathic TDP-43 aggregates in sarkosyl-insoluble extracts from neurodegenerative disease brains. The TDP-43 biosensors cells were kindly shared by Marc Diamond's group, however they have only been used in experiments with recombinant proteins. Dennis Dickson and Michael DeTure from The Mayo Clinic in Florida shared brain tissues from 40 cases of Alzheimer's disease with TDP-43 pathology, in addition to the 40 FTLD-TDP cases used in the study described in chapter 1. John Ravits and Annie Hiniker from UCSD, with the help of Anita Hermann, shared brain tissues from over 30 ALS cases. I saw these large cohorts of TDP-43 proteinopathy cases as optimal candidates for a study employing the TDP-43 biosensor cells. Hillary Hernandez, an undergraduate student, extracted sarkosyl-insoluble fractions from nearly 100 brain samples in the summer of 2023. Xinyi (Cindy) Cheng performed the TDP-43 biosensor seeding experiments. We demonstrated that fluorescent punta can only be induced in these biosensor cells by transfection of TDP-43 fibrils, not amorphous aggregates, monomers, or other amyloids. Detection of proteopathic TDP-43 is correlated with the burden of phosphorylated C-terminal fragment of TDP-43. We also noticed the correlation of seed-competency of cases to FTLD-TDP subtypes and HpScl. The cases identified to contain abundant proteopathic TDP-43, or an enrichment of fibrils, are ideal samples for subsequent structural and biochemical studies.

Chapter 3 is a reprint of a published manuscript that describes the SARS-CoV-2 nucleocapsid protein's ability to form amyloid fibrils, and our structural approach to design inhibitors against amyloid-facilitated phase separation that demonstrated antiviral activity in SARS-CoV-2

infected cells. The project began during the COVID-19 pandemic on a Zoom call. Gregory Rosenberg computationally analyzed the SARS-CoV-2 genome and found the nucleocapsid protein contains low-complexity domains which may be aggregation-prone. Einav Tayeb-Fligelman led the project to biochemically characterize the amyloid aggregation of nucleocapsid proteins, which may be involved in phase separation during the viral replication cycle. Michael Sawaya and a small team determined structures of three aggregation-prone segments. I used those crystal structures as templates to design capping inhibitors aimed at obstructing potential fibril elongation. Jeanette Bowler screened my top *in silico* candidates *in vitro*, and found that D-peptide G12 interferes with self-assembly of nucleocapsid proteins and viral replication of SARS-CoV-2 in cells. This work presents proof-of-concept for an alternative approach to drug development for viral diseases.

Chapter 4, the last chapter, describes my first project in the Eisenberg lab, in which I studied the potential aggregation of myocilin, a protein linked to glaucoma. Disease-associated mutations in myocilin cause its intracellular accumulation in trabecular meshwork cells in the eye, which leads to increased intraocular pressure, the main causative factor of glaucoma. Raquel Lieberman's group at the Georgia Institute of Technology had published several articles showing that recombinant myocilin proteins are destabilized by disease mutations, misfold and convert into amyloids. This prompted me to investigate the molecular basis of myocilin aggregation. I used ZipperDB to predict aggregation-driving segments in the C-terminal olfactomedin domain, confirmed six out of nine segments form amyloid fibrils *in vitro*, and determined crystal structures of two segments as steric zippers. These structures offer plausible models for how these segments may contribute energetic stability in an amyloid fibril. However, this work was not pursued further or published due to insufficient evidence of amyloid involvement in

glaucoma. Dr. Anne Coleman at UCLA Jules Stein Eye Institute provided ocular tissues extracted from her trabeculectomy surgeries (which makes a hole in the trabecular meshwork to alleviate pressure). I examined these glaucoma patient tissues by histology and Congo Red staining, but there was not conclusive evidence to indicate amyloid (data not shown). Nevertheless, I am grateful for the opportunity this project gave me to learn peptide crystallography.

CHAPTER 1

Discovery of TMEM106B amyloid fibrils in frontotemporal lobar degeneration with TDP-43 inclusions

Abstract

Frontotemporal lobar degeneration (FTLD) is the third most common neurodegenerative condition after Alzheimer's and Parkinson's diseases[1]. FTLD typically presents in 45 to 64 year olds with behavioural changes or progressive decline of language skills[2]. The subtype FTLD-TDP is characterized by certain clinical symptoms and pathological neuronal inclusions with TAR DNA-binding protein (TDP-43) immunoreactivity[3]. Here we extracted amyloid fibrils from brains of four patients representing four of the five FTLD-TDP subclasses, and determined their structures by cryo-electron microscopy. Unexpectedly, all amyloid fibrils examined were composed of a 135-residue carboxy-terminal fragment of transmembrane protein 106B (TMEM106B), a lysosomal membrane protein previously implicated as a genetic risk factor for FTLD-TDP[4]. In addition to TMEM106B fibrils, we detected abundant non-fibrillar aggregated TDP-43 by immunogold labelling. Our observations confirm that FTLD-TDP is associated with amyloid fibrils, and that the fibrils are formed by TMEM106B rather than TDP-43.

Introduction

Pathological deposits of amyloid proteins are associated with more than 50 systemic and neurodegenerative diseases[5,6], and include amyloid-β and tau in Alzheimer's disease, α-synuclein in Parkinson's disease, and amylin (also known as hIAPP) in type II diabetes[7]. Recent advances in cryo-electron microscopy (cryo-EM) techniques have enabled near-atomic resolution of structures of amyloid fibrils extracted from patients with these diseases[8,9,10,11,12,13]. Structural information from ex vivo fibrils offers insights into the molecular signatures of disease and can guide the design of therapeutic interventions intended to prevent, delay or reverse the aggregation of proteins in amyloid disorders.

FTLD causes presenile dementia in around 81 out of every 100,000 people between the ages of 45 and 64[14]. FTLD presents clinically as disorders of social behaviour and language ability[2]. The major subtype of FTLD is characterized by neuronal inclusions containing TDP-43, termed FTLD-TDP, which accounts for around 50% of all FTLD cases[15]. FTLD-TDP is further classified into types A to E according to morphology and neuroanatomical distribution of TDP-43 inclusions[16,17]: each type is associated with diverse clinical symptoms[18]. Although the defining neuropathological feature of FTLD-TDP is immunoreactivity for deposits of ubiquitinated, hyperphosphorylated TDP-43[3,19], amyloid fibrils in FTLD-TDP brains have not been studied structurally. Here we extracted amyloid fibrils from four donors diagnosed with FTLD-TDP types A, B, C and D, and determined 12 near-atomic resolution structures by cryo-EM. Using atomic model building, mass spectrometry and western blot, we confirmed that these fibrils are formed by TMEM106B, a protein previously identified as a genetic risk factor for FTLD-TDP[4]. Our study reveals the amyloidogenic nature of TMEM106B in FTLD-TDP, and focuses attention on a protein not previously associated with amyloid conditions[20,21,22].

Results

Fibril extraction from human brains

The Mayo Clinic Brain Bank contributed frozen brain tissues of patients with FTLD-TDP (40 donors) and age-matched, non-FTLD-TDP controls (8 donors) (Extended Data Fig. 1a and Extended Data Table 1). We inspected the detergent-insoluble fractions of these samples by negative stain transmission election microscopy. In most donors with FTLD-TDP (38 out of 40 donors), we observed amyloid fibrils as well as non-fibrillar aggregates (donors 1–4 in Fig. 1aand 34 other donors in Extended Data Fig. 2). By contrast, no fibrils were observed in the

eight non-FTLD-TDP donors (donor 5 in Fig. 1a and 7 other donors in Extended Data Fig. 2). We then selected four donors neuropathologically confirmed to be FTLD-TDP types A, B, C and D (donors 1–4, Table 1, Extended Data Fig. 3) for further study. We collected cryo-EM data on samples from each of the four donors and determined amyloid fibril structures (Fig. 2, Table 2, Extended Data Figs. 4–6 and Extended Data Table 2). Three fibril polymorphs—PM1, PM2 and PM3—were identified in each cryo-EM dataset (Extended Data Fig. 4a–c), with similar distributions in all four FTLD-TDP donors (Fig. 2 and Extended Data Fig. 4d).

Identification of the fibril-forming protein

We used our highest resolution cryo-EM map (2.9 Å, from donor 1, PM1) to begin model building. We initially presumed that the fibrils were composed of TDP-43; however, extensive incompatibility between the side chain density in the map and the sequence of TDP-43 invalidated this presumption. To identify the protein in the fibrils, we constructed query sequences by selecting amino acids that best fit the PM1 density without regard to their similarity to known protein sequences. We built two models with opposite directionality, thereby obtaining two query sequences (Extended Data Fig. 7 and Methods). A search for human proteins similar to these two queries revealed only one hit: residues 121–254 of TMEM106B, a lysosomal transmembrane protein.

Several observations lend confidence that TMEM106B is indeed the protein building block of these fibrils: (1) the sequence of TMEM106B fits the map well, leaving no unexplained density (Fig. 2); (2) density protruding from Asn145, Asn151, Asn164 and Asn183 (Extended Data Fig. 8a, b) accounts for the known *N*-linked glycosylation sites of TMEM106B23 (Fig. 3a); (3) the only density that connects two side chains is accounted for by a disulfide bond between Cys214 and Cys253 (Fig. 2); and (4) a previous study revealed that genetic variants of

TMEM106B are associated with FTLD-TDP4. No other protein could explain these features so completely. The models for PM2 and PM3 of donor 1 were built using rigid body fits of the PM1 model into the density maps, with minor adjustments and refinement applied subsequently. Cryo-EM maps of all three polymorphs make it clear that the fibrils are composed of a proteolytic fragment rather than the full-length TMEM106B. The amino terminus of the core, Ser120, is buried by residues Tyr158, Val160, His239 and Glu241, leaving no space for residues preceding Ser120 (Figs. 2, 3b). Consequently, proteolytic cleavage between residues Arg119 and Ser120 must precede fibril formation (Fig. 3a, Discussion). By contrast, the C terminus of the core, Gly254, is solvent-exposed, allowing the possibility that the remaining C-terminal 20 residues are attached but disordered. Notably, the range of the ordered core (residues 120–254) coincides closely with the definition of the lumenal domain of TMEM106B (Fig. 3a), which is known to be released from the lysosomal membrane by proteolytic cleavage at an undetermined location in the vicinity of Ser12024.

Stable fold of TMEM106B fibrils

Protofilaments of all three polymorphs from donor 1 exhibit the same fold (Fig. 2, bottom left) despite their different types of bundling. A single protofilament constitutes PM1, and a pair of protofilaments constitute PM2 and PM3 and intertwine with distinct symmetries: C_2 in PM2 and pseudo-2_1 in PM3. Moreover, the protofilament interfaces of PM2 and PM3 differ. In PM2, side chains of Arg180 point towards residual density at the centre of the interface, whereas PM3 features a hydrophobic interaction between pairs of Met207 and Tyr209 residues (Fig. 2 and Extended Data Fig. 9). The only notable structural differences between the folds in the three polymorphs are confined to the PM3 protofilament interface—the rotamers of Glu206, Met207,

Tyr209 and Tyr211 in PM3 vary from those in PM1 and PM2 (Fig. 2 and Extended Data Fig. 9c).

The conserved fold of TMEM106B contains 18 β-strands (Fig. 3b), varying in length (from 3 to 15 residues) and curvature. Notably, 18 β-strands exceeds the number typically associated with amyloid fibrils[25]; this abundance of β-strands contributes to an unusually large and stable fibril core. Indeed, these 18 β-strands mate together in pairs, forming multiple steric zippers that strongly stabilize the fibril (Supplementary Fig. 1). Compared with the average of the 63 known amyloid fibril structures (in 2021), the 'golf course' fold of TMEM106B is nearly three times more stable according to solvation energy estimates[25] (−62 versus −22 kcal mol^{-1} per chain). These crude estimates of energetic stability may explain the fibrils' resistance to 2% sarkosyl and 1% sodium dodecyl sulfate (SDS) during extraction. This stability suggests that formation of TMEM106B fibrils in the brain is irreversible, which is consistent with their possible pathogenic roles in FTLD-TDP.

Clues to the origin of the TMEM106B fibrils are given by our observation that four uncomplemented acid side chains face inward in the fibril core structure: Glu163, Glu205, Asp212 and Glu246 (Fig. 3b). Such uncomplemented, buried negative charges are rare among pathogenic amyloids, as electrostatic repulsion between closely spaced molecules with the identical charge would weaken the fibril. However, given that TMEM106B is a lysosomal protein and this is a lumenal domain, the pH of the environment might be low enough to protonate and neutralize some inward-facing acids. Neutralization of these negative charges may help to overcome the barrier to fibril nucleation.

FTLD-TDP donors share a conserved fold

Our study of four patients with FTLD-TDP contributes to mounting evidence that protein conformation is conserved among patients diagnosed with the same amyloid disease. In fact, it has been proposed that protein conformation could form the basis for classifying disease[26]. Accordingly, we found that the golf course fold of TMEM106B is conserved among all four FTLD-TDP donors and constitutes the basis of all three polymorphs. However, we do observe a small but significant shift in the protofilament–protofilament interface of PM2 among donors. The interface from donors 1 and 4 (types C and D) is centred on Lys178, whereas the interfaces from donors 2 and 3 (types B and A) are centred on Arg180 (Extended Data Fig. 6 and Extended Data Fig. 9a, b). It is unclear whether these structural differences are linked to the different FTLD-TDP subtypes or other patient attributes, such as age. PM2 accounts for around 10% of TMEM106B fibrils in all four donors

Genotyping suggests that donor 3 contains wild-type *TMEM106B*, whereas donors 1, 2 and 4 harbour heterozygous Thr185Ser variants in their *TMEM106B* gene (Table 1). Mass spectrometry identified a peptide with the Thr185Ser variant in donor 1 (Extended Data Fig. 10 and Extended Data Table 3). The Thr185Ser variant was detected in a genome-wide association study of patients with FTLD-TDP[4], and a follow-up study suggested that the Thr185Ser variant protects against FTLD-TDP[27]. The Ser185 isoform is associated with lower protein expression and is more rapidly degraded than the wild-type Thr185 isoform, which may contribute to its protective mechanism[28]. The observation that all four donors possess fibrils with the same fold suggests that this fold is compatible with wild-type TMEM106B protein; however, we cannot confirm whether the heterozygous donor's fibrils contain only wild-type TMEM106B, only variant TMEM106B, or a mixture of both. Our maps confirm that residue 185 is within the protofilament core (Fig. 2a and Extended Data Fig. 8c), but even the highest-

resolution cryo-EM map (donor 1 PM1) is insufficient to distinguish among Thr, Ser or an average of both sidechains (Extended Data Fig. 8c). Since the Thr/Ser185 side chain faces the solvent with no contributing interactions in the fibril structure, we believe that the fibril fold observed here can accommodate both wild-type and Thr185Ser TMEM106B proteins.

TMEM106B and TDP-43 in patient extracts

To verify that the protein in FTLD-TDP fibrils is indeed TMEM106B, we performed western blotting of patient extracts with a C-terminal TMEM106B antibody. We observed a TMEM106B-positive band at around 35 kDa (along with other high molecular weight bands; Methods) in sarkosyl-insoluble fractions from most of the FTLD-TDP donors that we examined, but not in extracts from the non-FTLD-TDP donors (Fig. 1b and Extended Data Fig. 1b), consistent with our observations by electron microscopy that TMEM106B fibrils are present only in patients with FTLD-TDP (Fig. 1a and Extended Data Fig. 2). These results support the hypothesis that the aggregation of TMEM106B is associated with FTLD-TDP. The 35-kDa band could be the fibril-forming cleavage product (TMEM106B 120–274), a 155-residue peptide with five glycans. To further confirm the presence of TMEM106B, we performed mass spectrometry on gel slices around the 35 kDa region and detected peptide fragments from only the lumenal domain (Extended Data Fig. 10 and Extended Data Table 3), consistent with our cryo-EM maps.

We attempted immunogold labelling of patient-extracted fibrils using three different TMEM106B antibodies (Methods), but none of them labelled the fibrils (Fig. 1c, left). This was anticipated, since the epitopes of the antibodies are either cleaved before fibril formation or buried inside the fibril core. It may also explain why TMEM106B aggregation has not been detected by immunohistochemical staining of patient brains. Future work is needed to develop

antibodies that probe TMEM106B fibrils, either by binding to the fibril core, or by recognizing residues 255–274 of TMEM106B, which are not included in the fibril core and may form an accessible fuzzy coat.

To verify the presence of TDP-43 aggregates, we also probed patient extracts with TDP-43 antibodies. Immunoblotting revealed that the FTLD-TDP-associated, phosphorylated form of TDP-43[29] is present with TMEM106B in the sarkosyl-insoluble fractions of FTLD-TDP donors but not the non-FTLD-TDP control (Fig. 1b). These results are consistent with a previous study[30] and suggest that TDP-43 forms aggregates in patients with FTLD-TDP. Immunogold labelling of sarkosyl-insoluble fractions with the TDP-43 antibody revealed that non-fibrillar aggregates were present and identified them as TDP-43 (Fig. 1c, right). Fibrils were also detected but were not labelled; these were presumably composed of TMEM106B.

Discussion

A decade ago, TMEM106B was identified as a genetic risk factor for FTLD-TDP in a genome-wide association study; ever since, its role in the disease has remained unknown. Our finding of TMEM106B amyloid fibrils in ex vivo patient brains gives us a firm, new grasp on this association. Indeed, TMEM106B amyloid appears to be a signature of FTLD-TDP as evidenced by its presence in four different FTLD-TDP types, both with and without the Thr185Ser variant, and in both familial and sporadic cases (Table 1). Each of the three polymorphic structures we observed from each patient reveals TMEM106B folded in the same golf-course-like fold. This discovery directs new focus on TMEM106B and the role that its aggregation may have on the aggregation of TDP-43, previously recognized as “the major disease protein”[31]. It remains unknown whether TMEM106B fibrils are pathogenic and, if so, whether their pathogenicity

arises from a gain or loss-of-function. Alternatively, they may be benign by-products downstream of an undiscovered, primary pathological pathway. Understanding these possible pathogenic roles of TMEM106B fibrils will require studies from a larger cohort.

Our structures reveal that proteolytic cleavage happens prior to, and is essential for, formation of the core of TMEM106B fibrils (Figs. 2, 3a)—this suggests that protease inhibition might offer a strategy for FTLD-TDP intervention. TMEM106B undergoes lumenal domain shedding, in which the amyloid core-forming C-terminal fragment is released within the lumen of the lysosome by a resident protease—the cleavage site and the identity of the protease is still unknown[24]. Our structural evidence that cleavage occurs between Arg119 and Ser120 (Supplementary Fig. 2) is consistent with the propensity of many known proteases to cleave between arginine and serine residues. The specificity of the TMEM106B cleavage site required to enable fibril growth would seem to be even greater than that of amyloid precursor protein[32], another single-pass transmembrane protein whose proteolytic cleavage by secretases is involved in the growth of amyloid fibrils in Alzheimer's disease. In the case of β-amyloid, cleavage at various sites leads to amyloid formation, whereas in TMEM106B, burial of its cleaved terminus in the fibril core constrains amyloid-competent fragments to those cleaved between Arg119 and Ser120.

Evidence of a ligand-binding site in the TMEM106B PM2 polymorph identifies another potential factor affecting fibril formation. The interface between PM2 protofilaments features positively charged residues (Lys178 or Arg180) that point towards an unmodelled residual density (Extended Data Fig. 9b). This positively charged environment suggests that the ligand bound here is negatively charged. Our observation that this ligand accommodates two slightly different protofilament arrangements of PM2 (Extended Data Fig. 9a) is consistent with the small size of

the interface and the dominance of electrostatic forces in its stabilization. Of note, residual densities are also found in similar positions in PM1 and PM3 (Extended Data Fig. 8a), and may correspond to the same ligand. However, the resolution of the residual densities does not allow us to identify the ligand. We cannot rule out the possibility that the density corresponds to sarkosyl—which is negatively charged at pH 7.5—that may have been incorporated during fibril extraction. If fibril formation is sensitive to the presence of this ligand, then ligand withdrawal may inhibit their formation. Ligand-binding sites have similarly been discovered in pathogenic fibrils of tau13 and α-synuclein12.

Our experiments show that aggregates of TDP-43 in our samples of FTLD-TDP are amorphous rather than amyloid-like. FTLD-TDP has been defined by inclusions of TDP-4315, and TDP-43-immunolabelled fibrils have been isolated from patients with FTLD-TDP33. However, sarkosyl-insoluble TDP-43 extracted from patients with FTLD-TDP exhibits non-fibrillar rather than fibrillary morphology, shown both here (Fig. 1) and in a previous study30. Our cryo-EM structures suggest that the amyloid fibrils that we imaged contain only TMEM106B; however, we cannot exclude the possibility that TDP-43 forms fibrils in patients with FTLD-TDP, and that our extraction method did not capture TDP-43 fibrils that were present. We believe that this is unlikely because we observed no fibrils under electron microscopy in the other fractions throughout our fibril-extraction protocol, and the sarkosyl and SDS treatment should not dissolve pathogenic irreversible fibrils. TDP-43 fibrils have recently been characterized in ALS with FTLD-TDP type B34, a related neurodegenerative disease. It is uncertain whether the presence of TDP-43 fibrils in those cases stems from the different neuropathological condition or from different extraction procedures. In our western blot and immunogold-labelling experiments, we found that TDP-43 aggregates can be co-extracted with TMEM106B fibrils from patients with

FTLD-TDP. It is unclear whether TDP-43 aggregates and TMEM106B fibrils co-localize in the patients' brains, and whether a cross-seeding mechanism is involved in FTLD-TDP pathogenesis.

Conclusion

Our structures of patient-derived TMEM106B fibrils reveal protein cleavage followed by amyloid formation of the C-terminal fragment of TMEM106B. This finding may refocus pathogenic studies of FTLD-TDP and perhaps other neurodegenerative diseases to include TMEM106B.

Methods

Post-mortem human brain samples

The Brain Bank for Neurodegenerative Disorders at Mayo Clinic Florida provided post-mortem brain tissues from patients with neuropathologically confirmed FTLD-TDP. Information on human donors is provided in Table 1. Autopsies were performed after consent by the next of kin or someone with legal authority to grant permission. The brain bank operates under protocols approved by the Mayo Clinic Institutional Review Board.

Immunohistochemistry staining

Post-mortem brains were immersion-fixed in 10% formalin, and sections of brain were embedded in paraffin, cut on a microtome at 5 μm thickness and mounted on positively charged glass slides. Sections were dried overnight and used for immunohistochemistry staining. Paraffin-embedded brain sections were deparaffinized in xylene, and rehydrated through a series of ethanol solutions, followed by washing in deionized H_2O. Antigen retrieval was performed by

steaming slides in deionized H_2O or Tris-EDTA (DAKO), pH 9.0 for 30 min followed by a 5-min incubation in DAKO Peroxidase Block (DAKO, catalogue (cat.) no. S2001) to block endogenous peroxidase activity. Slides were blocked with DAKO Protein Block Serum-Free (DAKO, cat. no. X0909) for 1 h, and incubated with Anti TAR DNA-Binding Protein 43 (TDP-43) phospho-Ser409/410 monoclonal antibody (Cosmo Bio USA, cat. no. CAC-TIP-PTD-M01, lot no. 11-9-20) diluted 1:1,000 for 45 min. After washing, sections were incubated for 30 min in DAKO Envision-Plus System HRP Labelled Polymer Anti-Mouse (DAKO, cat. no. K4001). Peroxidase labelling was visualized with the Liquid DAB+ Substrate Chromogen System (DAKO, cat. no. K3468). Glass-mounted tissue sections were viewed with an Olympus BX51 and microscope images were captured with Olympus DP73 digital camera.

Fibril extraction from FTLD-TDP patient brains

Frozen brain tissues were weighed, diced into small pieces, and resuspended in 10 ml g^{-1}homogenization-solubilization (HS) buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol) supplemented with 1:100 (v/v) Halt protease inhibitor (Thermo Scientific). Resuspended tissue was homogenized using a Polytron homogenizer (Thomas Scientific) for 45 s and mixed with 10 ml g^{-1} of 4% (w/v) *N*-lauroyl-sarcosine (sarkosyl, Sigma), 2 U μl^{-1} home-made benzonase and 4 mM $MgCl_2$. Mixed solution was incubated at 37 °C with constant shaking at 300 r.p.m. for 45 min. Sample solution was then mixed with 10 ml g^{-1} ice-cold HS buffer with 0.5% (w/v) sarkosyl, and centrifuged at 3,000*g* at 4 °C for 5 min. The supernatant was extracted and centrifuged at 21,000*g* for 30 min. The supernatant was discarded, and the pellet was resuspended in 1.5 ml g^{-1} Tris buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl) and incubated at 4 °C overnight. After incubation, the solution was centrifuged at 6,000*g* for 5 min. The supernatant was transferred into a new tube, mixed with 1 %

(w/v) SDS (Invitrogen), and gently rotated at room temperature for 15 min. The solution was then sonicated for 3 min and centrifuged at 6,000*g* for 10 min. The supernatant was centrifuged at 21,000*g* for 30 min, and the pellet was resuspended in 10 µl g^{-1} Tris buffer, sonicated for 3 min and used for EM. Approximately 1 g of brain tissue from each patient was used for cryo-EM study.

Fibrils were also extracted using ice-cold water following a previously described protocol35. In brief, brain tissues were diced and resuspended in 4 ml g^{-1} of Tris-calcium buffer (20 mM Tris, pH 8.0, 138 mM NaCl, 2 mM $CaCl_2$, 0.1% NaN_3), then centrifuged at 3,100*g* at 4 °C for 5 min. The supernatant was collected and the Tris-calcium buffer wash was repeated four more times. After the fifth wash, the pellet was resuspended in Tris-calcium buffer with collagenase and incubated at 37 °C with constant shaking at 500 r.p.m. overnight, then centrifuged at 3,100*g* at 4 °C for 30 min. The pellet was resuspended in 4 ml g^{-1} Tris-EDTA buffer (20 mM Tris, pH 8.0, 140 mM NaCl, 10 mM EDTA, 0.1% NaN_3), then centrifuged at at 3,100*g* at 4 °C for 5 min. The supernatant was collected and the Tris-EDTA buffer wash was repeated nine more times. After the tenth wash, the pellet was resuspended in 2 ml g^{-1} ice-cold water, then centrifuged at 3,100*g* at 4 °C for 5 min. The fibril-containing supernatant was collected and the ice-cold water extraction was repeated nine more times. Fibrils extracted using sarkosyl and ice-cold water exhibited similar morphologies when examined by negative-stain electron microscopy (Fig. 1d), suggesting that sarkosyl does not induce the fibril formation during the extraction process. Fibril samples extracted using sarkosyl were more abundant and contained less contaminants, thus were used for cryo-EM study.

Negative-stain transmission electron microscopy

Carbon-coated formvar support films (400-mesh) mounted on copper grids (Ted Pella) were glow-discharged for 30 s before sample preparation. Three microlitres of sample solution was applied to the grids and incubated for 3 min, then excess sample solution was blotted off using filter paper. Grids were stained with 3 µl of 2% uranyl acetate (Electron Microscopy Sciences) for 1 min, washed with an additional 3 µl of 2% uranyl acetate and air-dried for 2 min. The grids were imaged using a Tecnai T12 transmission electron microscope (FEI).

Cryo-EM data collection and processing

To prepare the cryo-EM grids, we applied 2.6 µl of sample solution onto Quantifoil 1.2/1.3 200 mesh electron microscope grids glow-discharged for 2 min before use. Grids were plunge frozen into liquid ethane using a Vitrobot Mark IV (FEI). Cryo-EM data were collected on a Titan Krios transmission electron microscope (FEI) equipped with a K3 Direct Detection Camera (Gatan), operated with 300 kV acceleration voltage and energy filter width of 20 eV. Super-resolution movies were collected with a nominal physical pixel size of 1.1 Å per pixel (0.55 Å per pixel in super-resolution movie frames) and a dose per frame of ~1 e^{-} $Å^{-2}$. A total of 36 frames with a frame rate of 12 Hz were taken for each movie, resulting in a final dose of ~36 e^{-} $Å^{-2}$ per image. Manual data collection was performed using Leginon software package[36]. Cryo-EM datasets of four FTLD-TDP donors were collected and processed separately.

CTF estimation was performed using CTFFIND4[37]. For the dataset of donor 1, particle picking was performed using CrYOLO[38], trained with ~150 micrographs that we picked manually using e2helixboxer.py from EMAN2[39]. Particle extraction, 2D classification, helical reconstruction, and 3D refinement were performed in RELION[40,41]. Particles were extracted using an inter-box distance of 102.4 Å and a box size of 1,024 pixels scaled down to a 432-pixel box size. Two-dimensional classification with tau_fudge 2 was performed with all particles, and we found that

all identifiable classes could be grouped into one of three fibril polymorphs, which we named PM1 (78.1% of all particles), PM2 (17.8% of all particles) and PM3 (4.1% of all particles). No other fibril morphologies were identified in this dataset. Particles from each polymorph were selected and used for 3D classification with $K = 3$ (for PM1) or $K = 1$ (for PM1 and PM2), using a Gaussian cylinder as the initial model. The best 3D classes were used as the initial model for subsequent 3D classifications with smaller box size particles. We re-extracted particles from all micrographs using an inter-box distance of 32 Å and a box size of 686 pixels scaled down to a 320 pixels box size. We manually selected 686-pixel particles for each polymorph based on the 2D classification of 1,024-pixel particles. For example, if a 1,024-pixel particle was classified as PM1 in 2D classification, then all 686-pixel particles extracted from that particular filament were selected for the PM1 polymorph. These manually selected particles were used for 2D classification for each polymorph separately, and the best 2D classes were selected for 3D classification. Two rounds (for PM1) or one round (for PM2 and PM3) of $K = 3$ 3D classification was performed. Particles from the best 3D class for each polymorph were re-extracted using an inter-box distance of 32 Å and a box size of 320 pixels (no scaling) and used for high-resolution 3D refinement. To improve resolution, additional 2D classifications were performed for PM1 and PM2, with tau_fudge starting at 2 and increased incrementally to 8 in the final iterations. The 2D classes with clear 4.8 Å separation were selected for further 3D refinement. High-resolution gold-standard refinement was performed for each polymorph and the initial near-atomic resolution maps with refined helical parameters were generated. CTF refinement and Bayesian polishing was performed, and the final reconstructions were generated by one or two rounds of additional golden-standard refinement. The resolution of each reconstruction was estimated

using the 0.143 Fourier shell correlation (FSC) resolution cut-off. See Table 2 for data collection and processing statistics for donor 1.

For the datasets of donor 2–4, similar strategies were applied for data collection and processing. All three datasets were processed independently up to and throughout 2D classification, with no information from other dataset introduced. 2D classification of all particles suggested that these three datasets also contain PM1, PM2, and PM3, with similar morphology and distribution as donor 1 (Extended Data Fig. 4), and no other identifiable fibril morphologies. During initial 3D reconstruction, we processed the donor 2 dataset independently, whereas for donors 3 and 4 we used the maps generated from the donor 1 dataset as initial model to expedite data processing. We believe these reconstructions were not biased because we used a 30 Å low-pass-filter on the initial model so that any higher-resolution information beyond 30 Å would originate from the dataset and not the initial model, and the final map of each polymorph was near-atomic (3.5–5.3 Å). After initial 3D reconstruction, a similar strategy was applied for each dataset to generate the final maps. See Extended Data Table 2 for data collection and processing statistics for donors 2–4.

Atomic model building

Our first model building efforts focused on PM1 of donor 1. We attempted to model the sequence of TDP-43 in the map using two methods: manual building with Coot[42] and automatic building using phenix.sequence_from_map[43]. No satisfactory model could be made with the TDP-43 sequence; multiple sidechains were inconsistent with densities even taking into consideration potential post-translational modifications such as phosphorylation of serine residues (the best model is shown in Extended Data Fig. 7a).

We then applied an unbiased strategy for model building: we built two polyalanine backbones with opposite directionality of N- and C-termini. For each model, we mutated the residues to amino acids that best fit the side chain densities (Extended Data Fig. 7a). The sequences of two resulting models were used as queries in BLAST to identify the protein in our maps. We restricted the BLAST database to human sequences (NCBI:txid9606), knowing that the brain samples that produced the fibrils were human. No other parameters of the search were altered from the default values. TMEM106B stood out as the only significant match to our query sequence.

We obtained five hits in total, but all five hits were sequences of the same protein, TMEM106B (including the protein labelled as hypothetical protein FLJ11273) (Extended Data Table 4). The percent identity of the hits was low (~16%), but the coverage was high, covering 125 of the 134 residues in the query (93%) without gaps or insertions in the alignment. All of these TMEM106B hits score above the default expect threshold value of 10. The 'expect' value that we obtained for the highest scoring hit, TMEM106B, was 0.003. This value means the number of times that a match as good or better would occur by chance is 0.003 in a database of this size. We interpret this value to indicate that the match is significant. If we increase the permissiveness of the expect threshold to higher values, we find the next best hit (6th) yields an expect value of 224, indicating this sequence (nuclear receptor subfamily 2 group E member) is an insignificant match to our query.

The TMEM106B sequence was threaded onto the initial model with phenix.sequence_from_map. The TMEM106B model fit the density better than the query model (Extended Data Fig. 7b). We extended the model to include five layers by applying the helical symmetry operators of the map and the model was refined with phenix.real_space_refine44. The

orientations of mainchain oxygen atoms were manually adjusted to form inter-layer hydrogen bonds. The model was refined again with phenix.real_space_refine and the final model was validated using MolProbity[45]. The model was built and adjusted with Coot[42].

For PM2 and PM3 of donor 1, we fit a single layer of PM1 model into PM2 or PM3 maps with rigid body fit, and manually adjusted the model with Coot. A five-layer model (containing ten chains) was generated with the helical symmetry of PM2 or PM3. Model refinement was performed with phenix.real_space_refine and the final model was validated using MolProbity[46]. No major conformational changes were observed during model refinement.

For all polymorphs of donor 2–4, we did not build models de novo. Instead, we fit a single chain of each donor 1 polymorph into corresponding donor 2–4 maps as a rigid body (Extended Data Fig. 6). For PM2 and PM3, a one-layer model was generated by applying their helical symmetry. All models fit the maps well, suggesting that TMEM106B forms fibrils with the same fold in different FTLD-TDP patients.

Standard free energy calculation

Standard free energy values were estimated using a solvation energy algorithm as described previously[47]. We note that our free energy calculations neglect the contribution of glycans, as we cannot confirm the composition or conformation of the sugar groups.

Patient genotyping

Genomic DNA was extracted from patient brain tissues by the UCLA Technology Center for Genomics and Bioinformatics. Primers were designed to amplify the 7 exonic DNA fragments encoding TMEM106B (hg38 chr7:12,214,811–12,231,975, exons 3 to 9); primers were synthesized by Integrated DNA Technologies. PCR amplification was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and the PCR products were sequenced

by Genewiz. For donor 3, sequencing results suggest the genotype of TMEM106B is wild type. For donors 1, 2 and 4, the sequencing results indicate a heterozygous Thr185Ser variant: at position hg38 chr7:12,229,791 in exon 6 of the TMEM106B gene, there is equal detection of the wild-type cytosine nucleotide and the variant guanine nucleotide. This single nucleotide polymorphism would induce a point mutation from wild-type Thr185 (ACC) to variant Ser185 (AGC). From the 50–50 distribution of cytosine and guanine nucleotides measured in the sequencing chromatogram, we conclude that one allele contains the wild-type gene and the other harbours the variant.

Western blotting

The supernatant and pellet from the first 21,000*g* centrifugation step of fibril extraction were used as samples for immunoblotting. Sample solutions were mixed with SDS–PAGE loading dye containing 2 M urea and 1 M β-mercaptoethanol, sonicated for 10 min in ice water, then boiled at 100 °C for 10 min. Samples were loaded onto a NuPAGE 4–12%, Bis-Tris, 1.0 mm, 12-well Mini Protein Gel (Invitrogen) and electrophoresis was performed at 200 V for 30 min. Proteins were wet transferred onto 0.2 μm nitrocellulose membranes (Bio-Rad) by application of a constant 35 V overnight, in transfer buffer consisting of 25mM Tris, pH8.3, 192mM glycine, 20% (w/v) methanol. Membranes were incubated with gentle rocking in 5% (w/v) Blocking-Grade Blocker (milk, Bio-Rad) in 1× Tris-buffered saline, with 0.1% (v/v) Tween20 (TBST). Membranes were incubated for 1 h with TMEM106B antibody (Novus Biologicals, cat. no. NBP1-91311, lot no. QC18333-42825) diluted 1:300, TDP-43 monoclonal antibody (Proteintech, cat. no. 60019-2-IG, lot no. 10011784) diluted 1:500, or anti TAR DNA-binding protein 43 phospho-Ser409/410 monoclonal antibody (Cosmo Bio, cat. no. CAC-TIP-PTD-M01, lot no. 11-9-20) diluted 1:300 in 2% milk in TBST. Membranes were washed three times in

TBST with gentle rocking for 5 min. Then, membranes were incubated for 1 h with either goat anti-rabbit IgG HRP (Invitrogen, cat. no. A27036, lot no. 2116291) or goat anti-mouse IgG HRP (Abcam, cat. no. ab205719, lot no. GR3271082-2) diluted 1:4,000 in 2% milk in TBST. Membranes were washed three times in TBST with gentle rocking for 5 min. Pierce ECL Plus Western Blotting Substrate (Thermo Scientific) was applied to membranes. Membranes were imaged using an Azure 600 (Azure Biosystems). Western blotting was repeated three times independently with similar results.

In addition to the ~35 kDa band on the TMEM106B western blot (Fig. 1b, left), bands with molecular weights (>110 kDa) higher than what is expected for TMEM106B were observed in the sarkosyl-soluble (3 bands) and insoluble fractions (1 band). These bands could correspondto different cleavage products or different post translational modification states of TMEM106B. The high molecular weights could be due to TMEM106B oligomerization or association with other macromolecules to form SDS-resistant complexes. Although the identify of these species cannot be confirmed, the conclusion of this study is unaffected as the ~35kDa band has been validated by mass spectrometry.

Immunogold labelling

Four-hundred-mesh carbon coated copper grids were glow discharged for 30 s before sample preparation. Three microlitres of sample solution was applied to the grids and incubated for 3 min, then excess sample solution was blotted off using filter paper. Blocking buffer (phosphate buffered saline, pH 7.4, 0.1% w/v gelatin) was applied to the grids and incubated for 10 min; excess solution was blotted off. TMEM106B antibody specific for residues 2–53 (Atlas Antibodies, cat. no. HPA058342, lot no. 22721), residues 204–253 (Novus Biologicals, cat. no. NBP1-91311, lot no. QC18333-42825) or residues 218-252 (antibodies-online, cat. no.

ABIN6578799, lot no. SA160811DF), or TDP-43 Monoclonal Antibody (Proteintech, cat. no. 60019-2-IG, lot no. 10011784) diluted 1:100 in blocking buffer was applied to the grids and incubated for 30 min; excess solution was blotted off. Grids were washed five times with blocking buffer, and excess liquid was blotted off between each wash. 6 nm Colloidal Gold AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch Laboratories, cat. no. 115-195-146, lot no. 150545) or 6 nm Colloidal Gold AffiniPure Goat Anti-Rabbit IgG (Jackson ImmunoResearch Laboratories, cat. no. 111-195-144, lot no. 146470) diluted 1:8 in blocking buffer was applied to the grids and incubated for 30 min; excess solution was blotted off. Grids were washed five times with distilled water, and excess liquid was blotted off between each wash. Finally, grids were stained with 3 µl of 2% uranyl acetate for 1 min, washed with an additional 3 µl of 2% uranyl acetate and air-dried for 10 min. The grids were imaged using a Tecnai T12 electron microscope.

Mass spectrometry

The pellet from the first 21,000*g* centrifugation step of fibril extraction was analysed by SDS-PAGE as previous described. Gels were stained by InstantBlue Coomassie Protein Stain (Abcam) and gel bands were excised for mass spectrometry analysis. Protein digestion and peptide identification were adapted from a previously described protocol[48,49]. In brief, proteins trapped in gel bands were reduced with 10 mM dithiothreitol (Sigma) at 60 °C for 1 h, alkylated with 50 mM iodoacetamide (Sigma) at 45 °C for 45 min in the dark, digested with 200 ng trypsin (Promega) at 37 °C overnight and then 100 ng endoproteinase GluC (New England BioLabs) was added for another overnight incubation at 25 °C. Peptides were extracted from the gel bands in 50% acetonitrile/49.9% water/ 0.1% trifluoroacetic acid (TFA) and cleaned with C18 StageTip[50] before mass spectrometry analysis. Digested peptides were separated on an EASY-

Spray column (25 cm × 75 μm internal diameter, PepMap RSLC C18, 2 μm, Thermo) connected to an U3000 RSLCnano HPLC (Thermo) and eluted using a gradient of 3%–32% acetonitrile in 0.1% formic acid and a flow rate of 300 nl min^{-1} (total time 45 min). Tandem mass spectra were collected in a data-dependent manner with an Orbitrap Exploris 480 mass spectrometer (Thermo) interfaced to a nano-ESI source (Thermo). Raw MS/MS data were analysed using Mascot (version 2.5) and files were searched against the UniProt human database (as of 19 December 2018; 20,387 entries) supplemented with common laboratory contaminants. The search parameters were: enzyme specificity: trypsin and GluC; maximum number of missed cleavages: 6, precursor mass tolerance, 10 ppm; product mass tolerance, 0.02 Da; variable modifications included cysteine carbamidomethylation and methionine oxidation. Peptide-spectrum matches (PSMs) were filtered to 1% false discovery rate using the target–decoy strategy. Additional searches were performed applying identical parameters to Mascot's error-tolerant algorithm. The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via MassIVE partner repository with the dataset identifier PXD029876.

Mass spectrometric analysis identified peptide LNNISIIGPLDMK (corresponding to TMEM106B 181–193) from the sarkosyl-insoluble fraction of donor 1. The Thr185Ser variant is present in this peptide, suggesting the variant protein is present in the fibrils. The peptide also contains the N184 glycosylation site without a glycan modification. Detection of this unglycosylated peptide does not contradict the glycosylation of N184 suggested by the cryo-EM structures. The type and size of the glycan could not be determined, thus it was much more difficult to identify the glycosylated peptide using mass spectrometry and only the unglycosylated peptide was detected.

Data availability

Cryo-EM maps and atomic models of proteins from FTLD-TDP donor 1 have been deposited in the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB) with accession codes 7SAQ and EMD-24953 for PM1, 7SAR and EMD-24954 for PM2, and 7SAS and EMD-24955for PM3, respectively. Mass spectrometry data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier PXD029876. Any other relevant data are available from the corresponding author upon reasonable request.

Code availability

Energetic calculations were performed using custom written software; the code is available at https://doi.org/10.5281/zenodo.6321286.

Acknowledgements

We thank Dr. Z. Hong Zhou for the use of Electron Imaging Center for Nanomachines (EICN) instruments supported by the NIH (1S10RR23057 and IS10OD018111), NSF (DBI-1338135) and California NanoSystems Institute at UCLA. We thank Drs. Sjors Scheres and Michel Goedert for discussion of TMEM106B antibodies. The authors acknowledge NIH AG048120, NIH AG070895, DOE DE-FC02-02ER63421, NIH GM103479, NIH GM007185, NSF DGE-1650604, ADRC (P30 AG062677), Einstein Aging Study (P01 AG003949), ALLFTD (U19 AG063911), C9ORF72 P01 (P01 NS084974), and the National Facility for Translational Medicine (Shanghai) for support.

Author contributions

Y.X.J. and Q.C. extracted fibrils from the material donated by patients with FTLD-TDP, prepared cryo-EM grids, and collected and processed cryo-EM data. Y.X.J., Q.C. and M.R.S. built the atomic models. P.G. assisted in cryo-EM data collection. Y.X.J. and R.A. performed western blotting and immunolabelling assays. Y.X.J. performed patient genotyping. J.Y.F., R.R.O.L. and J.A.L. performed mass spectrometry. M.D. and D.W.D. prepared frozen brain samples and performed immunohistochemistry staining. Q.C. and M.R.S. performed solvation energy calculations. All authors analysed the results and wrote the manuscript. D.S.E. supervised the project.

Competing interests

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

Peer review

Nature thanks Henning Stahlberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

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

The online version is available at https://www.nature.com/articles/s41586-022-04670-9

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

Figures and Tables

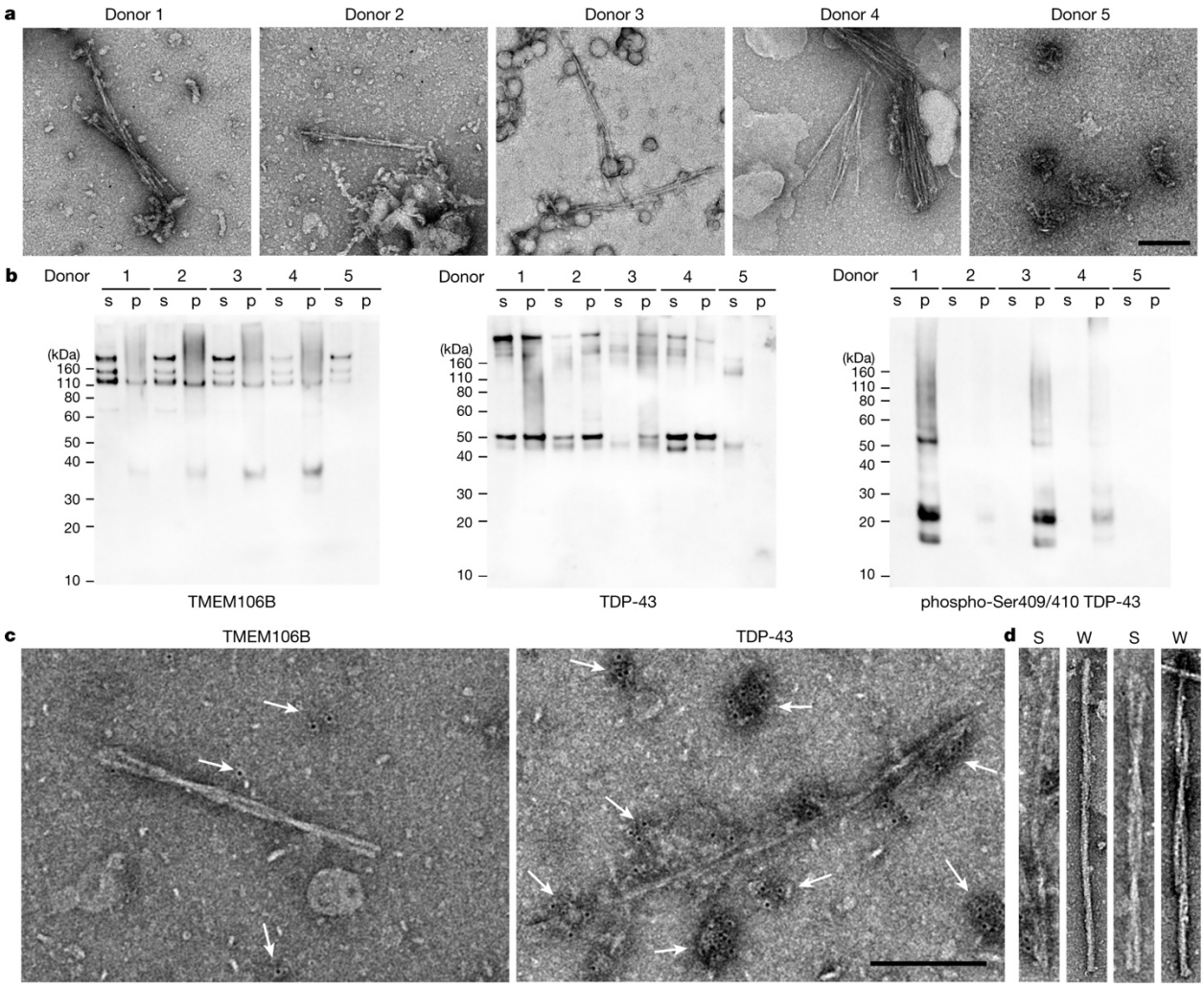

Figure 1-1: Characterization of TMEM106B and TDP-43 in FTLD-TDP brain extracts. a, Negative-stain transmission electron microscopy (TEM) images of sarkosyl-insoluble fractions from FTLD-TDP donors 1 to 4 and a non-FTLD-TDP donor (donor 5). Scale bar, 200 nm. **b**, Western blots of sarkosyl-soluble (s) and sarkosyl-insoluble (p) fractions from donors 1 to 5 probed with TMEM106B (left), TDP-43 (middle) or phospho-Ser409/410 TDP-43 (right) antibody. The original, uncropped blots are shown in Supplementary Fig. 3. **c**, Representative negative-stain TEM images of the sarkosyl-insoluble fraction from FTLD-TDP donor 1 showing

immunogold labelling with TMEM106B (left) and TDP-43 (right) antibodies. White arrows label immunogold beads, showing low-density, probably nonspecific binding of TMEM106B antibody (left) and high-density, non-fibrillar TDP-43 aggregates (right). Scale bar, 200 nm. **d**, Comparison of fibrils extracted from FTLD-TDP donor 1 using the sarkosyl-based protocol (S; same fibrils as in **c**) and a water-based protocol (W) imaged by negative-stain TEM. PM1 (left two images) and PM2 or PM3 (right two images; PM2 and PM3 are difficult to distinguish by negative-stain TEM). Fibrils extracted from both protocols exhibit similar morphology.

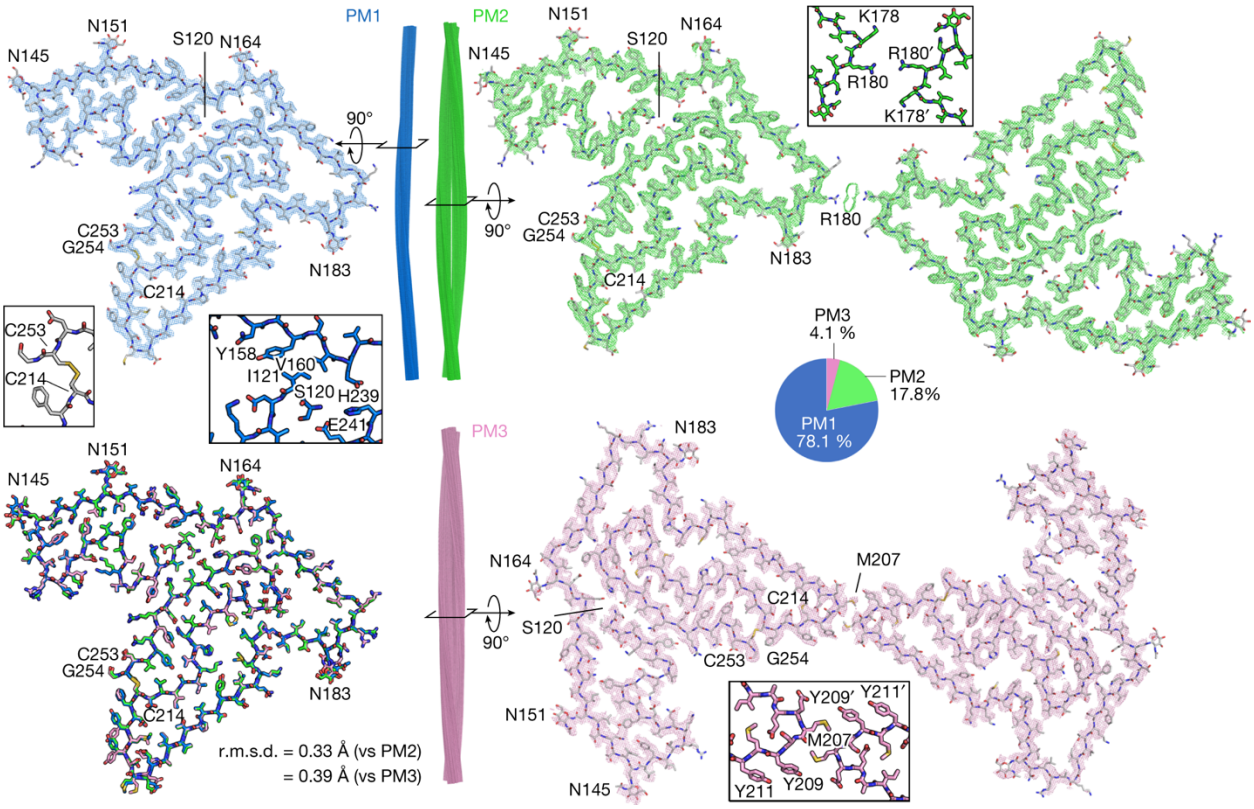

Figure 1-2: Cryo-EM structures of TMEM106B fibrils from FTLD-TDP donor 1. Cryo-EM maps and atomic models of one cross-sectional layer of PM1 (blue, top left), PM2 (green, top right) and PM3 (pink, bottom right). Side view of fibril reconstructions of PM1, PM2 and PM3 (middle). Superimposition of a single chain from PM1, PM2 and PM3 (bottom left). Enlarged

views of the C214–C253 disulfide bond, N-terminal Ser120 in PM1, and dimer interfaces of PM2 and PM3 are shown as insets. The distribution of PM1, PM2 and PM3 polymorphs in the donor 1 dataset is shown as a pie chart.

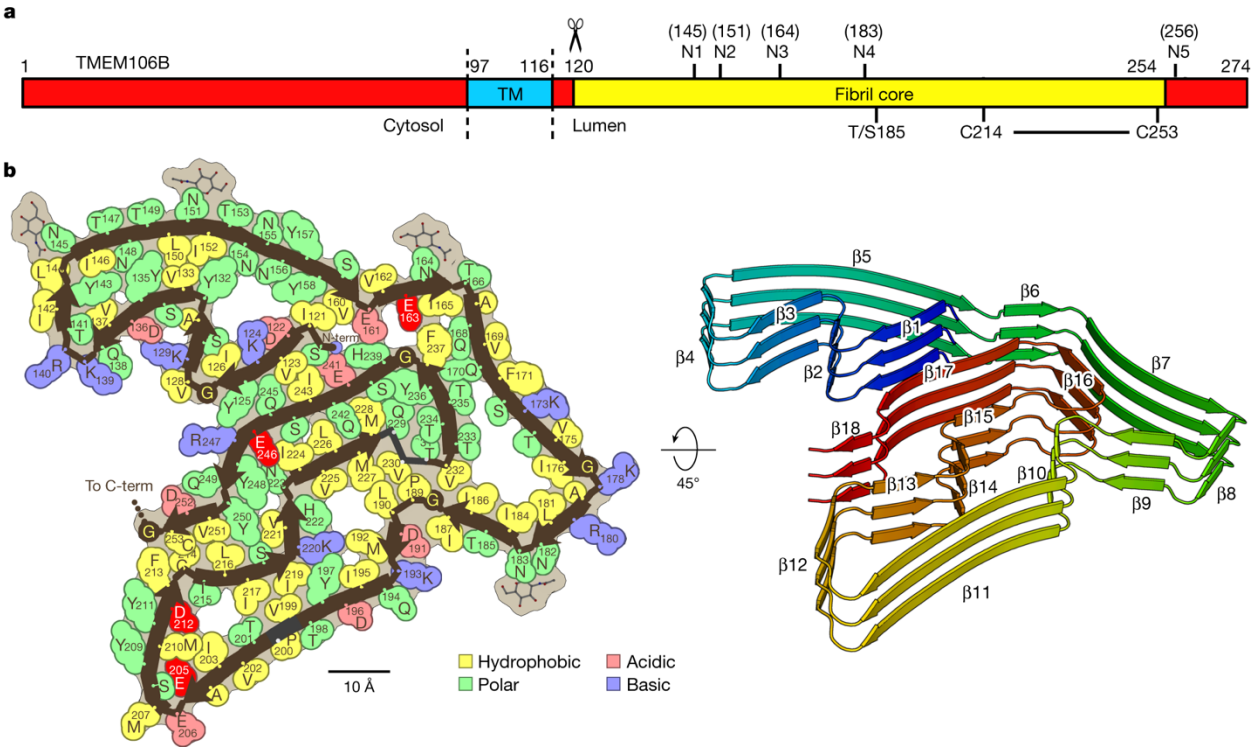

Figure 1-3: The conserved golf-course-like fold of TMEM106B fibrils. a, Schematic of TMEM106B showing the cytosolic, transmembrane (TM) and lumenal domains, as well as the fibril core of the structures observed here. Labelled are the five known glycosylation sites (N1–N5), the heterozygous Thr185Ser variant and the C214–C253 disulfide bond. **b**, Space-filling representation (left) and cartoon model (right) of the conserved protofilament core represented by PM1 from FTLD-TDP donor 1. Uncomplemented acidic residues buried inside the fibril core are highlighted in bright red. N-term, N terminus; C-term, C terminus.

Donor ID	Gender	Age	Braak	PathDx (type)	FHx	Brain	Distribution (%)			TMEM106B gene
						region	PM1	PM2	PM3	
Donor 1	Male	65	2	FTLD-TDP (C)	False	MF	78.1	17.8	4.1	Heterozygous T185S
Donor 2	Female	76	3	FTLD-TDP (B)	False	MF	84.9	10.0	5.1	Heterozygous T185S
Donor 3	Male	86	0	FTLD-TDP (A)	False	MF	85.3	12.8	1.8	Wild type
Donor 4	Female	64	3	FTLD-TDP (D)	True	MF	85.3	10.7	4.0	Heterozygous T185S
Donor 5	Male	59	0	Normal	False	MF	n.a.	n.a.	n.a.	n.a.

Table 1-1: Information on FTLD-TDP and non-FTLD-TDP donors. PathDx, pathological diagnosis; FHx, family history; MF, medial frontal gyrus; n.a., not available.

	PM1 (EMD-24953, PDB 7SAQ)	PM2 (EMD-24954, PDB 7SAR)	PM3 (EMD-24955, PDB 7SAS)
Data collection and processing			
Magnification	×81,000	×81,000	×81,000
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	C_1	C_2	C_1
Helical rise (Å)	4.91	4.905	2.452
Helical twist (°)	−0.42	179.59	179.79
Initial particle images (no.)	883,793	883,793	883,793
Final particle images (no.)	60,806	16,883	9,203
Map resolution (Å)	2.9	3.2	3.7
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	200-2.9	200-3.2	200-3.7
Refinement			
Initial model used (PDB code)	De novo	7SAQ	7SAQ
Model resolution (Å)	3.1	3.4	4.3
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)	200-3.1	200-3.4	200-4.3
Map sharpening *B* factor (Å^2)	124	93	98
Model composition			
Nonhydrogen atoms	5,710	11,420	11,420
Protein residues	675	1,350	1,350
Ligands	20	40	40
B factors (Å^2)			
Protein	7.10	6.57	29.53
Ligand	11.64	10.61	52.32
R.m.s. deviations			
Bond lengths (Å)	0.005	0.004	0.004
Bond angles (°)	0.793	0.732	0.818
Validation			
MolProbity score	2.2	2.3	2.4
Clashscore	11.9	15.1	18.6
Poor rotamers (%)	0	0	0
Ramachandran plot			
Favoured (%)	87.2	88.6	88.0
Allowed (%)	12.8	11.4	12.0
Disallowed (%)	0	0	0

Table 1-2: Cryo-EM data collection, refinement and validation statistics of FTLD-TDP donor 1.

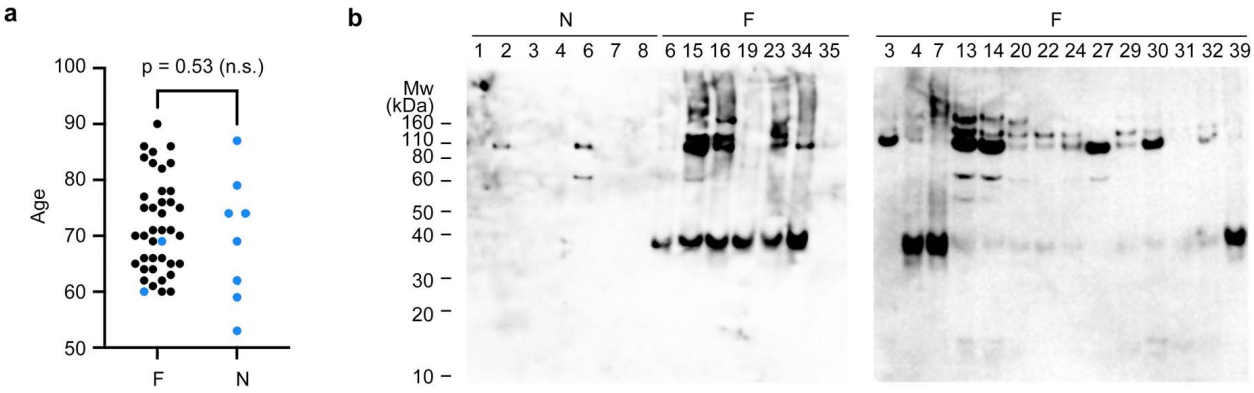

Figure 1-ED1: Comparison of FTLD-TDP and non-FTLD-TDP donors. a, Age distribution of FTLD-TDP donors (F, n = 40) and non-FTLD-TDP controls (N, n = 8). Donors with and without fibrils detected under negative stain EM are coloured black and blue, respectively. P-value of 0.53 (n.s., not significant) from an unpaired, two-tailed t-test suggests that the presence of fibrils is disease-dependent but not age-dependent. **b**, Western blots of sarkosyl-insoluble fractions from FTLD-TDP (F) and non-FTLD-TDP donors (N) probed by TMEM106B antibody. The ~35 kDa TMEM106B-positive band was found in none of the non-FTLD-TDP donors (N5 shown in Fig. 1b as donor 5) and all of the FTLD-TDP donors except F35, F3, and F27. P-value of less than 0.0001 was obtained from an unpaired, two-tailed t-test comparing presence of the ~35kDa band (value of "1" for present, "0" for absent) in non-FTLD-TDP donors (0 out of 8, n = 8) and FTLD-TDP donors (22 out of 25, n = 25). Fibrils were not observed in F3 and F27 by negative stain EM, consistent with the western blot. Fibrils were observed in F35, which suggests that western blot may not always be accurate in detecting TMEM106B aggregation. Western blot membranes were prepared in parallel and exposed with equal time. The original, uncropped blots are shown in Supplementary Fig. 3.

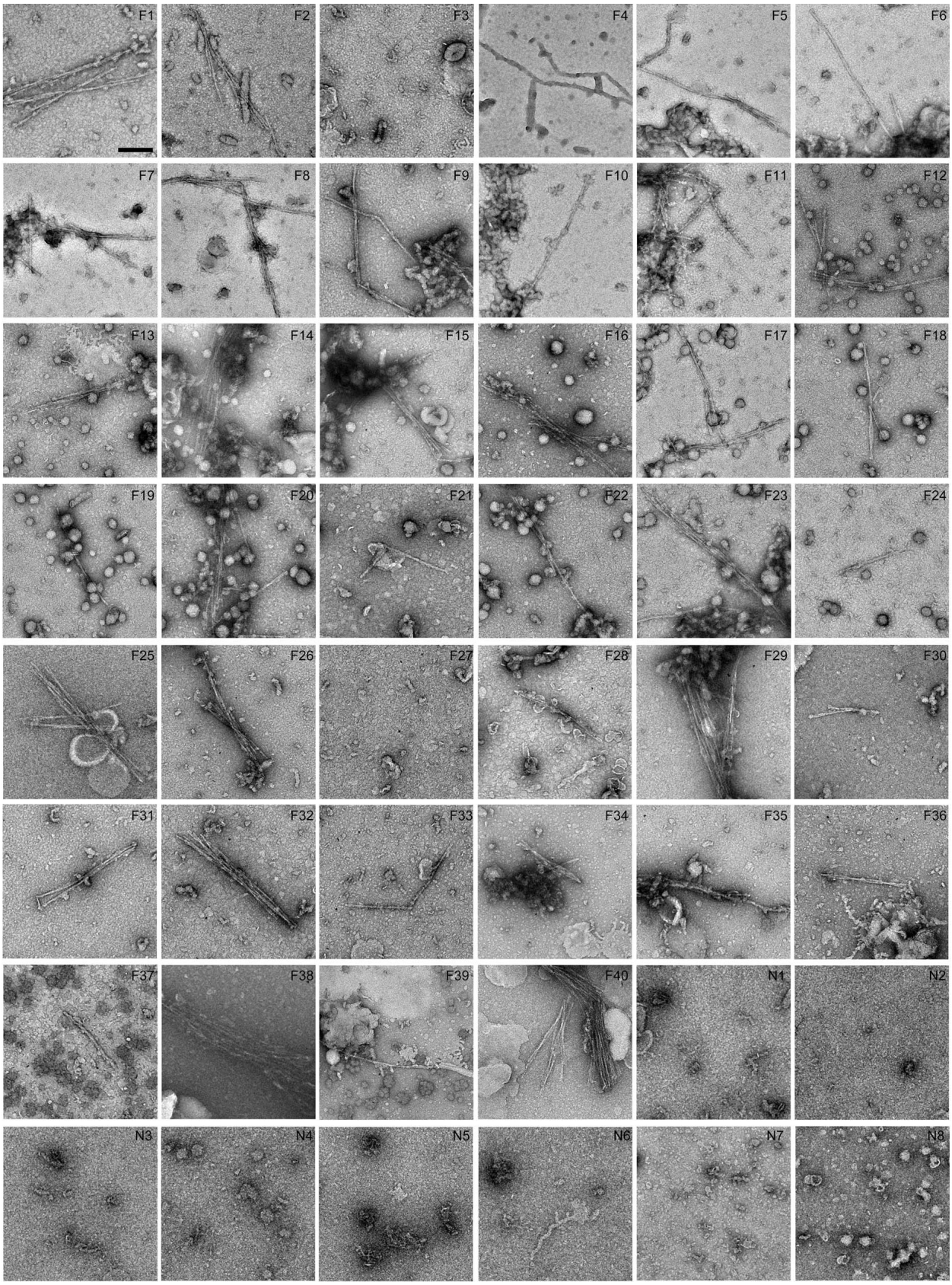
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
F13
F14
F15
F16
F17
F18
F19
F20
F21
F22
F23
F24
F25
F26
F27
F28
F29
F30
F31
F32
F33
F34
F35
F36
F37
F38
F39
F40
N1
N2
N3
N4
N5
N6
N7
N8

Figure 1-ED2: Fibril screen of FTLD-TDP and non-FTLD-TDP donors by negative stain EM. Negative stain EM images of sarkosyl-insoluble fractions from all donors. Scale bar 200 nm. Donors 1–5 (F26, F36, F17, F40 and N5, respectively) are also shown in Fig. 1a. Fibrils with similar morphologies were found in all FTLD-TDP donors except F3 and F27. No fibrils were found in any non-FTLD-TDP donors. P-value of less than 0.0001 was obtained from an unpaired, two-tailed t-test comparing fibrils detected by EM (value of 1 for present, 0 for absent) in FTLD-TDP donors (38 out of 40, n = 40) and non-FTLD-TDP donors (0 out of 8, n = 8).

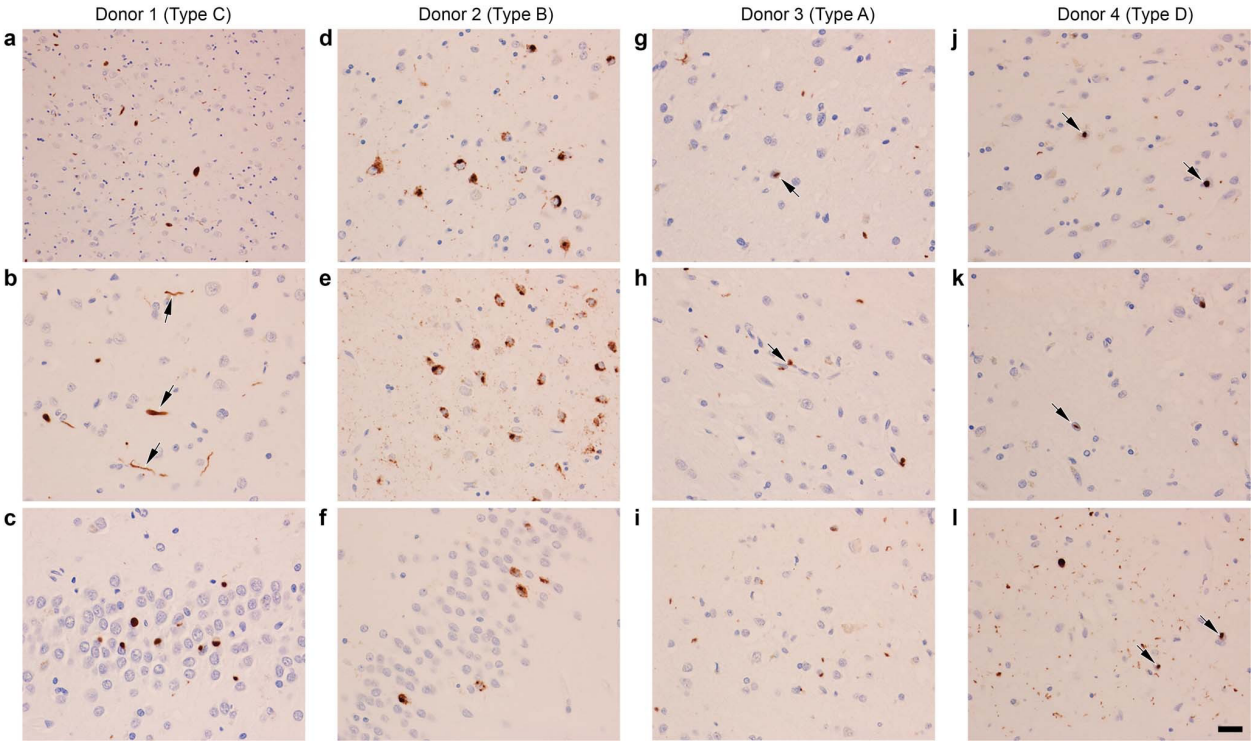

Figure 1-ED3: Neuropathological diagnosis of donors 1–4 as FTLD-TDP. Immunohistochemistry staining using a phosphor-Ser409/410 TDP-43 antibody was performed for brain sections from donors 1–4. All four donors were confirmed to be FTLD-TDP cases, representing the 4 subtypes A to D, respectively. For all figures, scale bar 20 µm. **a**, **b**, **c**, Donor 1 is FTLD-TDP type C, with long, thick neurites (arrows in **b**) and 'Pick-body like NCI' in dentate

fascia (**c**). **d**, **e**, **f**, Donor 2 is FTLD-TDP type B, displaying characteristic granular cytoplasmic NCI in cortex (**d**), hippocampus (**e**) and dentate fascia (**f**). **g**, **h**, **i**, Donor 3 is FTLD-TDP type A, exhibiting small dense neuronal cytoplasmic inclusions (NCI), sparse neuronal intranuclear inclusions (NII, arrow in **g**), and perivascular glial inclusions (arrow in **h**). **j**, **k**, **l**, Donor 4 is FTLD-TDP type D, shown by frequent NII (arrows in **j**, **k** and **l**) and small NCI and neurites. **a**, **b**, **d**, **e**, **g**, **h**, **i**, **j**, **k**, **l** are brain sections from the temporal cortex. **c** and **f** are brain sections from dentate fascia.

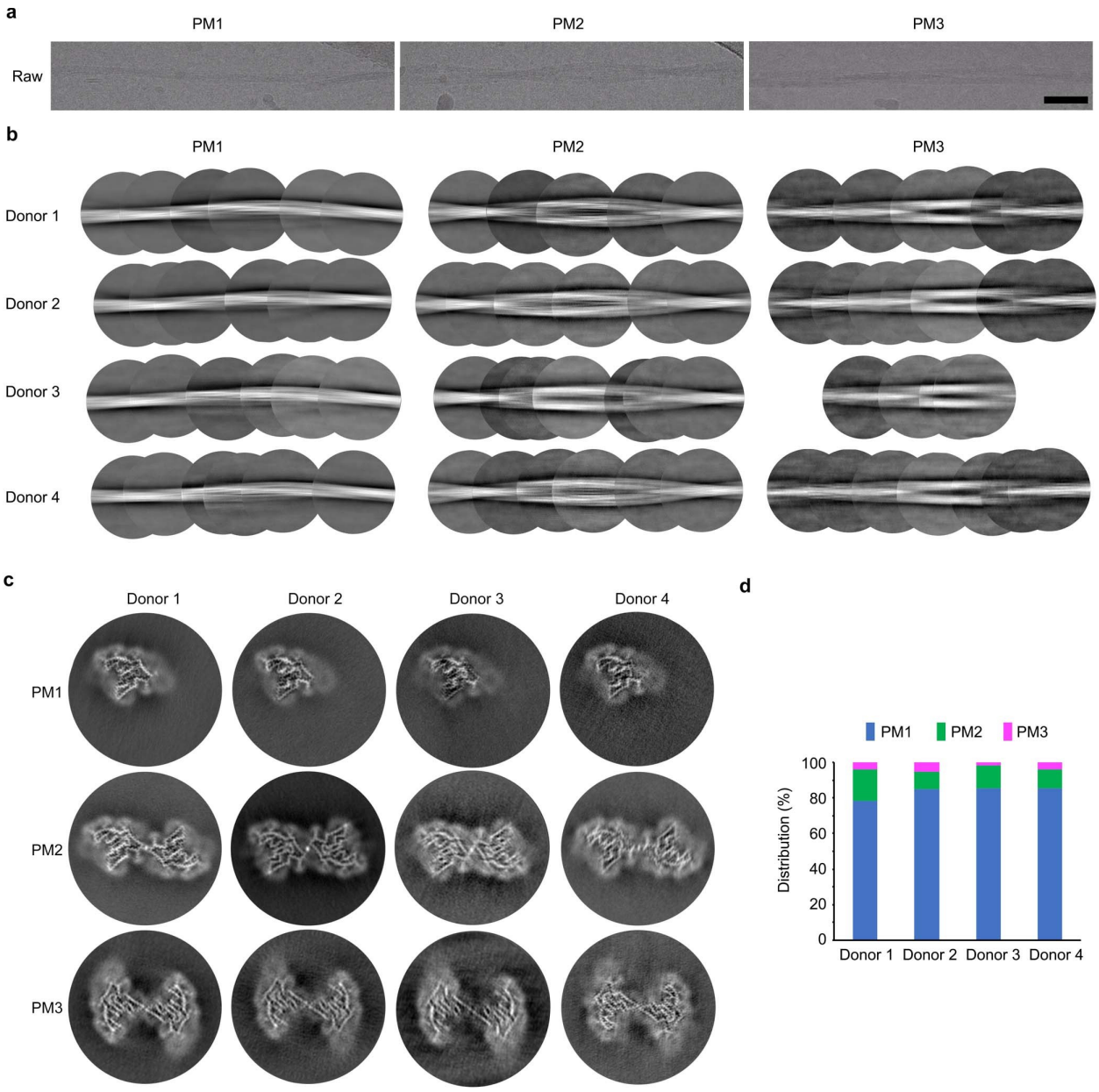

Figure 1-ED4: Cryo-EM data processing of amyloid fibrils from FTLD-TDP donors 1 to 4. **a**, Representative micrographs of PM1-3 from FTLD-TDP donor 1. Scale bar 500 Å. **b**, Representative 2D classes of PM1-3 from FTLD-TDP donors 1 to 4. 2D classes are stitched together to show a full cross-over of each morphology. **c**, 3D reconstructions of PM1-3 from the four FTLD-TDP donors. **d**, Distributions of the three fibril polymorphs in the four FTLD-TDP donors.

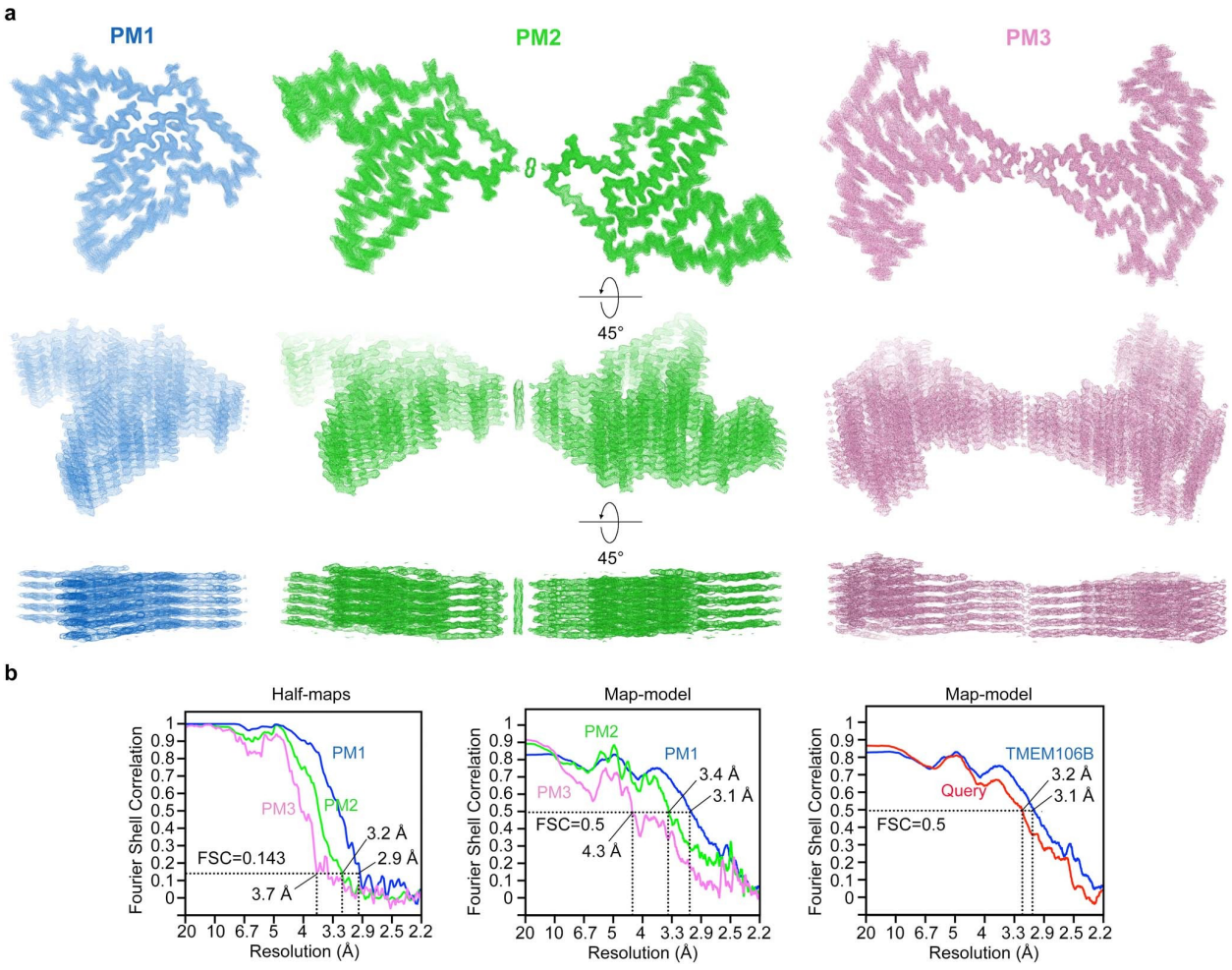

Figure 1-ED5: Cryo-EM maps and FSC curves of FTLD-TDP donor 1. a, Views of the cryo-EM maps from FTLD-TDP donor 1 with five layers shown. **b**, FSC curves between two half-maps (left) and the cryo-EM reconstruction and refined atomic model (middle) of each polymorph PM1 (blue), PM2 (green), and PM3 (pink). FSC curves between cryo-EM reconstruction and the query model (red, Direction 1 chain of Extended Data Fig. 5a) and the atomic model of PM1 from FTLD-TDP donor 1 (blue) are compared on the right.

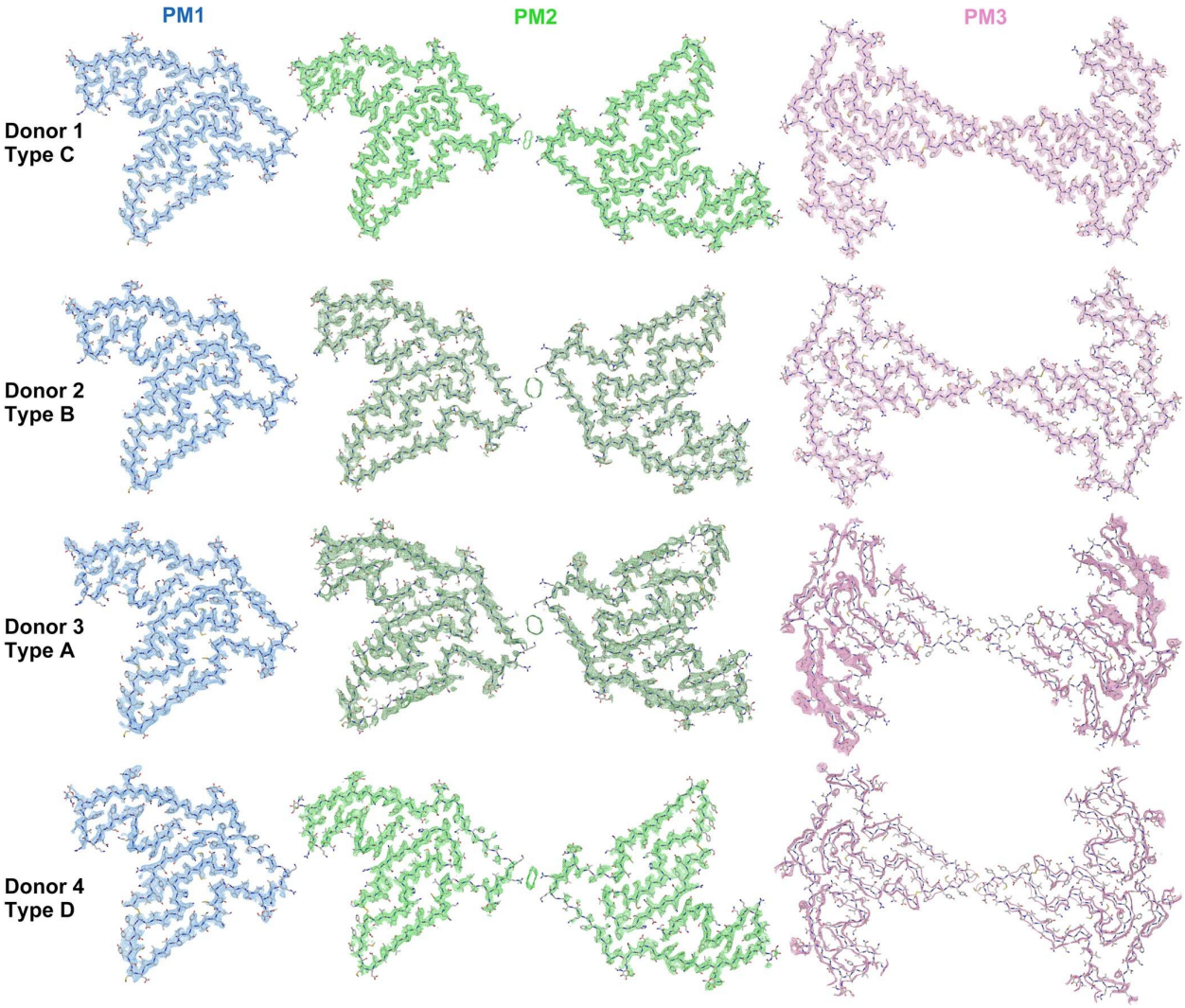

Figure 1-ED6: Cryo-EM maps of TMEM106B fibrils from FTLD-TDP donors 1 to 4. The models of PM1-3 from donor 1 were rigid body fitted into the maps of PM1-3 from donors 2 to 4. These cryo-EM structures reveal that the polymorphs from all four donors share the same protofilament fold. Two subtypes of PM2 are exhibited by donors 1 and 4 (light green) and donors 2 and 3 (dark green), respectively.

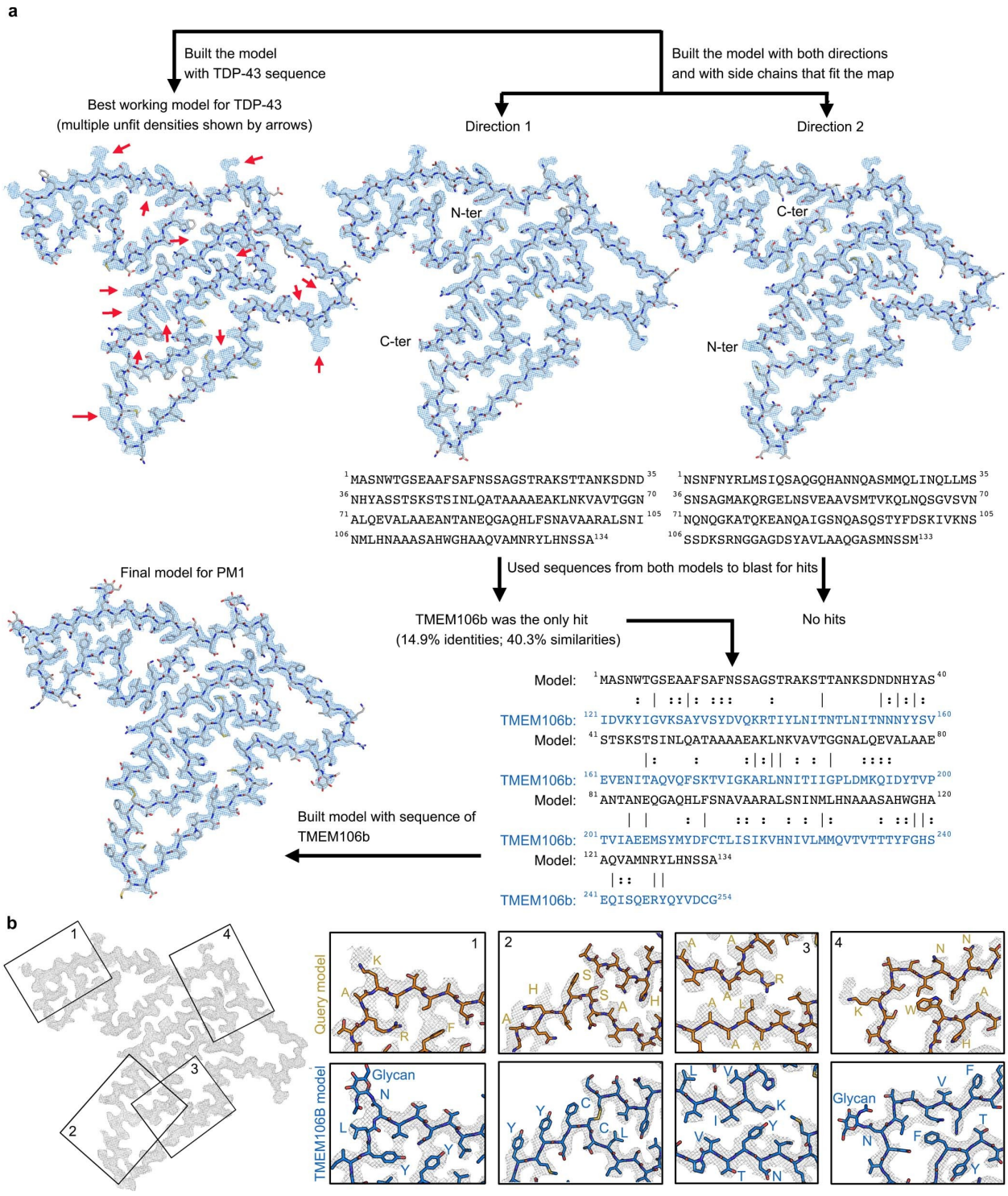

Figure 1-ED7: Identification of the TMEM106B molecule by atomic model building. a, Atomic model building flowchart for PM1 of FTLD-TDP donor 1. In the sequence alignment (bottom right), lines indicate identical residues and two dots indicate similar residues. **b**,

Comparison of query model (Direction 1 chain, orange) with final model built with TMEM106B sequence (blue). Four representative regions of the cryo-EM map are shown. Residues that show clear differences in side chain density fitting are labelled in both models.

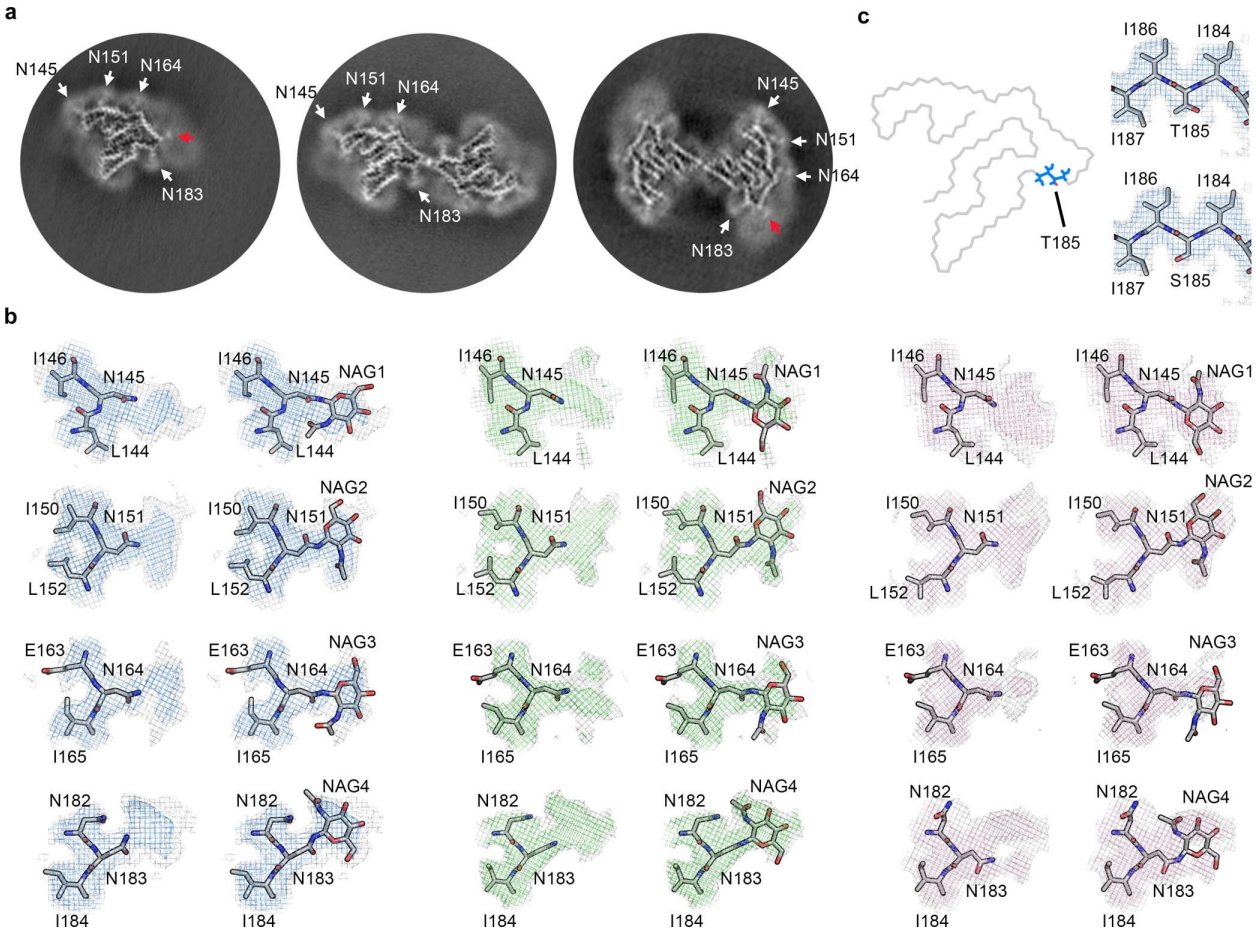

Figure 1-ED8: Evidence of glycosylation and genetic polymorphism in maps of TMEM106B fibrils. a, Three-dimensional reconstructions of PM1-3 from FTLD-TDP donor 1. White arrows point to the four glycosylation sites within the fibril core. Red arrows point to the residual densities outside the fibril core in PM1 and PM3, which may correspond to the binding of the same undefined, negatively charged ligand present in the PM2 dimer interface (Extended Data Fig. 9). **b**, Maps and models of the four glycosylation sites with or without the sugar group

in PM1 (left), PM2 (middle), and PM3 (right) from FTLD-TDP donor 1. **c**, Position of Thr/Ser185 genetic polymorphism in the conserved fibril fold (left) and the map and model of the Thr/Ser185 environment (right, represented by PM1 from donor 1).

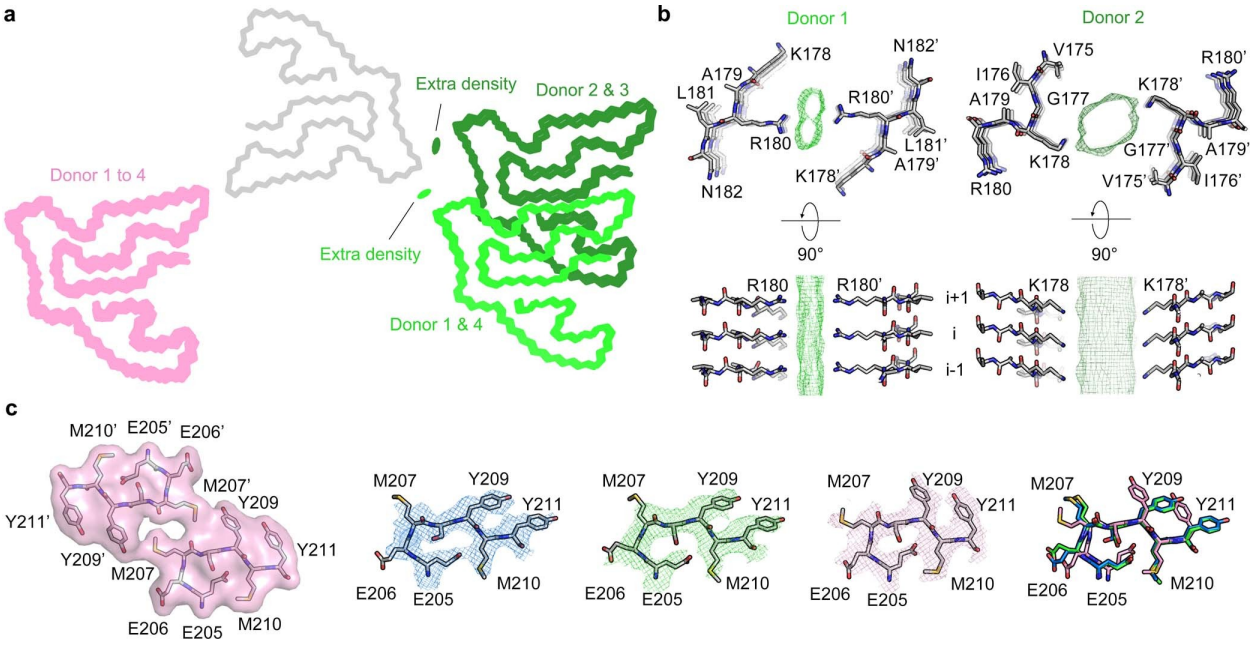

Figure 1-ED9: Diverse dimer interfaces of PM2 and PM3 of TMEM106B fibrils. a, Dimer arrangements of PM2 and PM3 from FTLD-TDP donors 1 to 4. PM2 and PM3 from all donors are aligned at chain A (grey, represented by PM2 of donor 1). Chain B of PM2 (light green for donors 1 and 4, dark green for donors 2 and 3) and PM3 (pink) from each donor is shown. Residual densities in the PM2 dimer interfaces are shown as green ovals. The dimer arrangement of PM3 is consistent among all donors, whereas there are two subtypes of dimer arrangements for PM2. **b**, Atomic model and the residual density in the dimer interface of donor 1 (left, represents donors 1 and 4) or donor 2 (right, represents donors 2 and 3). In donors 1 and 4, Arg180 from each protofilament is on the opposite sides of an extra density in the middle of the PM2 dimer interface; in donors 2 and 3, the dimer interface is shifted to Lys178. Although two

slightly different dimer interfaces were observed, we consider PM2 in all four FTLD-TDP donors to be the same morphology because of the similarity in dimer formation (see Discussion). **c**, Comparison of the residues near the PM3 interface (far left) from PM1 (blue), PM2 (green), PM3 (pink), and the superimposition of those residues from PM1-3 (far right). PM1-3 are all represented by FTLD-TDP donor 1.

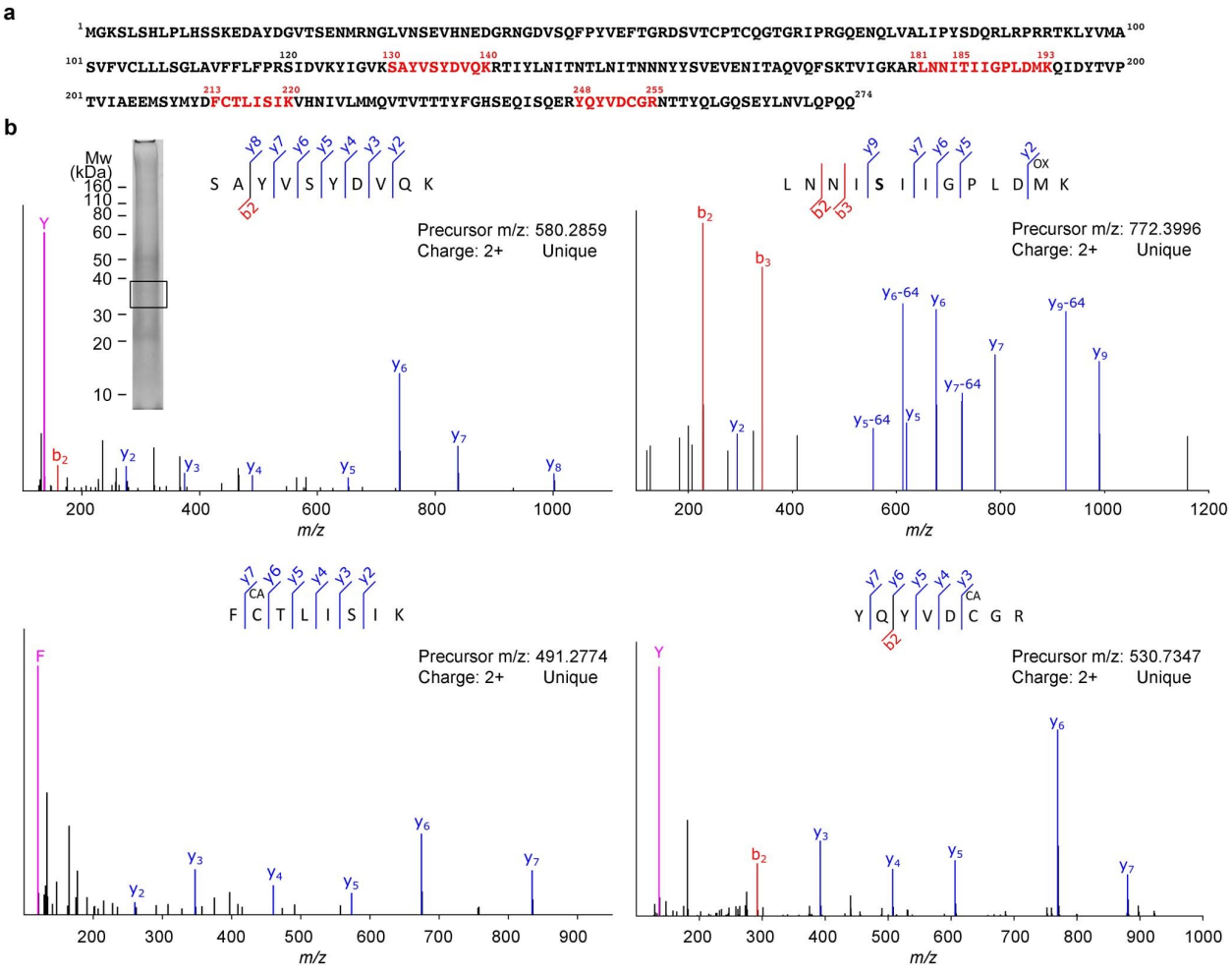

Figure 1-ED10: Detection of TMEM106B peptides by mass spectrometry. a, Sequence map of TMEM106B. Sequences in red indicate unique peptides detected by LC-MS/MS from excised gel bands. **b**, Fragmentation spectra (MS/MS) of detected peptides from TMEM106B. Detected fragment ions (b, y, and immonium ions) are labelled accordingly. The peptide modifications

methionine oxidation (OX) and cysteine carbamidomethylation (CA) were observed. SDS-PAGE gel of sarkosyl-insoluble fraction of donor 1 shown as an insert. Box indicates the gel region, which corresponds to the ~35kDa band from the TMEM106B western blot (Fig. 1b), excised for LC-MS/MS analyses.

Original ID	Gender	Age	Braak	PathDx (type)	FHx	Brain region	Fibrils under EM	Band in western blot	Donor ID reported
F1	Male	66	2	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F2	Male	65	2.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F3	Male	60	3	FTLD-TDP (A)	FALSE	MF	No	No	
F4	Female	71	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F5	Male	71	2.5	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F6	Male	61	1	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F7	Male	60	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F8	Female	84	2.5	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F9	Male	76	1	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F10	Female	78	1.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F11	Female	85	1	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F12	Male	63	0	FTLD-TDP (A)	TRUE	MF	Yes	n.a.	
F13	Male	86	2.5	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F14	Male	70	1	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F15	Female	83	3	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F16	Male	60	1	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F17	Male	86	0	FTLD-TDP (A)	FALSE	MF	Yes	Yes	Donor 3
F18	Female	90	2.5	FTLD-TDP (A)	FALSE	MF	Yes	n.a.	
F19	Female	64	2	FTLD-TDP (A)	TRUE	MF	Yes	Yes	
F20	Male	83	3	FTLD-TDP (A)	FALSE	MF	Yes	Yes	
F21	Female	71	0	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F22	Male	70	0.5	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F23	Male	75	0	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F24	Male	74	2.5	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F25	Female	75	2.5	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F26	Male	65	2	FTLD-TDP (C)	FALSE	MF	Yes	Yes	Donor 1
F27	Male	66	3	FTLD-TDP (C)	TRUE	MF	No	No	
F28	Male	75	2	FTLD-TDP (C)	FALSE	MF	Yes	n.a.	
F29	Male	65	1	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F30	Male	78	1	FTLD-TDP (C)	TRUE	MF	Yes	Yes	
F31	Female	69	3	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F32	Male	66	0	FTLD-TDP (C)	FALSE	MF	Yes	Yes	
F33	Female	69	2	FTLD-TDP (C)	TRUE	MF	Yes	n.a.	
F34	Female	82	3	FTLD-TDP (C)	TRUE	MF	Yes	Yes	
F35	Male	70	2.5	FTLD-TDP (B)	TRUE	MF	Yes	No	
F36	Female	76	3	FTLD-TDP (B)	FALSE	MF	Yes	Yes	Donor 2
F37	Female	62	0	FTLD-TDP (B)	FALSE	MF	Yes	n.a.	
F38	Female	62	1	FTLD-TDP (B)	TRUE	MF	Yes	n.a.	
F39	Male	77	2	FTLD-TDP (B)	TRUE	MF	Yes	Yes	
F40	Female	64	3	FTLD-TDP (D)	TRUE	MF	Yes	Yes	Donor 4
N1	Male	69	2	Normal	No	MF	No	No	
N2	Male	62	2	VaD	Yes	MF	No	No	
N3	Male	79	1	VaD	No	MF	No	No	
N4	Female	87	2	VaD	No	MF	No	No	
N5	Male	59	0	Normal	No	MF	No	No	Donor 5
N6	Male	53	1	PART	No	MF	No	No	
N7	Female	74	2	PART	No	MF	No	No	
N8	Male	74	3	PART	No	MF	No	No	

Table 1-ED1: PathDx, pathological diagnosis; FHx, family history; VaD, vascular dementia; PART, primary age-related tauopathy; MF, medial frontal gyrus; band in western blot, detection

of the ~35 kDa TMEM106B antibody-positive band in the sarkosyl-insoluble fraction; n.a., not available. All donors were numbered by original ID (F1-40 for FTLD-TDP cases and N1-8 for non-FTLD-TDP cases) during initial fibril screen. F26, F36, F17, F64 and N5 were selected for further study and renumbered as donors 1–5 in the main text.

	Donor 2			Donor 3			Donor 4		
	PM1	PM2	PM3	PM1	PM2	PM3	PM1	PM2	PM3
Magnification	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000	×81,000
Voltage (kV)	300	300	300	300	300	300	300	300	300
Electron exposure ($e^-/Å^2$)	37	37	37	36	36	36	36	36	36
Defocus range (μm)	1.0-5.1	1.0-5.1	1.0-5.1	0.5-5.1	0.5-5.1	0.5-5.1	0.8-4.9	0.8-4.9	0.8-4.9
Pixel size (Å)	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Symmetry imposed	C_1	C_2	C_1	C_1	C_2	C_1	C_1	C_2	C_1
Helical rise (Å)	4.896	4.9	2.446	4.897	4.906	2.452	4.897	4.904	2.4281
Helical twist (°)	-0.413	179.619	179.807	-0.413	179.6	179.8	-0.413	179.605	179.81
Initial particle images	438,393	438,393	438,393	419,107	419,107	419,107	228,432	228,432	228,432
Final particle images	38,894	24,420	5,317	26,533	15,587	5,335	14,407	13,482	5,798
Map resolution (Å)	3.8	4.0	4.6	3.6	4.1	5.3	3.5	3.6	5.0
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	200-3.8	200-4.0	200-4.6	200-3.6	200-4.1	200-5.3	200-3.5	200-3.6	200-5.0

Table 1-ED2: Cryo-EM data collection and processing statistics of FTLD-TDP donors 2–4.

Sequence	Modifications	Protein Group Accessions	Position in TMEM106B	# PSMs	IonScore	E-value	m/z (Da)	Δ m/z (ppm)
SAYVSYDVQK		Q9NUM4	130-140	3	39	0.0044	580.28595	+1.34
LNNISIIGPLDMK[a]	M12(Oxidation)	Q9NUM4	181-193	—	—	—	722.39960	+6.35
FCTLISIK	C2(Carbamidomethyl)	Q9NUM4	213-220	3	59	2.1e-005	491.27747	+3.89
YQYVDCGR	C6(Carbamidomethyl)	Q9NUM4	248-255	3	34	0.0025	530.73291	+6.41

Table 1-ED3: TMEM106B peptides identified by mass spectrometry. [a]Peptide identified from Mascot's error-tolerant algorithm.

Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Percent Identity	Accession Length	Accession
TMEM106B protein	Homo sapiens	39.3	39.3	93%	0.003	16.00%	274	AAH39741.1
transmembrane protein 106B	Homo sapiens	39.3	39.3	93%	0.003	16.00%	274	NP_001127704.1
transmembrane protein 106B, isoform CRA_b	Homo sapiens	39.3	39.3	92%	0.003	16.13%	285	EAW93639.1
TMEM106B protein	Homo sapiens	38.9	38.9	93%	0.004	16.00%	314	AAH28108.1
hypothetical protein FLJ11273 variant	Homo sapiens	37.0	37.0	92%	0.019	16.13%	274	BAD96983.1
nuclear receptor subfamily 2 group E member 1 isoform b	Homo sapiens	25.4	25.4	50%	224	32.89%	385	NP_003260.1

Table 1-ED4: BLAST search results.

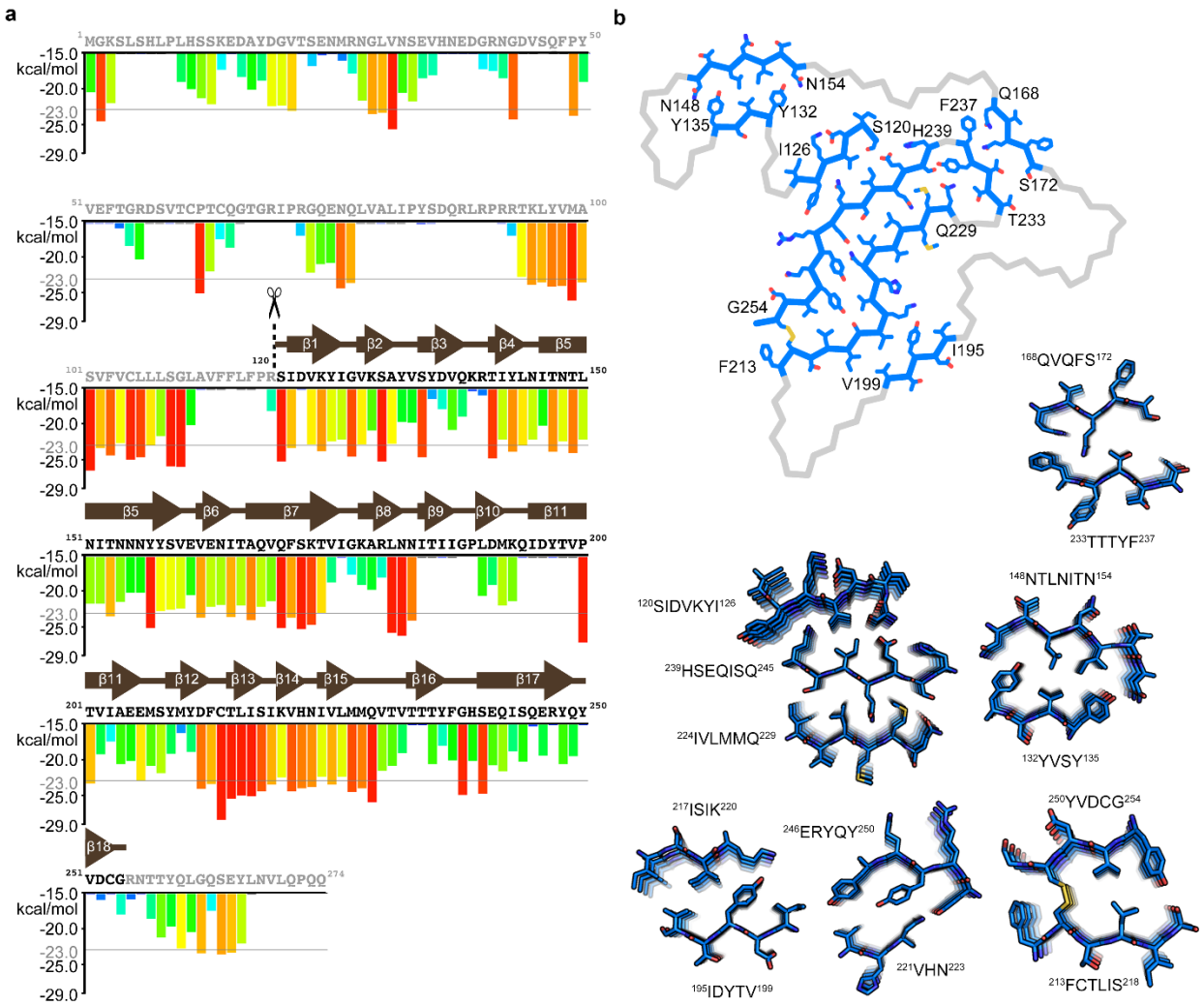

Figure 1-S1: Steric zippers of TMEM106B fibrils. a, Amino acid sequence of TMEM106B and ZipperDB prediction. Red bars indicate high propensity for the six-residue fragments to form

homotypic steric zippers. Residues that form β-strands in the TMEM106B protofilament core are designated by arrows. **b**, Heterotypic steric zippers in the conserved protofilament core of TMEM106B fibrils, represented by PM1 from donor 1.

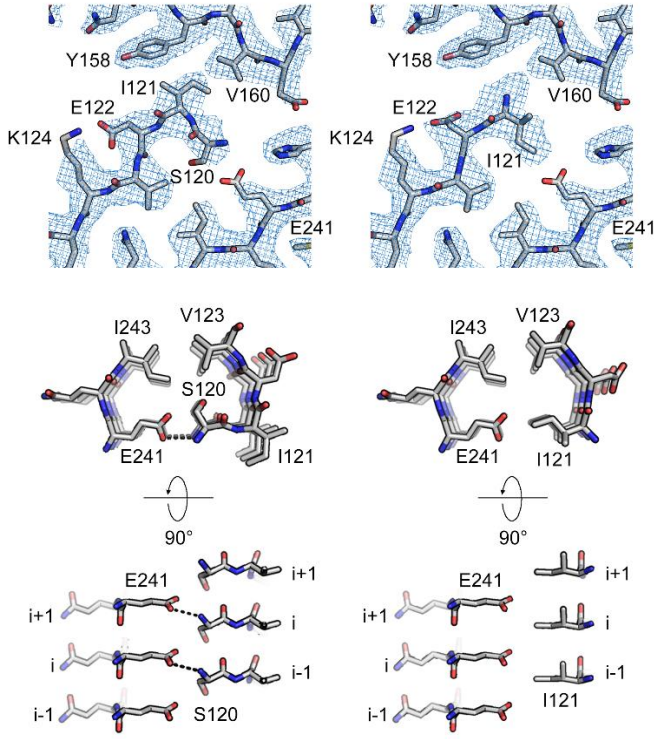

Figure 1-S2: Comparison of models that start with Ser120 or Ile121 at the N- terminus. The current model starting with Ser120 (left) and an alternative model starting with Ile121 (right) are juxtaposed. We believe the N-terminus starting with Ser120 is the better model because i) Ser120 fits the density well, whereas the density is too large for only the nitrogen of Ile121; ii) Ser120 hydrogen-bonds with Glu241, whereas Ile121 does not participate in interactions that favor the fibril fold.

References

1. Mohandas, E. & Rajmohan, V. Frontotemporal dementia: an updated overview. Indian J. Psychiatry 51, S65–S69 (2009).
2. Neary, D. et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 51, 1546–1554 (1998).
3. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314, 130–133 (2006).
4. Van Deerlin, V. M. et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat. Genet. 42, 234–239 (2010).
5. Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. Annu. Rev. Biochem. 86, 27–68 (2017).
6. Benson, M. D. et al. Amyloid nomenclature 2020: update and recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. Amyloid Int. J. Exp. Clin. Invest. 27, 217–222 (2020).
7. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. Cell 148, 1188–1203 (2012).
8. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. Nature 547, 185–190 (2017).
9. Falcon, B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. Nature 561, 137–140 (2018).
10. Falcon, B. et al. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. Nature 568, 420–423 (2019).
11. Kollmer, M. et al. Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. Nat. Commun. 10, 4760 (2019).

12. Schweighauser, M. et al. Structures of α-synuclein filaments from multiple system atrophy. Nature 585, 464–469 (2020).
13. Zhang, W. et al. Novel tau filament fold in corticobasal degeneration. Nature 580, 283–287 (2020).
14. Ratnavalli, E., Brayne, C., Dawson, K. & Hodges, J. R. The prevalence of frontotemporal dementia. Neurology 58, 1615–1621 (2002).
15. Goldman, J. S. et al. Frontotemporal dementia: genetics and genetic counseling dilemmas. Neurologist 10, 227–234 (2004).
16. Mackenzie, I. R. A. et al. A harmonized classification system for FTLD-TDP pathology. Acta Neuropathol. 122, 111–113 (2011).
17. Lee, E. B. et al. Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. Acta Neuropathol. 134, 65–78 (2017).
18. Lashley, T., Rohrer, J. D., Mead, S. & Revesz, T. Review: an update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. Neuropathol. Appl. Neurobiol. 41, 858–881 (2015).
19. Hasegawa, M. et al. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Ann. Neurol. 64, 60–70 (2008).
20. Lu, R.-C., Wang, H., Tan, M.-S., Yu, J.-T. & Tan, L. TMEM106B and APOE polymorphisms interact to confer risk for late-onset Alzheimer's disease in Han Chinese. J. Neural Transm. 121, 283–287 (2014).
21. Rutherford, N. J. et al. TMEM106B risk variant is implicated in the pathologic presentation of Alzheimer disease. Neurology 79, 717–718 (2012).

22. Vass, R. et al. Risk genotypes at TMEM106B are associated with cognitive impairment in amyotrophic lateral sclerosis. Acta Neuropathol. 121, 373–380 (2011).
23. Lang, C. M. et al. Membrane orientation and subcellular localization of transmembrane protein 106B (TMEM106B), a major risk factor for frontotemporal lobar degeneration. J. Biol. Chem. 287, 19355–19365 (2012).
24. Brady, O. A., Zhou, X. & Hu, F. Regulated intramembrane proteolysis of the frontotemporal lobar degeneration risk factor, TMEM106B, by signal peptide peptidase-like 2a (SPPL2a). J. Biol. Chem. 289, 19670–19680 (2014).
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. Shi, Y. et al. Structure-based classification of tauopathies. Nature 598, 359–363 (2021).
27. Cruchaga, C. et al. Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. Arch. Neurol. 68, 581–586 (2011).
28. Nicholson, A. M. et al. TMEM106B p.T185S regulates TMEM106B protein levels: implications for frontotemporal dementia. J. Neurochem. 126, 781–791 (2013).
29. Inukai, Y. et al. Abnormal phosphorylation of Ser409/410 of TDP-43 in FTLD-U and ALS. FEBS Lett. 582, 2899–2904 (2008).
30. Laferrière, F. et al. TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. Nat. Neurosci. 22, 65–77 (2019).
31. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314, 130–133 (2006).

32. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's Disease. Annu. Rev. Neurosci. 34, 185–204 (2011).

33. Nonaka, T. et al. Prion-like properties of pathological TDP-43 aggregates from diseased brains. Cell Rep. 4, 124–134 (2013).

34. Arseni, D. et al. Structure of pathological TDP-43 filaments from ALS with FTLD. Nature 601, 139–143 (2022).

35. Schmidt, M. et al. Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. Nat. Commun. 10, 5008 (2019).

36. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151, 41–60 (2005).

37. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221 (2015).

38. Wagner, T. et al. Two particle-picking procedures for filamentous proteins: SPHIRE–crYOLO filament mode and SPHIRE-STRIPER. Acta Crystallogr. D 76, 613–620 (2020).

39. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).

40. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. J. Struct. Biol. 198, 163–176 (2017).

41. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).

42. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).

43. Terwilliger, T. C. Automated side-chain model building and sequence assignment by template matching. Acta Crystallogr. D 59, 45–49 (2003).

44. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D 74, 531–544 (2018).

45. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D 66, 12–21 (2010).

46. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315 (2018).

47. 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).

48. Thevis, M., Ogorzalek Loo, R. R. & Loo, J. A. In-gel derivatization of proteins for cysteine-specific cleavages and their analysis by mass spectrometry. J. Proteome Res. 2, 163–172 (2003).

49. McConnell, S. A. et al. Protein labeling via a specific lysine–isopeptide bond using the pilin polymerizing sortase from Corynebacterium diphtheriae. J. Am. Chem. Soc. 140, 8420–8423 (2018).

50. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906 (2007)

CHAPTER 2

Detection of proteopathic TDP-43 in neurodegenerative brain extracts.

Abstract

Pathological deposits of TDP-43 are found in numerous neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD), Alzheimer's disease (AD), various tauopathies, Lewy-body-related diseases, limbic-predominant age-related TDP-43 encephalopathy (LATE) and hippocampal sclerosis (HpScl). The role of TDP-43 aggregation in these distinctive diseases is not clear, and there lacks a proficient method for the study of patient-derived TDP-43 aggregates. Here, we employed a biosensor cell line for specific and efficient detection of proteopathic TDP-43 in sarkosyl-insoluble brain extracts. We confirmed the presence and examined the prevalence of proteopathic TDP-43 in ALS, FTLD-TDP, AD-TDP and HpScl cases, which correlate with the immunogenic detection of the phosphorylated C-terminal fragment of TDP-43. We observed that FTLD-TDP type C contains less seed-competent aggregates compared to FTLD-TDP types A and B. Concurrence of HpScl with FTLD-TDP increases the abundance of proteopathic TDP-43, however HpScl is not a modifying variable in the same way for AD-TDP. The TDP-43 biosensor cell line is a tool for high-throughput characterization of TDP-43 aggregates in brain extracts, and can identify fibril-enriched cases for subsequent biochemical and structural studies.

Introduction

Transactive response DNA-binding protein 43 (TDP-43) is a highly conserved, ubiquitously expressed heterogeneous nuclear ribonucleoprotein involved in essential functions of RNA metabolism, including splicing, transport and translation[1–5]. TDP-43 is primarily located in the nucleus, but can be shuttled into the cytoplasm to undergo functional and reversible aggregation in stress granules[6,7]. However in neurodegenerative diseases, TDP-43 is hyperphosphorylated,

cleaved, mislocalized to the cytoplasm, and undergoes pathogenic and irreversible aggregation[8,9]. TDP-43 was discovered as the major aggregated protein in ubiquitin-positive, tau- and α-synuclein-negative diseases FTLD and ALS[10,11]. Neuropathologists have since categorized FTLD with TDP-43 inclusions (FTLD-TDP) into subtypes A through E, based on the morphology and neuroanatomical location of TDP-43 deposits[12–14]. Once it was on the radar, TDP-43 inclusions has been observed in AD[15–18], various tauopathies[11,19–23], Lewy body-related diseases[24–26], LATE[27–30], and HpScl[15,31] (Fig. S1), typically correlating with exacerbated patient outcomes. For example, TDP-43 pathology, found in 50-70% of AD cases[32], is associated with hippocampal sclerosis, increased tau burden, greater brain atrophy, more severe and rapid cognitive decline[17]. The role of TDP-43 aggregation in pathogenesis is unknown, and the molecular mechanisms underlying different diseases may be unique. Another subject of interest is the interplay TDP-43 pathology may have with the "main" aggregated proteins with which it is comorbid, such as tau in AD or α-synuclein in Lewy body-related diseases.

Immunogold electron microscopy has revealed that TDP-43 forms fibrils in neuronal inclusions of a variety of neurodegenerative diseases[33,34]. Fibrillar TDP-43 extracted from ALS and FTLD-TDP brains were shown to exhibit prion-like properties, inducing aggregation which transmits between cells[35]. However, the proteopathic nature of TDP-43 aggregates in other diseases have not been tested. In addition, the isolation of TDP-43 fibrils from patient brains remains a challenge. Thus, the abundance of fibrillar TDP-43, in various disease cohorts, is a topic of interest. To address these questions, we used a FRET-based TDP-43 biosensor cell line, similar to those established for tau and α-synuclein[36,37], for the specific, sensitive, and high-throughput detection of seed-competent TDP-43 in brain extracts. We extracted sarkosyl-insoluble pellets

from ALS, FTLD-TDP and AD with TDP-43 pathology (AD-TDP) brains and screened them for proteopathic species of TDP-43.

Results

Sensitivity of biosensor cells to proteopathic TDP-43

Marc Diamond's group created the TDP-43 biosensor cells by stably expressing FRET pairs of fluorescently tagged TDP-43 low-complexity domain 262-414 in HEK293T cells. TDP-43 biosensors facilitate the spatiotemporal observation of TDP-43 aggregation in live cells: upon addition of proteopathic seeds, the aggregation of endogenous TDP-43 is visualized as fluorescent puncta ("seeding", Fig. 1a). The TDP-43 biosensors had been used to detect preformed fibrils of recombinant TDP-43[38], but they have not been used to study brain extracted samples. First, we tested the specificity of the TDP-43 biosensors using recombinant protein seeds. Fluorescent puncta in TDP-43 biosensors were induced by TDP-43 low-complexity domain fibrils (LCDf) but not the monomeric form of the protein (LCDm); TDP-43 full length (FL) and C-terminal fragment (CTF) constructs which formed amorphous aggregates as reported previously[39], also did not induce fluorescent puncta (Fig. 1b,c,f). As expected, recombinant tau dGAE paired helical filaments and α-synuclein fibrils did not generate fluorescent puncta in TDP-43 biosensors. Then, we confirmed the specificity of the TDP-43 biosensors using sarkosyl-insoluble brain extracts containing amyloid fibrils confirmed by negative stain TEM (Fig. 1d). Fluorescent puncta in TDP-43 biosensors were induced only by extracts from TDP-43-positive ALS cases, and not by extracts from TDP-43-negative tauopathy, synucleinopathy and non-neurodegenerative cases (Fig. 1e,g, Table 1). Taken together, the aggregation of endogenous TDP-43 in biosensor cells can only be induced by fibrillar TDP-43, and not other amyloids.

Detection of proteopathic TDP-43 in ALS extracts

We extracted sarkosyl-insoluble fractions from the frontal cortices of 33 sporadic ALS cases (Table 2), sonicated them to generate proteopathic seeds, and transfected them into TDP-43 biosensor cells. After 4 days, the majority of ALS extracts showed low to undetectable seeding; only four cases showed strong seeding (Fig. 2a, Fig. S2). Immunoblotting of patient extracts revealed that full-length TDP-43 is present in almost all sarkosyl-insoluble fractions; however the extracts containing phosphorylated Ser409/410 and truncated ~25 kDa fragments of TDP-43 exhibited robust seeding of biosensor cells (Fig. 2b,c). We integrated the western blot band intensities and correlated them with biosensor seeding puncta count for each case; the quantitative measurements support our qualitative observations (Fig. S3a). The strongest seeding case, SD96, contains "frequent" TDP-43 neuronal cytoplasmic inclusions (NCIs) in the frontal cortex, while many cases with weak seeding contained "sparse" or "moderate" TDP-43 NCIs (Table 2). The correlation of biosensor cell seeding results to neuropathological reports and western blot detection of phosphorylated species gives confidence that this tool is an effective screen for protepathic TDP-43 in brain extracts.

Detection of proteopathic TDP-43 in FTLD-TDP, AD-TDP and HpScl extracts

We extracted sarkosyl-insoluble fractions from the middle frontal gyrus of 26 FTLD-TDP cases (Table 3) from the same cohort of our previous study[40], and transfect sonicated seeds into TDP-43 biosensor cells. After 5 days, most FTLD-TDP type A and B extracts seeded proficiently; strikingly, FTLD-TDP type C extracts, with the exception of case M133, exhibited undetectable seeding (Fig. 3a). We also extracted sarkosyl-insoluble fractions from the occipitotemporal lobe and hippocampus of 38 AD cases, 36 of which have TDP-43 pathology, and 2 TDP-43-positive HpScl cases (Table 4), and transfect sonicated seeds into TDP-43 biosensor cells. After 5 days,

approximately a quarter of the TDP-43-positive AD extracts induced seeding; as expected, the TDP-43-negative AD extracts did not induce seeding (Fig. 4a). One of the two HpScl extracts seeded TDP-43 biosensor cells. Western blotting of FTLD-TDP, AD-TDP and HpScl extracts yielded observations similar to those made for ALS cases. Full-length TDP-43 is consistently found sarkosyl-insoluble fractions, however the accumulation of phospho-Ser409/410 C-terminal TDP-43 is more closely associated with seed-competent TDP-43 (Fig. 3b,c, Fig. 4b,c), supported by the correlation of integrated western blot band intensities with biosensor seeding puncta count (Fig. S3b,c). Interestingly, FTLD-TDP type C cases which did not seed still contained detectable amounts of phosphorylated C-terminal fragment of TDP-43.

Finally, we used immunogold labeling TEM to confirm the presence of TDP-43 fibrils in ALS, FTLD-TDP, AD-TDP and HpScl cases that demonstrated strong seeding, further validating detection by the biosensor cells (Fig. 5).

Patient variables affecting proteopathic TDP-43 content

We examined the TDP-43 biosensor seeding data by grouping FTLD-TDP, AD-TDP and HpScl cases based on pathological diagnoses and patient information (Fig. S4, Table 3 and 4). Generally, FTLD-TDP extracts seeded more consistently than AD-TDP extracts. Extracts from both TDP-43 type A or B inclusions were capable of biosensor cell seeding, in FTLD-TDP and AD-TDP cohorts independently (Fig. S4a). FTLD-TDP type C extracts seeded much less effectively than types A and B; there were no cases of AD-TDP with type C in the study. Interestingly, FTLD-TDP cases with HpScl are more likely to contain proteopathic TDP-43 compared to FTLD-TDP cases without HpScl, while the diagnosis of HpScl in AD was not a decisive factor (Fig. S4b). This may suggest that HpScl has different relationships to the pathogenesis of FTLD-TDP and AD-TDP, whether it is a driving factor or a secondary

phenomenon. Family history of disease and gender of the patient were not influential factors for whether their extracts seeded TDP-43 biosensor cells (Fig. S4c,d).

Discussion

TDP-43 was discovered as the major component of neuronal inclusions in FTLD-U and ALS, however its lack of reactivity to amyloid-binding dyes once raised the question of whether TDP-43 aggregates are amyloid[41,42]. Now, there is biochemical and structural evidence of TDP-43 amyloid fibrils extracted from patient brains[33–35,43,44]. Still, the contribution TDP-43 aggregation to pathogenesis is not well understood. Is it through the gain-of-toxicity of TDP-43 aggregates, or by the loss of TDP-43's RNA-binding function? If it is the former, which type(s) of pathological TDP-43 is the toxic species – ordered oligomers, amorphous aggregates, or fibrils? If it is the latter, TDP-43 in any pathological aggregated state would prohibit its normal function. The type of TDP-43 aggregation in various diseases may not be uniform, and there is the possibility of a mixture of species.

To help to characterize TDP-43 aggregates in neurodegenerative diseases, we employed a biosensor cell line that can differentiate between fibrillar and non-fibrillar recombinant TDP-43, and discriminate between TDP-43-positive and TDP-43-negative brain extracts. We added sarkosyl-insoluble extracts from ALS, FTLD-TDP, AD-TDP and HpScl cases to biosensor cells to assess their proteopathic content. We found the presence of seeding-competent TDP-43 correlated strongly with phosphorylated, C-terminal fragments of TDP-43, in particular for the ALS cohort. This is consistent with the idea that hyperphosphorylation of TDP-43 is critical for its conversion into pathological fibrils[8,9]. However, many TDP-43-positive cases from various diseases induced low to undetectable seeding in biosensor cells. Even some AD-TDP cases with

abundant phosphorylated, truncated TDP-43 did not seed well, such as M224 and M230. Why do cases with the same neuropathological diagnosis contain variable proteopathic content? One interpretation is that TDP-43 in neuropathological inclusions are not all fibrillar, consisting at least in part of non-proteopathic species, such as amorphous aggregates or larger oligomeric assemblies. Underlying factors may dictate the relative abundance of fibrillar and non-fibrillar TDP-43, which may account for variability within the same disease.

Sarkosyl-insoluble extracts from FTLD-TDP types A, B, D seeded biosensor cells fairly consistently, while type C extracts exhibited markedly low seeding potential. This trend is similar to observations made by Nonaka et al.[35], suggesting FTLD-TDP type C aggregates are less proteopathic. Multiple explanations can be considered. First, TDP-43 fibrils in type C could be scarce relative to other subtypes. Second, type C inclusions could contain TDP-43 fibrils that are inherently less effective at templating aggregation, which has intriguing implications for the disease mechanism of this subtype. FTLD-TDP type C is characterized by long dystrophic neurites and the absence of NCIs, which are enriched in both FTLD-TDP type A and B[14]. The third and technical explanation is that type C TDP-43 fibrils could be incompatible with the TDP-43 construct expressed, and their proteopathic competence was not measured in this biosensor cell system. Recently, distinct TDP-43 fibril structures have been determined by cryogenic-electron microscopy, from ALS/FTLD-TDP type B and FTLD-TDP type A patients[43,44]. M133, the only FTLD-TDP type C case which showed strong seeding, is a priority candidate for cryo-EM study for which the fibril structure of the subtype is unknown.

Our UCSD collaborators are completing the neuropathological analyses of ALS cases, which will provide details regarding abundance of TDP-43 NCIs, site of onset and motor neuron predominancy (Table 2). With the information, we can better assess and rationalize the TDP-43

biosensor seeding data. For example, low protepathic content was observed for many cases. An explanation is those cases may have lower motor neuron predominancy, that is, greater TDP-43 burden in the brain stem and spinal cord compared to the frontal and motor cortex of the upper motor neurons (which are the tissues we extracted from). It would be intriguing to extract and seed samples from the lower motor neuron tissues of the same cases.

Methods

Recombinant aggregates and amyloid fibrils

Recombinant constructs of TDP-43 were expressed and purified as described[39]. Full length (1-414, FL), C-terminal fragment (208-414, CTF) and low complexity domain (274-414, LCD) TDP-43 proteins were diluted to 50 μM monomer equivalent concentration in 20mM Tris-HCl pH 8.0, 150 mM NaCl, 10 μM DTT, and shaken at 900 r.p.m. (Torrey Pines Scientific) at 37 °C for 3 days. Tau dGAE fibrils were prepared as described[45]. dGAE fibrils were formed at 6 mg/mL monomer equivalent concentration, in 10 mM sodium phosphate buffer with 10 mM DTT and 100 mM $MgCl_2$ pH 7.2, with continuous single orbital shaking at 200 r.p.m. (FLUOstar Omega, BMG Labtech) at 37 °C for 20 hours. α-synuclein fibrils were formed at 40 μM monomer equivalent concentration, in PBS buffer pH 7.4, with double orbital shaking at 600 r.p.m. (FLUOstar Omega, BMG Labtech) for 10 seconds every 10 minutes (otherwise quiescent) at 37 °C for 72 hours.

Post-mortem brains

Post-mortem brain tissues from patients neuropathologically diagnosed with ALS were provided by the University of California San Diego (UCSD) ALS tissue repository, which follows HIPAA-compliant procedures approved by UCSD Institutional Review Board (CA IRB no.

120056). All patients provided informed consent, were de-identified, and received no compensation. Post-mortem brain tissues from patients neuropathologically diagnosed with FTLD-TDP, AD-TDP and HpScl were provided by the Brain Bank for Neurodegenerative Disorders at Mayo Clinic Florida, which operates under protocols approved by the Mayo Clinic Institutional Review Board. Autopsies were performed after consent by the next of kin or someone with legal authority to grant permission. Mayo Clinic Florida also provided PART, VaD and ND brain tissues. Harry Vinters and Christopher K. Williams from UCLA provided the PD, DLB and MSA brain tissues.

Patient-extracted amyloid fibrils

ALS brain tissues were diced finely and transferred into 40 mL/g tissue of extraction buffer: 10 mM Tris-HCl pH 7.4, 0.8 M NaCl, 10% (w/v) sucrose and 1 mM EGTA. Tissues were homogenized by hand using Biomasher II disposable micro-tube homogenizer (Research Products International). 2% (w/v) of *N*-lauroyl-sarcosine (sarkosyl, Sigma) was added to the homogenate, and the mixture was incubated with 200 r.p.m. shaking at 37 °C for 45 minutes. The sample was centrifuged at 10,000 *g* at 25 °C for 10 minutes. The supernatant was taken and centrifuged at 166,000 *g* at 25 °C for 30 minutes. The sarkosyl-insoluble pellet was resuspended in 200 μL/g tissue of 20 mM Tris-HCl pH 7.4, 100 mM NaCl.

Amyloid fibril extraction from FTLD-TDP and AD-TDP brain tissues was performed with a similar protocol with two changes. First, the extraction buffer used was 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, supplemented with 1:100 (v/v) Halt protease inhibitor (Thermo Scientific). Second, the final centrifugation speed was 400,000 *g* instead of 166,000 *g*. The protocols were otherwise identical.

Negative stain transmission electron microscopy

Carbon-coated formvar support films mounted on copper grids with 400-mesh (Ted Pella) were glow-discharged for 30 seconds. 3 µL of recombinant aggregates (undiluted) or patient-extracts (20X diluted) was applied to the grids and incubated for 2 minutes, then excess solution was blotted off using filter paper. Grids were stained with 3 µl of 2% uranyl acetate (Electron Microscopy Sciences) for 2 minutes, then washed with an additional 3 µl of 2% uranyl acetate and air-dried. The grids were imaged using a Tecnai T12 transmission electron microscope (FEI).

TDP-43 biosensor cells

HEK293T cell lines stably expressing TDP-43 262-414 fused to mClover3 and mRuby3 were obtained from the Marc Diamond laboratory[38]. Cells were maintained in DMEM (Life Technologies, cat. 11965092) supplemented with 10% (v/v) FBS (Life Technologies, cat. A3160401), 1% penicillin/streptomycin (Life Technologies, cat. 15140122) and 1% Glutamax (Life Technologies, cat. 35050061) at 37 °C and 5% CO_2 in a humidified incubator.

For seeding experiments, 1 µL of recombinant aggregates or 2 µL of patient-extracts were diluted in 18 µL of OptiMEM (Life Technologies, cat. 31985062) and sonicated in a cuphorn water bath for 5 minutes, then mixed with 1 µL of Lipofectamine 2000 (Life Technologies, cat. 11668019) in 19 µL of OptiMEM. After 20 minutes, 10 µL of the fibril mixture were added to 100 µL of TDP-43 biosensor cells. The number of seeded aggregates was determined by imaging the entire well of a 96-well plate in triplicate using a Celigo Image Cytometer (Nexcelom Bioscience) in the green fluorescence channel.

The number of fluorescent puncta in an image was determined using an ImageJ 2.3.0[46] script, which subtracts the background fluorescence from unseeded cells and counts the number of puncta with fluorescence peaks above the background using the built-in Particle Analyzer. The

number of puncta was normalized to the confluence of each well, and plots were generated by calculating the average and standard deviations from triplicate measurements. Quantification of the number of puncta was done on day 1, 3, 4 and 5 after seeding; the data with optimal signal and cell confluence was shown. Fluorescence microscopy images are taken on day 3 on an AXIO Observer D1 Inverted Fluorescence Motorized Phase Contrast Microscope (Zeiss) at 10X magnification in the EGFP channel.

Western blotting

1 µL of patient-extracts was mixed with SDS–PAGE loading dye containing 8 M urea and 1 M β-mercaptoethanol, sonicated in a cuphorn ice water bath for 10 minutes, then boiled at 95 °C for 10 minutes. Samples were loaded onto Bolt Bis-Tris Plus Mini Protein Gels, 4–12%, 1.0 mm, 15-well (Invitrogen) and electrophoresis was performed at 200 V for 30 minutes. Proteins were transferred onto 0.2 µm nitrocellulose membranes (Bio-Rad) by application of 100 V for 1 hour, in transfer buffer consisting of 25 mM Tris pH 8.3, 192 mM glycine, 20% (w/v) methanol. Membranes were incubated with gentle rocking in 5% (w/v) Blocking-Grade Blocker (milk, Bio-Rad) in TBS with 0.1% (v/v) Tween20 (TBST) for 1 hour. Membranes were incubated with TDP-43 monoclonal antibody (Proteintech, cat. 60019-2-IG, lot 10023187) diluted 1:1000, or anti TAR DNA-binding protein 43 phospho-Ser409/410 monoclonal antibody (Cosmo Bio, cat. CAC-TIP-PTD-M01A, lot 11-9-19-33-1) diluted 1:500 in 2% milk in TBST for 1 hour. Membranes were washed three times in TBST with gentle rocking for 5 minutes. Membranes were incubated with goat anti-mouse IgG secondary antibody HRP (Invitrogen, cat. 31430, lot XH3637J2) diluted 1:4000 in 2% milk in TBST for 1 hour. Membranes were washed three times in TBST with gentle rocking for 5 min. Pierce ECL Plus Western Blotting Substrate (Thermo Scientific) was applied to membranes. Membranes were imaged using an Azure 600 (Azure

Biosystems). Western blot band intensities were quantified in ImageJ[46]. The ~43 kDa and ~25 kDa (phospho-Ser409/410 blot only) bands corresponding to full length and C-terminal fragment TDP-43 were boxed and measured for integrated intensities. The relative peak intensities are plotted for each sample.

Immunogold labeling

Carbon-coated formvar support films mounted on copper grids with 400-mesh (Ted Pella) were glow-discharged for 30 seconds. 3 µL of patient-extracts (20X diluted) was applied to the grids and incubated for 3 minutes, then excess solution was blotted off using filter paper. PBS pH 7.4 with 0.1% w/v gelatin (blocking buffer) was applied to the grids and incubated for 10 minutes, then excess solution was blotted off. Anti TAR DNA-binding protein 43 phospho-Ser409/410 monoclonal antibody (Cosmo Bio, cat. CAC-TIP-PTD-M01A, lot 11-9-19-33-1) diluted 1:100 in blocking buffer was applied to the grids and incubated for 2 hours, then excess solution was blotted off. Grids were washed five times with blocking buffer; excess solution was blotted off between each wash. 6 nm Colloidal Gold AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch Laboratories, cat. 115-195-146, lot 163780) diluted 1:20 in blocking buffer was applied to the grids and incubated for 30 minutes, then excess solution was blotted off. Grids were washed five times with water; excess solution was blotted off between each wash. Grids were stained with 3 µl of 2% uranyl acetate (Electron Microscopy Sciences) for 2 minutes, then washed with an additional 3 µl of 2% uranyl acetate and air-dried. The grids were imaged using a Tecnai T12 transmission electron microscope (FEI).

Acknowledgements

We thank the donors and their families, without whom this work would not have been possible. We thank Marc I. Diamond and Jaime Vaquer-Alicea from UT Southwestern for creating and sharing the TDP-43 biosensor cells. We thank Harry V. Vinters and Christopher K. Williams from UCLA for providing the PD, DLB and MSA brains samples. We acknowledge funding support from NIH AG048120 and NIH AG070895. Graphics were created with Biorender.com.

Author contributions

Y.X.J., X.C. and D.S.E conceived of and designed the project. Y.X.J. and H.H. extracted amyloid fibrils from patient brains and performed western blotting. Y.X.J. and L.L. prepared recombinant aggregates. X.C. performed TDP-43 biosensor cell seeding assays. Y.X.J. and X.C. performed immunogold labeling and negative stain TEM. J.R., A.Hiniker and A.Hermann prepared frozen brain samples of ALS cases. M.D. and D.W.D. prepared frozen brain samples of FTLD-TDP and AD-TDP cases. Y.X.J. and D.S.E. prepared the manuscript with contributions from all authors.

Competing interests

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

Figures and Tables

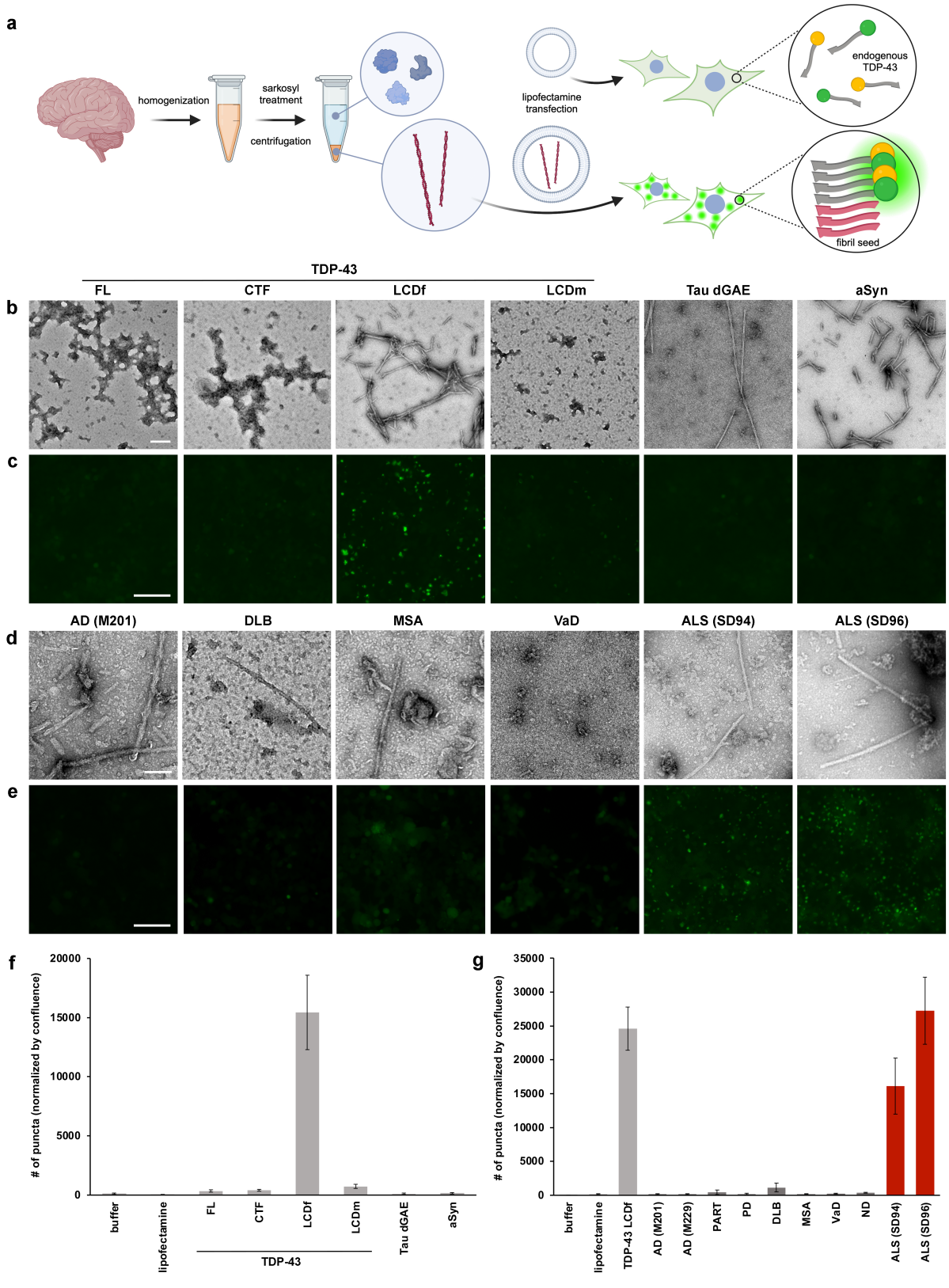
a
homogenization
sarkosyl treatment
centrifugation
lipofectamine transfection
endogenous TDP-43
fibril seed
TDP-43
FL
CTF
LCDf
LCDm
Tau dGAE
aSyn
b
c
AD (M201)
DLB
MSA
VaD
ALS (SD94)
ALS (SD96)
d
e
f
# of puncta (normalized by confluence)
20000
15000
10000
5000
0
buffer
lipofectamine
FL
CTF
LCDf
LCDm
Tau dGAE
aSyn
TDP-43
g
35000
30000
25000
20000
15000
10000
5000
0
TDP-43 LCDf
AD (M201)
AD (M229)
PART
PD
DLB
MSA
VaD
ND
ALS (SD94)
ALS (SD96)

Figure 1: TDP-43 biosensor cells are sensitive to fibrillar aggregates of TDP-43. a, Graphical summary of TDP-43 biosensor cell seeding assay, using recombinant and patient-extracted amyloid fibrils. Exogenous fibril seeds induce the aggregation of endogenously expressed, fluorescent-labeled TDP-43. **b**, Negative stain TEM images of recombinant aggregates used to seed TDP-43 biosensor cells. Full length (1-414, FL) and C-terminal fragment (208-414, CTF) TDP-43 proteins formed amorphous aggregates after shaking at 37 °C for 3 days, while the low complexity domain (274-414, LCD) formed fibrils (LCDf). LCD monomers (LCDm), not subjected to shaking, were also tested. Tau dGAE paired helical filaments and α-synuclein fibrils served as non-TDP-43 fibril controls. Scale bar 200 nm. **c**, Florescence microscopy images of TDP-43 biosensor cells 3 days after seeding with recombinant aggregates. Scale bar 100 μm. Fluorescent puncta formed only with TDP-43 LCD fibril seeding. **d**, Negative stain TEM images of amyloid fibrils in sarkosyl-insoluble patient extracts from various TDP-43-negative tauopathy, synucleinopathy, non-neurodegenerative cases, as well as TDP-43-positive amyotrophic lateral sclerosis (ALS) cases used to seed TDP-43 biosensor cells. Cases include Alzheimer's disease (AD), primary age-related tauopathy (PART), Parkinson's disease (PD), dementia with Lewy bodies (DLB), multiple systems atrophy (MSA), vascular dementia (VaD) and non-dementia (ND). Scale bar 100 nm. **e**, Fluorescence microscopy images of TDP-43 biosensor cells 3 days after seeding with patient-extracted amyloid fibrils. Scale bar 100 μm. Fluorescent puncta were observed only in TDP-43-positive ALS cases. **f**,**g**, Quantification of fluorescent puncta, normalized by cell confluence, of TDP-43 biosensor cells seeded by recombinant aggregates after 5 days (**f**) and sarkosyl-insoluble patient extracted amyloid fibrils after 4 days (**g**).

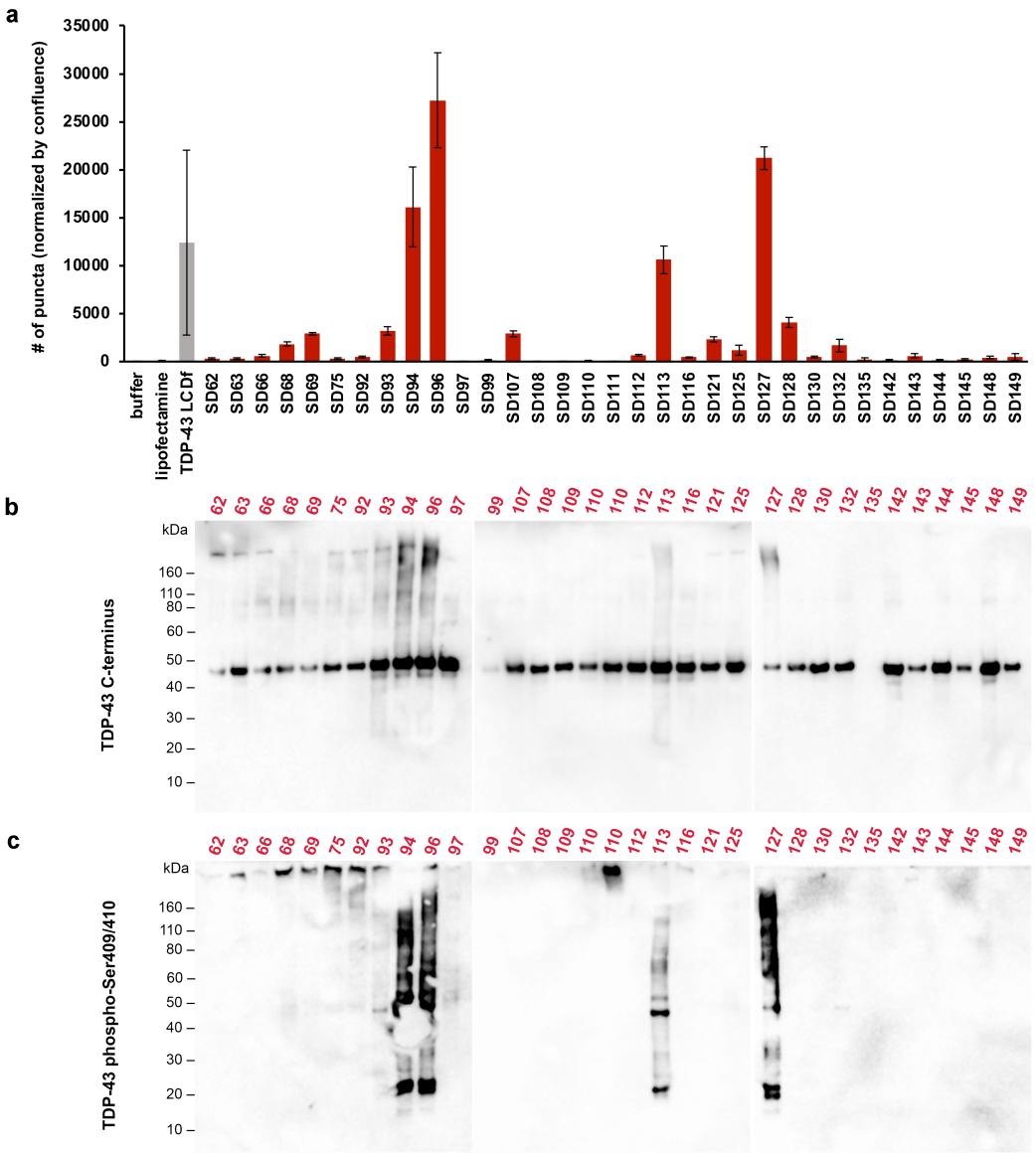

Figure 2: TDP-43 biosensor cell seeding and immunoblotting of ALS patient extracts. a, Quantification of fluorescent puncta, normalized by cell confluence, of TDP-43 biosensor cells seeded by ALS sarkosyl-insoluble extracts after 4 days. **b**,**c**, Western blots of ALS sarkosyl-insoluble extracts probed with C-terminus TDP-43 (**b**) and phospho-Ser409/410 TDP-43 (**c**) antibodies.

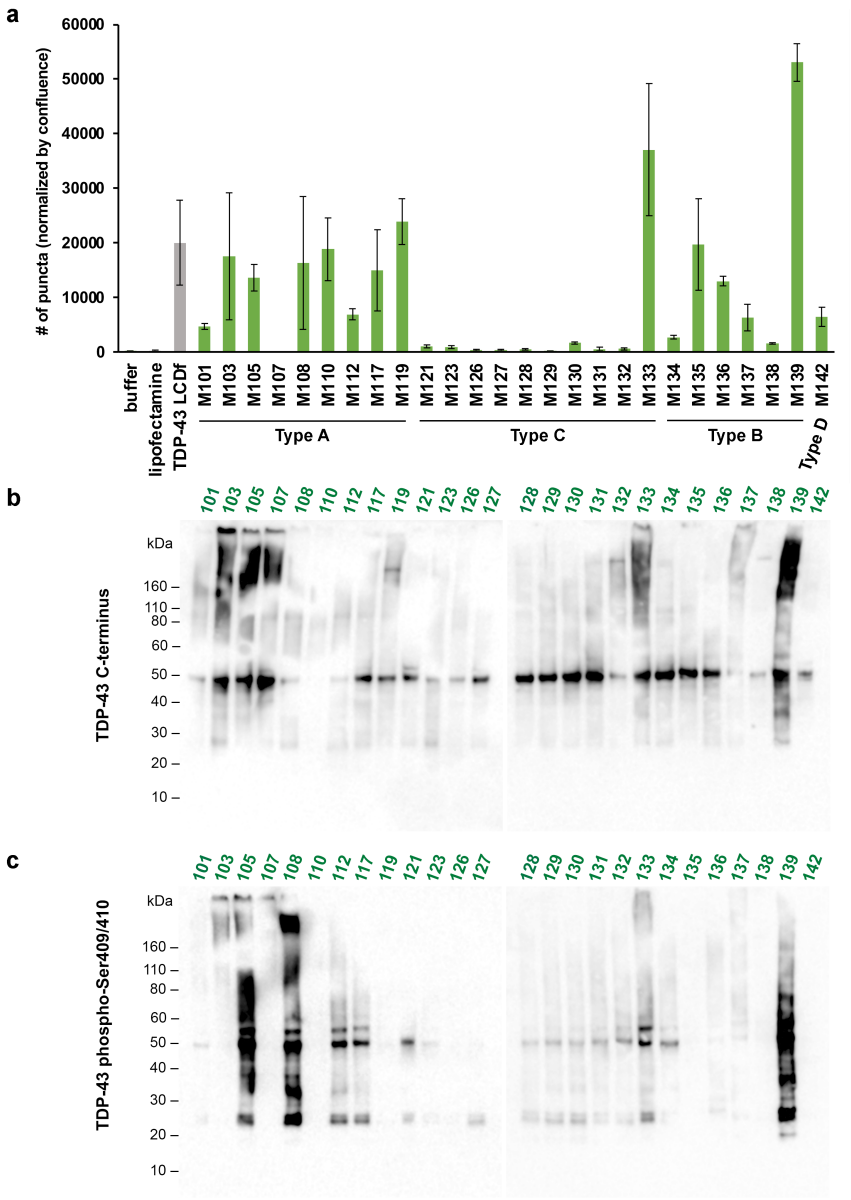

Figure 3: TDP-43 biosensor cell seeding and immunoblotting of FTLD-TDP patient extracts. a, Quantification of fluorescent puncta, normalized by cell confluence, of TDP-43 biosensor cells seeded by FTLD-TDP sarkosyl-insoluble extracts after 5 days. **b**,**c**, Western blots of FTLD-TDP sarkosyl-insoluble extracts probed with C-terminus TDP-43 (**b**) and phospho-Ser409/410 TDP-43 (**c**) antibodies.

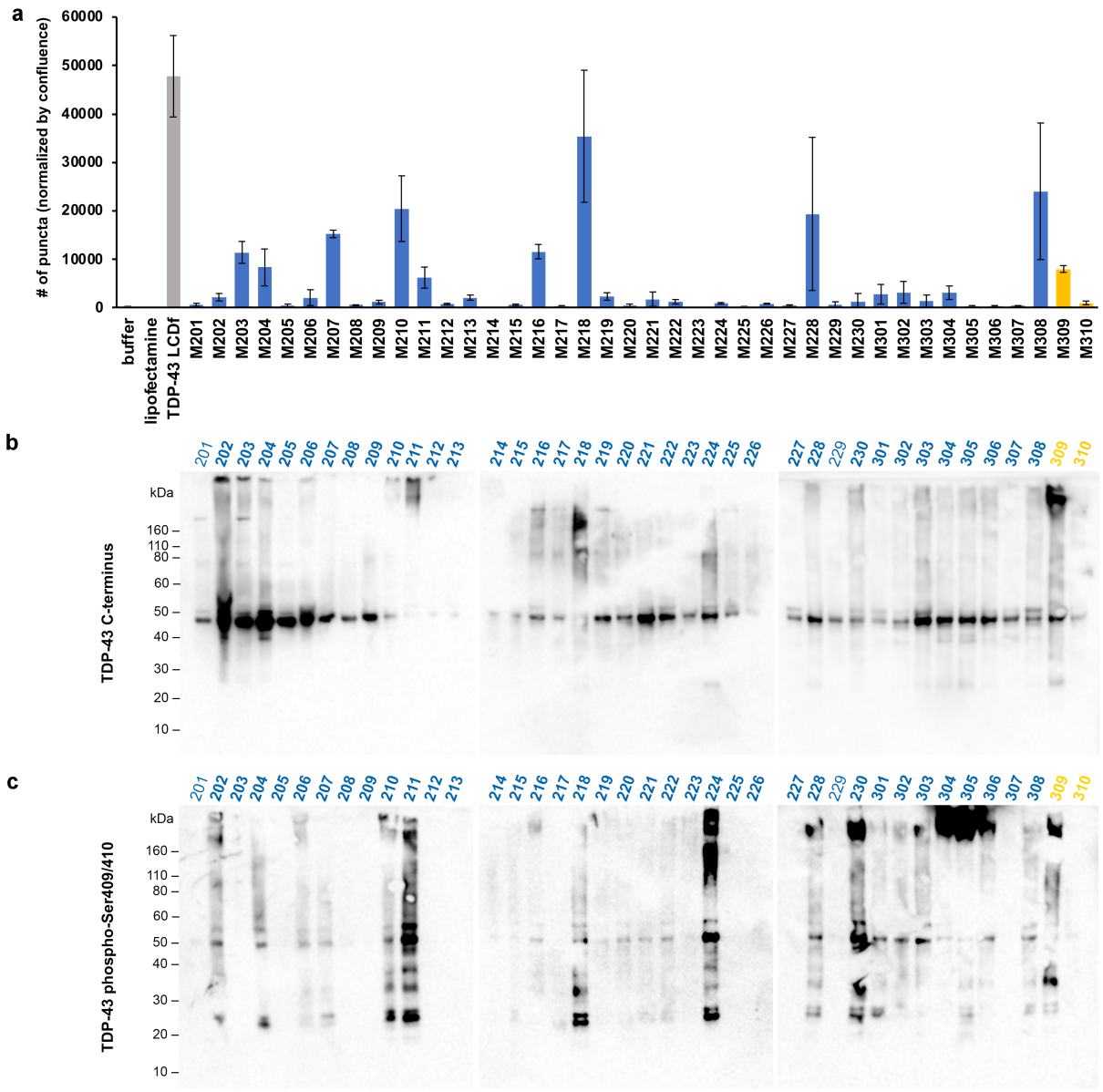

Figure 4: TDP-43 biosensor cell seeding and immunoblotting of AD-TDP and HpScl patient extracts. a, Quantification of fluorescent puncta, normalized by cell confluence, of TDP-43 biosensor cells seeded by AD-TDP (blue) and HpScl (yellow) sarkosyl-insoluble extracts after 5 days. **b**,**c**, Western blots of AD-TDP sarkosyl-insoluble extracts probed with C-terminus TDP-43 (**b**) and phospho-Ser409/410 TDP-43 (**c**) antibodies.

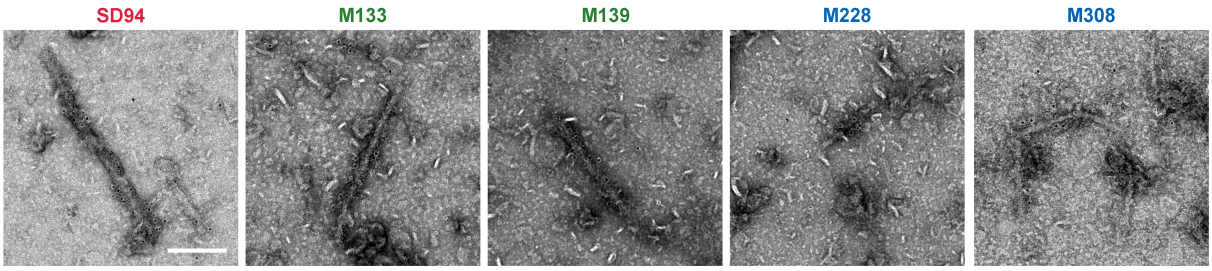

Figure 5: Immunogold labeling of TDP-43 amyloid fibrils extracted from various neurodegenerative diseases. Amyloid fibrils from sarkosyl-insoluble extracts from ALS (SD94), FTLD-TDP type C (M133) and type B (M139), AD-TDP (M228, M308) were immunolabeled with phospho-Ser409/410 TDP-43 antibodies. Scale bar 200 nm.

Patient ID	PathDx	Braak stage	Gender	Age	FHx	Brain region
M201	AD	6	male	84	false	OT
M229	AD	5	female	77	false	OT
M1010	PART	1	male	53	false	MF
	PD					
	DLB					
	MSA					
M1004	VaD	2	male	62	true	MF
M1007	ND	0	female	85	false	MF
SD94	sALS		male	62	false	FC
SD96	sALS		female	67	false	FC

Table 1: Control patient information. PathDx, pathological diagnosis; AD, Alzheimer's disease; PART, primary age-related tauopathy; PD, Parkinson's disease; DLB, dementia with Lewy bodies; MSA, multiple systems atrophy; VaD, vascular dementia; ND, non-dementia; sALS, sporadic amyotrophic lateral sclerosis; FHx, family history; OT, occipitotemporal lobe;

MF, middle frontal gyrus; FC, frontal cortex. sALS cases are from the ALS cohort (Table 2) and TDP-43-negative AD cases are from the AD-TDP cohort (Table 4).

Patient ID	PathDx	Gender	Age	Site of onset	MN predom	TDP-43 NCIs in FC	Brain region
SD62	sALS	male	72	arm	L>U		FC
SD63	sALS	male	52	arm	L>U		FC
SD66	sALS	male	46	arm	L>U		FC
SD68	sALS	female	72	leg	U=L		FC
SD69	sALS	male	60	generalized	L>>U		FC
SD75	sALS	male	52	arm	U=L		FC
SD92	sALS	female	61	leg	U=L		FC
SD93	sALS	male	71	bulbar	U>L	sparse	FC
SD94	sALS	male	62	bulbar	U=L	sparse	FC
SD96	sALS	female	67	bulbar	U>L	frequent	FC
SD97	sALS	female	63	arm	U=L	moderate	FC
SD99	sALS	female	70	bulbar	U>>L		FC
SD107	sALS	male	74	trunk	L>>U	moderate	FC
SD108	sALS	male	69	leg	L>>U		FC
SD109	sALS	male	49	leg	L>>U		FC
SD110	sALS	male	53	unknown	unknown		FC
SD111	sALS	female	59	leg	U=L		FC
SD112	sALS	female	54	leg	L>>U		FC
SD113	sALS	male	66	leg	U=L		FC
SD116	sALS	male	72	leg	L>U		FC
SD121	sALS	male	67	R. arm	U=L		FC
SD125	sALS	male	67	L. arm	L>U		FC
SD127	sALS	male	67	R. arm, respiratory	L>>U		FC

SD128	sALS	female	69	R. foot	L>U		FC
SD130	sALS	male	63	leg	L>U		FC
SD132	sALS	female	64	R. arm	U=L	sparse	FC
SD135	sALS	male	77	respiratory, thoraxic	L>U	none	FC
SD142	sALS	male	79	leg	U>>L	sparse	FC
SD143	sALS	male	67	bulbar	U>L	none	FC
SD144	sALS	male	57	leg	L>>U		FC
SD145	sALS	female	51	leg	U=L		FC
SD148	sALS	male	71	limb	U=L		FC
SD149	sALS	male	76	R. arm	L>U		FC

Table 2: ALS patient information. PathDx, pathological diagnosis; sALS, sporadic amyotrophic lateral sclerosis; MN, motor neuron; U, upper motor neuron; L, lower motor neuron; NCIs, neuronal cytoplasmic inclusions; FC, frontal cortex; MC, motor cortex.

Patient ID	**PathDx**	**HpScl**	**Braak stage**	**FHx**	**Gender**	**Age**	**Brain region**
M101	FTLD-TDP A	–	2	false	male	66	MF
M103	FTLD-TDP A	+	3	false	male	60	MF
M105	FTLD-TDP A	+	2.5	true	male	71	MF
M107	FTLD-TDP A	+	2	true	male	60	MF
M108	FTLD-TDP A	+	2.5	true	female	84	MF
M110	FTLD-TDP A	+	1.5	false	female	78	MF
M112	FTLD-TDP A	+	0	true	male	63	MF
M117	FTLD-TDP A	+	0	false	male	86	MF
M119	FTLD-TDP A	+	2	true	female	64	MF
M121	FTLD-TDP C	–	0	false	female	71	MF
M123	FTLD-TDP C	–	0	false	male	75	MF

M126	FTLD-TDP C	–	2	false	male	65	MF
M127	FTLD-TDP C	–	3	true	male	66	MF
M128	FTLD-TDP C	–	2	false	male	75	MF
M129	FTLD-TDP C	–	1	false	male	65	MF
M130	FTLD-TDP C	+	1	true	male	78	MF
M131	FTLD-TDP C	–	3	false	female	69	MF
M132	FTLD-TDP C	–	0	false	male	66	MF
M133	FTLD-TDP C	+	2	true	female	69	MF
M134	FTLD-TDP C	–	3	true	female	82	MF
M135	FTLD-TDP B	–	0	false	female	50	MF
M136	FTLD-TDP B	–	2	false	male	69	MF
M137	FTLD-TDP B	–	2.5	true	male	70	MF
M138	FTLD-TDP B	–	3	false	female	76	MF
M139	FTLD-TDP B	+	0	false	female	62	MF
M142	FTLD-TDP D	–	3	true	female	64	MF

Table 3: FTLD-TDP patient information. PathDx, pathological diagnosis; HpScl, hippocampal sclerosis; FHx, family history; MF, middle frontal gyrus.

Patient ID	**PathDx**	**TDP-43 type**	**HpScl**	**Braak stage**	**FHx**	**Gender**	**Age**	**Brain region**
M201	AD	negative	–	6	false	male	84	OT
M202	AD	A	–	5.5	false	female	91	OT
M203	AD	A	–	6	false	female	80	OT
M204	AD	A	–	5.5	true	male	76	OT
M205	AD	B	–	6	false	female	75	OT
M206	AD	B	–	5.5	false	male	73	OT
M207	AD	B	–	5	false	male	85	OT
M208	AD	B	–	6	true	male	79	OT
M209	AD	B	–	6	false	female	74	OT

M210	AD	A	–	6	false	female	69	OT
M211	AD	B	–	6	true	female	57	OT
M212	AD	A	–	5	false	male	87	OT
M213	AD	A	+	5	false	female	85	OT
M214	AD	A	+	5.5	false	female	85	OT
M215	AD	B	–	5.5	false	female	88	OT
M216	AD	A	+	4.5	true	male	87	OT
M217	AD	A	+	4.5	false	female	97	OT
M218	AD	A	+	4.5	false	male	94	OT
M219	AD	B	+	4	false	male	92	OT
M220	AD	B	–	6	false	female	79	OT
M221	AD	A	+	6	true	male	80	OT
M222	AD	A	+	5	true	male	84	OT
M223	AD	A	+	6	false	female	86	OT
M224	AD	A	+	5.5	true	male	86	OT
M225	AD	B	+	6	false	female	96	OT
M226	AD	B	–	6	false	female	79	OT
M227	AD	B	–	6	true	male	66	OT
M228	AD	A	–	5.5	true	female	91	OT
M229	AD	negative	–	5	false	female	77	OT
M230	AD	B	–	6	false	female	86	OT
M301	AD	A	+	5	false	female	100	Hip
M302	AD	A	+	6	false	female	99	Hip
M303	AD	A	+	5	false	female	88	Hip
M304	AD	A	+	6	true	female	87	Hip
M305	AD	A	+	5.5	false	female	89	Hip
M306	AD	A	+	6	false	female	86	Hip
M307	AD	B	+	6	true	female	96	Hip
M308	AD	B	+	6	true	female	88	Hip
M309	HpScl	A	+	2.5	true	female	95	Hip

M310	HpScl	A	–	2	false	female	95	Hip

Table 4: AD-TDP and HpScl patient information. PathDx, pathological diagnosis; HpScl, hippocampal sclerosis; FHx, family history; OT, occipitotemporal lobe; Hip, hippocampus.

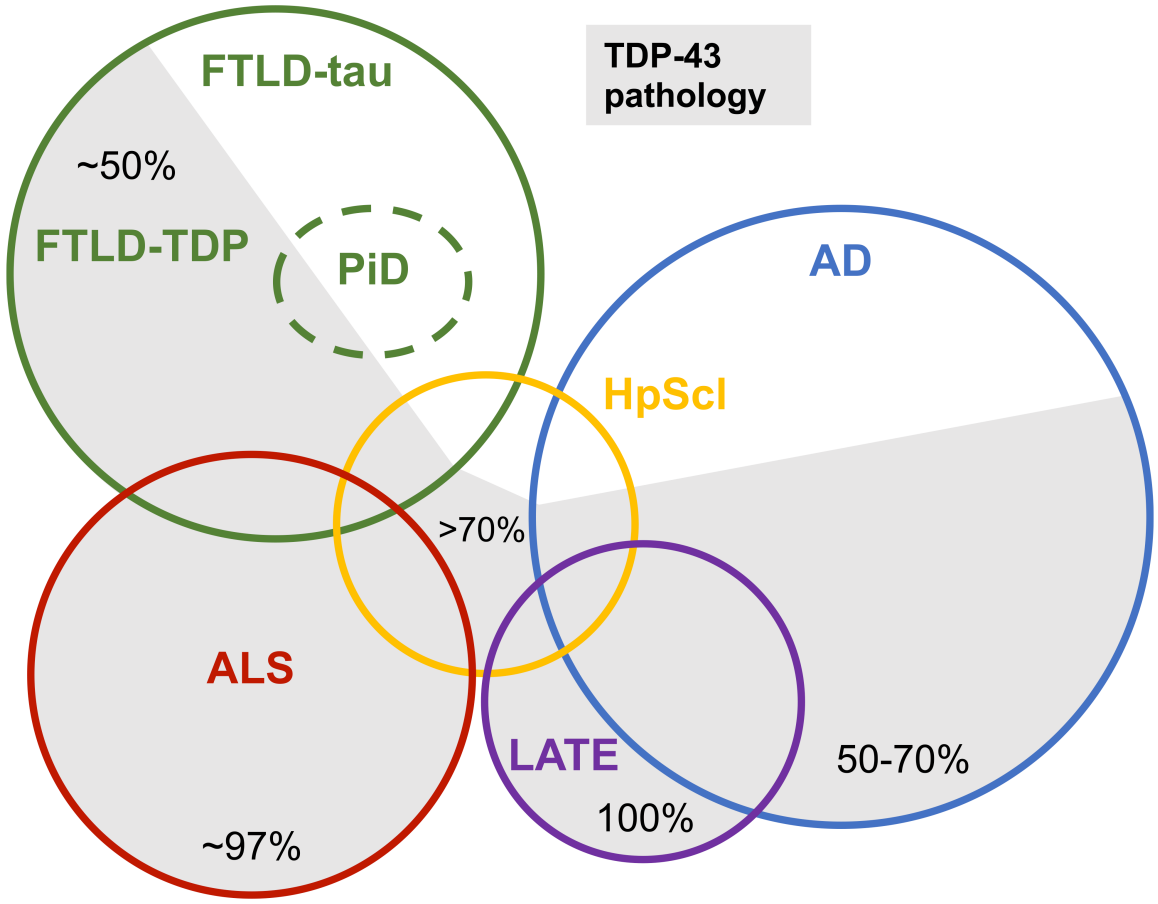

Figure S1: TDP-43 pathology in neurodegenerative diseases.

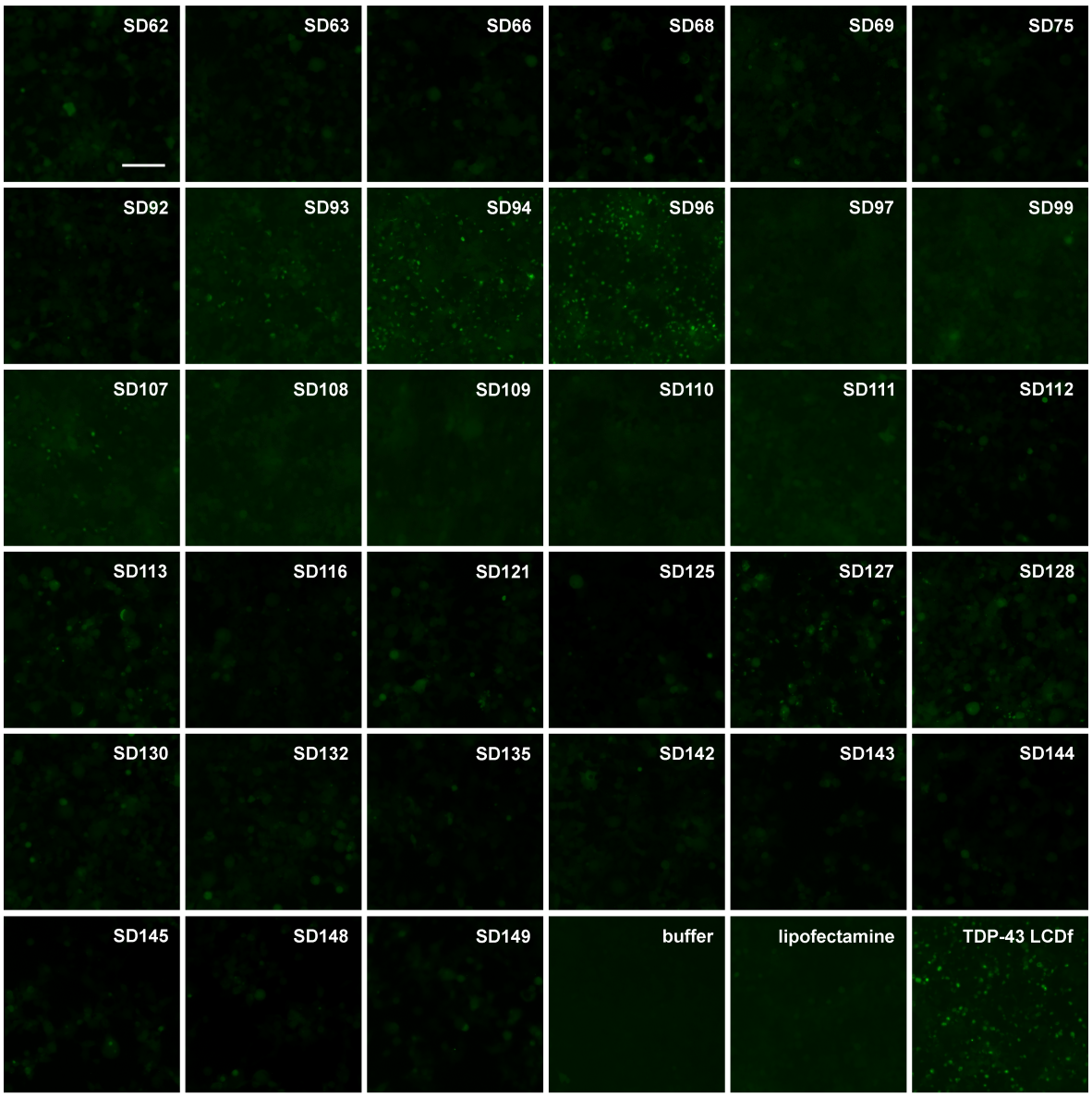

Figure S2: Florescence microscopy images of TDP-43 biosensor cells 3 days after seeding with ALS patient-extracts. Scale bar 100 µm.

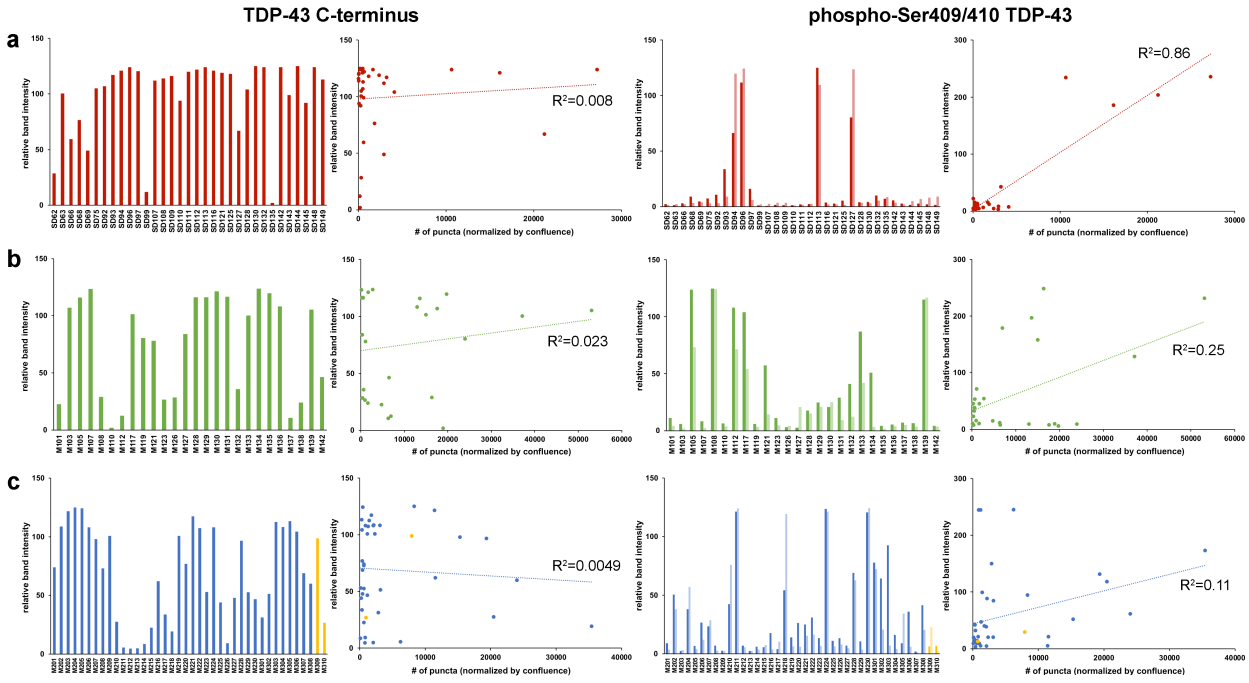

Figure S3: Quantification of western blots bands and correlation with TDP-43 biosensor cell seeding. Western blot band intensity quantification and correlation with fluorescent puncta in TDP-43 biosensor cell seeding are plotted for ALS (**a**), FTLD-TDP (**b**), AD-TDP and HpScl (**c**). The TDP-43 C-terminus western blot ~43 kDa band and the phospho-Ser409/410 TDP-43 western blot ~43 kDa (darker bars) and ~25 kDa (lighter bars) bands were quantified using ImageJ by relative peak intensity. The correlation graph for phospho-Ser409/410 plots the sum of ~43kDa and ~25 kDa band intensities (y-axis) against the number of fluorescent puncta from TDP-43 biosensor cell seeding (x-axis). Linear trendlines and R^2 values are shown for each correlation graph.

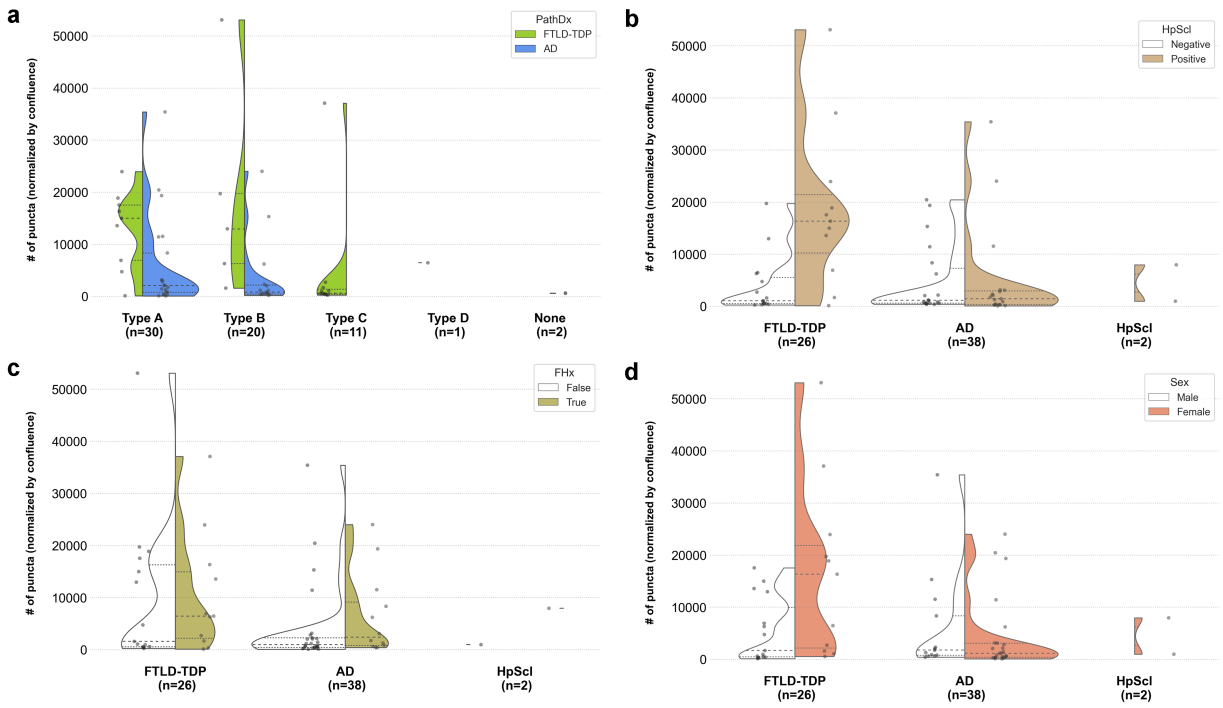

Figure S4: Violin plots of FTLD-TDP, AD-TDP and HpScl biosensor cell seeding data sorted by neuropathological diagnoses and patient information. Quantification of fluorescent puncta in TDP-43 biosensor cell seeding by FTLD-TDP, AD-TDP and HpScl extracts sorted by TDP-43 type (**a**), hippocampal sclerosis (**b**), family history (**c**) and gender (**d**).

References

1. Buratti, E. *et al.* Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* **20**, 1774–1784 (2001).
2. Buratti, E. & Baralle, F. E. Characterization and Functional Implications of the RNA Binding Properties of Nuclear Factor TDP-43, a Novel Splicing Regulator of CFTR Exon 9. *Journal of Biological Chemistry* **276**, 36337–36343 (2001).
3. Buratti, E. & Baralle, F. E. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Frontiers in Bioscience* **13**, 867–878 (2008).

4. Ayala, Y. M. *et al.* Structural determinants of the cellular localization and shuttling of TDP-43. *J Cell Sci* **121**, 3778–3785 (2008).

5. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* **30**, 277–288 (2011).

6. Colombrita, C. *et al.* TDP-43 is recruited to stress granules in conditions of oxidative insult. *J Neurochem* **111**, 1051–1061 (2009).

7. 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).

8. Inukai, Y. *et al.* Abnormal phosphorylation of Ser409/410 of TDP-43 in FTLD-U and ALS. *FEBS Lett* **582**, 2899–2904 (2008).

9. Hasegawa, M. *et al.* Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol* **64**, 60–70 (2008).

10. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).

11. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun* **351**, 602–611 (2006).

12. Mackenzie, I. R. A. *et al.* Heterogeneity of ubiquitin pathology in frontotemporal lobar degeneration: Classification and relation to clinical phenotype. *Acta Neuropathol* **112**, 539–549 (2006).

13. Sampathu, D. M. *et al.* Pathological Heterogeneity of Frontotemporal Lobar Degeneration with Ubiquitin-Positive Inclusions Delineated by Ubiquitin Immunohistochemistry and Novel Monoclonal Antibodies. *Am J Pathol* **169**, 1343–1352 (2006).

14. Mackenzie, I. R. A. *et al.* A harmonized classification system for FTLD-TDP pathology. *Acta neuropathologica* vol. 122 111–113 Preprint at https://doi.org/10.1007/s00401-011-0845-8 (2011).

15. Amador-Ortiz, C. *et al.* TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann Neurol* **61**, 435–445 (2007).

16. Uryu, K. *et al.* Concomitant TAR-DNA-binding protein 43 pathology is present in Alzheimer disease and corticobasal degeneration but not in other tauopathies. *J Neuropathol Exp Neurol* **67**, 555–564 (2008).

17. Josephs, K. A. *et al.* TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol* **127**, 811–824 (2014).

18. Josephs, K. A. *et al.* Updated TDP-43 in Alzheimer's disease staging scheme. *Acta Neuropathol* **131**, 571–585 (2016).

19. Kouri, N. *et al.* Corticobasal degeneration with olivopontocerebellar atrophy and TDP-43 pathology: An unusual clinicopathologic variant of CBD. *Acta Neuropathol* **125**, 741–752 (2013).

20. Koga, S. *et al.* Corticobasal degeneration with TDP-43 pathology presenting with progressive supranuclear palsy syndrome: a distinct clinicopathologic subtype. *Acta Neuropathol* **136**, 389–404 (2018).

21. Freeman, S. H., Spires-Jones, T., Hyman, B. T., Growdon, J. H. & Frosch, M. P. TAR-DNA Binding Protein 43 in Pick Disease. *J Neuropathol Exp Neurol* **67**, 62–67 (2008).

22. McKee, A. C. *et al.* TDP-43 Proteinopathy and Motor Neuron Disease in Chronic Traumatic Encephalopathy. *J Neuropathol Exp Neurol* **69**, 918–929 (2010).

23. Yang, C., Nag, S., Xing, G., Aggarwal, N. T. & Schneider, J. A. A Clinicopathological Report of a 93-Year-Old Former Street Boxer With Coexistence of Chronic Traumatic Encephalopathy, Alzheimer's Disease, Dementia With Lewy Bodies, and Hippocampal Sclerosis With TDP-43 Pathology. *Front Neurol* **11**, 493123 (2020).

24. Nakashima-Yasuda, H. *et al.* Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathol* **114**, 221–229 (2007).

25. Higashi, S. *et al.* Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. *Brain Res* **1184**, 284–294 (2007).

26. McAleese, K. E. *et al.* TDP-43 pathology in Alzheimer's disease, dementia with Lewy bodies and ageing. *Brain Pathology* **27**, 472–479 (2017).

27. Amador-Ortiz, C., Ahmed, Z., Zehr, C. & Dickson, D. W. Hippocampal sclerosis dementia differs from hippocampal sclerosis in frontal lobe degeneration. *Acta Neuropathol* **113**, 245–252 (2007).

28. Pao, W. C. *et al.* Hippocampal sclerosis in the elderly: genetic and pathologic findings, some mimicking Alzheimer disease clinically. *Alzheimer Dis Assoc Disord* **25**, 364 (2011).

29. Brenowitz, W. D., Monsell, S. E., Schmitt, F. A., Kukull, W. A. & Nelson, P. T. Hippocampal Sclerosis of Aging is a Key Alzheimer's Disease Mimic: Clinical-Pathologic Correlations and Comparisons with both Alzheimer's Disease and Non-Tauopathic Frontotemporal Lobar Degeneration. *Journal of Alzheimer's Disease* **39**, 691–702 (2014).

30. Nelson, P. T. *et al.* Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain* **142**, 1503–1527 (2019).

31. Nag, S. *et al.* Hippocampal sclerosis and TDP-43 pathology in aging and Alzheimer disease. *Ann Neurol* **77**, 942–952 (2015).

32. Meneses, A. *et al.* TDP-43 Pathology in Alzheimer's Disease. *Molecular Neurodegeneration 2021 16:1* **16**, 1–15 (2021).

33. Lin, W. L. & Dickson, D. W. Ultrastructural localization of TDP-43 in filamentous neuronal inclusions in various neurodegenerative diseases. *Acta Neuropathol* **116**, 205–213 (2008).

34. Thorpe, J. R., Tang, H., Atherton, J. & Cairns, N. J. Fine structural analysis of the neuronal inclusions of frontotemporal lobar degeneration with TDP-43 proteinopathy. *J Neural Transm* **115**, 1661–1671 (2008).

35. Nonaka, T. *et al.* Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. *Cell Rep* **4**, 124–134 (2013).

36. Holmes, B. B. *et al.* Proteopathic tau seeding predicts tauopathy in vivo. *Proc Natl Acad Sci U S A* **111**, E4376–E4385 (2014).

37. Yamasaki, T. R. *et al.* Parkinson's disease and multiple system atrophy have distinct -synuclein seed characteristics. *Journal of Biological Chemistry* **294**, 1045–1058 (2019).

38. Zhu, J. *et al.* VCP suppresses proteopathic seeding in neurons. *Mol Neurodegener* **17**, (2022).

39. 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).

40. Jiang, Y. X. *et al.* Amyloid fibrils in FTLD-TDP are composed of TMEM106B and not TDP-43. *Nature* **605**, 304–309 (2022).

41. Kwong, L. K., Uryu, K., Trojanowski, J. Q. & Lee, V. M. Y. TDP-43 Proteinopathies: Neurodegenerative Protein Misfolding Diseases without Amyloidosis. *Neurosignals* **16**, 41–51 (2007).

42. Johnson, B. S. *et al.* TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *Journal of Biological Chemistry* **284**, 20329–20339 (2009).

43. Arseni, D. *et al.* Structure of pathological TDP-43 filaments from ALS with FTLD. *Nature 2021 601:7891* **601**, 139–143 (2021).

44. Arseni, D. *et al.* TDP-43 forms amyloid filaments with a distinct fold in type A FTLD-TDP. *Nature* **620**, 898–903 (2023).

45. Lövestam, S. *et al.* Assembly of recombinant tau into filaments identical to those of Alzheimer's disease and chronic traumatic encephalopathy. *Elife* **11**, (2022).

46. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).

CHAPTER 3

Structure-based design of D-peptide disruptor of SARS-CoV-2 nucleocapsid phase separation

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 interfaces[10]. 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 ions[2], 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], condensates[20], and membrane-less organelles[19]. 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 cells[19].

Motivating our study is a medical experience that even efficient vaccines rarely eradicate viral diseases and their legacies of morbidity and mortality[21], so COVID-19 therapies are needed. Along with others[10] 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 assembly[10]. NCAP is also evolutionarily conserved in the coronavirus genus[10], 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 algorithm[22] 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 complexity[22]. 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).jh

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$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: $_{179}$GSQASS$_{184}$, $_{217}$AALALL$_{222}$, and $_{243}$GQTVTK$_{248}$. 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. 30,31,32,33). 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 ~6–16 μM, G12 exhibits dose-dependent antiviral activity while reducing the amount of virus detected in the culture by up to ~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 7–11 μM (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. 1and 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 vRNA25, 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], hnRNPA237, TDP-4338, and other RNA-binding proteins that are involved in RNA metabolism in eukaryotic cells[13,17], and under certain circumstances, also in amyloid-associated pathologies19. 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 $ZnCl_2$ (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: $_{179}$GSQASS$_{184}$, $_{217}$AALALL$_{222}$,

and $_{243}$GQTVTK$_{248}$(Fig. 4 and Supplementary Fig. 3a). $_{179}$GSQASS$_{184}$ and $_{243}$GQTVTK$_{248}$, are predominantly polar (Fig. 4c), similar to the highly polar reversible amyloid fibrils formed by the LCDs of FUS and hnRNPA240. The segment $_{179}$GSQASS$_{184}$ 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 assembly[8]. Both S180 and S184 face the dry, tight interface formed between the β-sheets in the structure of $_{179}$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, $_{217}$AALALL$_{222}$, is highly hydrophobic and produces the most stable steric-zipper structure (Supplementary Table 2). $_{217}$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 $_{217}$AALALL$_{222}$ render it an important target for the disruption of NCAP's self-assembly. The $_{243}$GQTVTK$_{248}$ segment, however, resembles sequences in LCDs found in the human proteome[29], and is therefore a poor target for drug design.

The self-assembly of NCAP is crucial for RNA packaging and SARS-CoV-2 replication[10]. 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 $_{217}$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 $_{217}$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: $_{179}$GSQASS$_{184}$, $_{217}$AALALL$_{222}$, and $_{243}$GQTVTK$_{248}$. Nevertheless, some mutations were detected within the central LCD, including the prevalent R203K/M, G204R/M, and T205I substitutions42·43·44. 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) substututions42·43·44 that are adjacent to the $_{217}$AALALL$_{222}$ steric-zipper segment. The Delta variant spread faster and caused more infection compared to its predecessors46·47·48·49. 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 proteins35·50·51·52. This is important to explore since amyloid fibrils are associated with numerous dementias and movement disorders53·54. 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 puncta[61]. 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 oligomers[62]. 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 fibrils[63]. The RIP-homotypic interaction motif containing segments of the herpes simplex virus 1 (HSV-1) protein ICP6[64], the murine cytomegalovirus protein M45[65] and the varicella-zoster virus protein ORF20[66] 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-assembly[67], peptides from the fiber protein of adenovirus[68,69], and a nine-residue peptide from the C-terminus of the SARS-CoV envelope protein[70]. Recent studies also showed the amyloidogenic properties of various segments of the spike protein[71], and other regions in the proteome[72] 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 NCAP[5] were identified in over two million eukaryotic

viruses[73]. 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; UniProtKB[74]accession number: P0DTC9 [https://covid-19.uniprot.org/uniprotkb/P0DTC9#Sequence]) was evaluated using SEG[22] 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 residues[22].

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 (UniprotKB[74] 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 tag[75] using splicing by overlap extension (SOE) technique[76]. 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 software[34]. 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 AREAIMOL[77] and Sc[78], respectively, from the CCP4 suite of crystallographic programs[79]. 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 PyMOL[80]. 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 $_{179}$GSQASS$_{184}$ and $_{243}$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 $_{217}$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^{TM} 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 $HClO_4$, and a 2:3 water to acetonitrile solution, buffered by 5.6 mM of $HClO_4$, 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′-TAATACGACTCACTATAGGCTGTGTGGCTGTCACTCG-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-µm 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 mL 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×*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 mL 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 $ZnCl_2$ 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 $ZnCl_2$ were mixed in PBS at final concentrations of 30 μM NCAP with 0 or 0.75 μM S2hp vRNA, and 0 or 20 μM $ZnCl_2$ 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×*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×*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 ×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 ×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):

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

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 px^2 were removed to reduce segmentation errors. For small regions, defined by an area less than 10,000 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-µm 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-µm 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 µM concentration. S2hp RNA was diluted by 10 mM HEPES pH 7.0 buffer to 75 µM concentration. The proteins, RNA and ThT were then mixed to final concentrations of 300 µM ThT, 30 µM protein, and 0 or 7.5 µM 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 μM concentration. The proteins and ThT were then mixed to final concentrations of 300 μM ThT and 100 μM 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 μL of the samples were taken out of the tubes at days 1, 6, and 11of 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-μm 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 μM protein and 0 or 7.5 μM 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-µm syringe filter. Preparations were done under sterile conditions and using sterile filter pipette tips to ensure RNA preservation. For grid preparation and screening, 4 µL 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 µL 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 °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 °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 °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 μM final concentration of LCD, in 50 μL 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 μM protein and 236 μM 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×*g* for 10 min at RT, then pelleted again and resuspended in 5 μL of deionized water. Fibrils were aligned by pipetting 2 μL 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 μL 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 (Milliepore 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 beamline24-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 $_{217}$AALALL$_{222}$ 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 μL of TFA and then diluting with 225 μL of acetonitrile and 225 μL 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 $_{217}$AALALL$_{222}$ 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 $_{179}$GSQASS$_{184}$ 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 Li_2SO_4, and Tris pH 7.0. Needle-shaped crystals appeared immediately after setting up the tray. The purified NCAP segment $_{243}$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 $(NH_4)_2SO_4$, 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 XSCALE[84]. Data collection statistics are reported in Table 1. Initial phases for AALALL and GSQASS were obtained by molecular replacement with the program Phaser[85] using a search model consisting of an ideal β-strand with sequence AAAAAA. Phases for GQTVTK were obtained by direct methods using the program ShelxD[86]. Simulated annealing composite omits maps[57] were calculated using Phenix[87]. Refinement was performed using the program Refmac[88]. Model building was performed using the graphics program Coot[89]. Structure illustrations were created using PyMOL[80]. Residue hydrophobicity of the steric zipper segments was assigned and colored according to the Kyte and Doolittle hydrophobicity scale embedded in UCSF Chimera[90].

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×*g* for 30 min to remove large aggregates, then aliquoted and stored at −20 °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 receptor[91] 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 µg/mL puromycin (Gibco cat no. A1113803) for selection, at 37 °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 °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 ~15 µM 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 µl 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% CO_2, 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% NaN_3 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 µg/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 IC_{50} 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% CO_2 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 μM 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_SARS-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.

Peer review

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.

Additional information

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

Select supplementary figures and tables are included, numbered the same as in the full supplementary. The online version containing additional supplementary material is available at https://www.nature.com/articles/s41467-023-37865-3

Figures and Tables

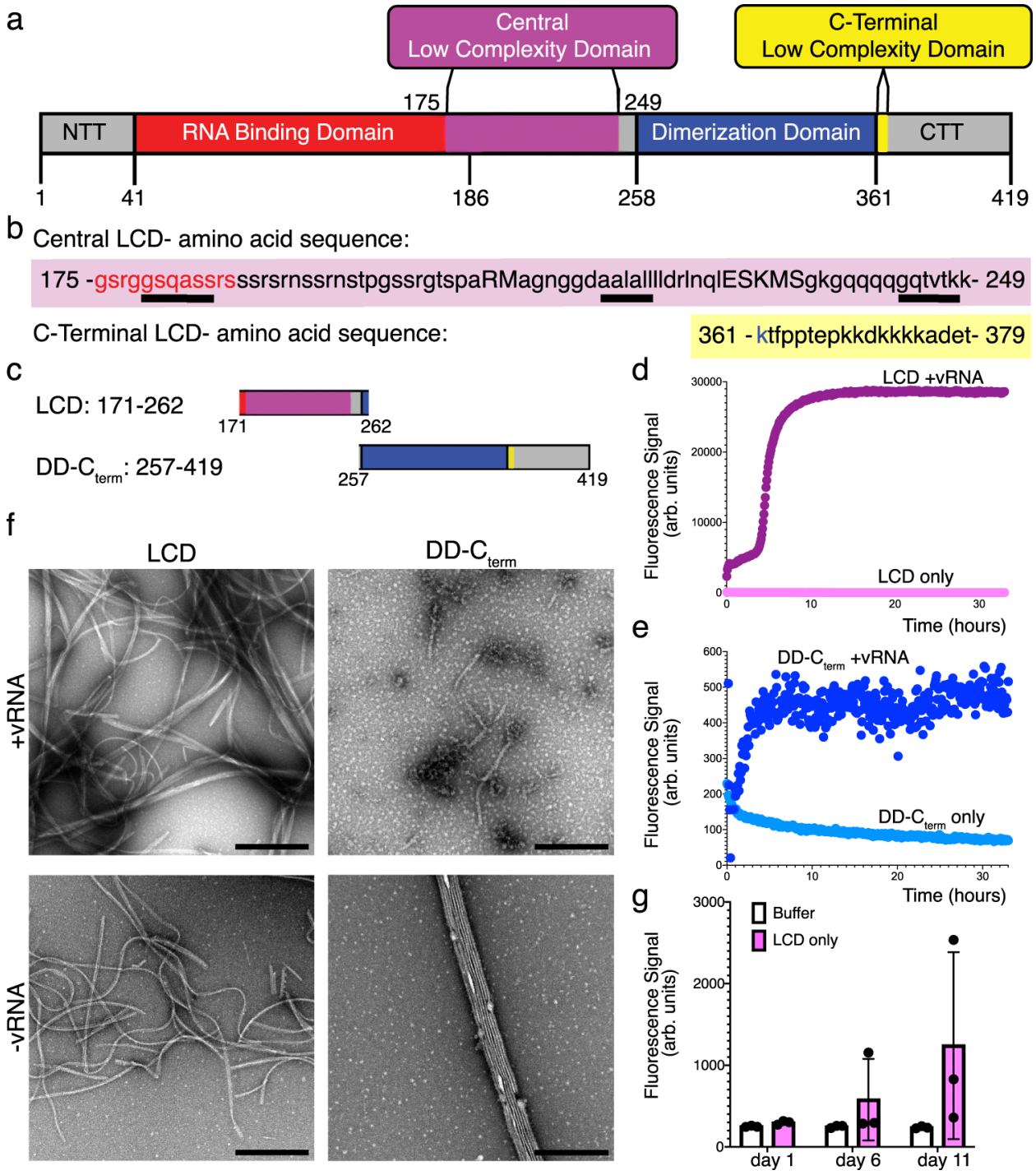

Figure 3-1: NCAP's LCDs form fibrils and ThT-positive species. a NCAP's domain organization. Domain definitions: N-terminal tail (NTT, gray), RNA-binding domain (red); Central low complexity domain (LCD, purple; residues 175–249), Dimerization domain (blue); C-terminal tail (CTT, gray). The C-terminal 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 ± SD. $n = 3$ samples. Source data for panels **d**, **e**, and **g** are provided as a Source Data file.

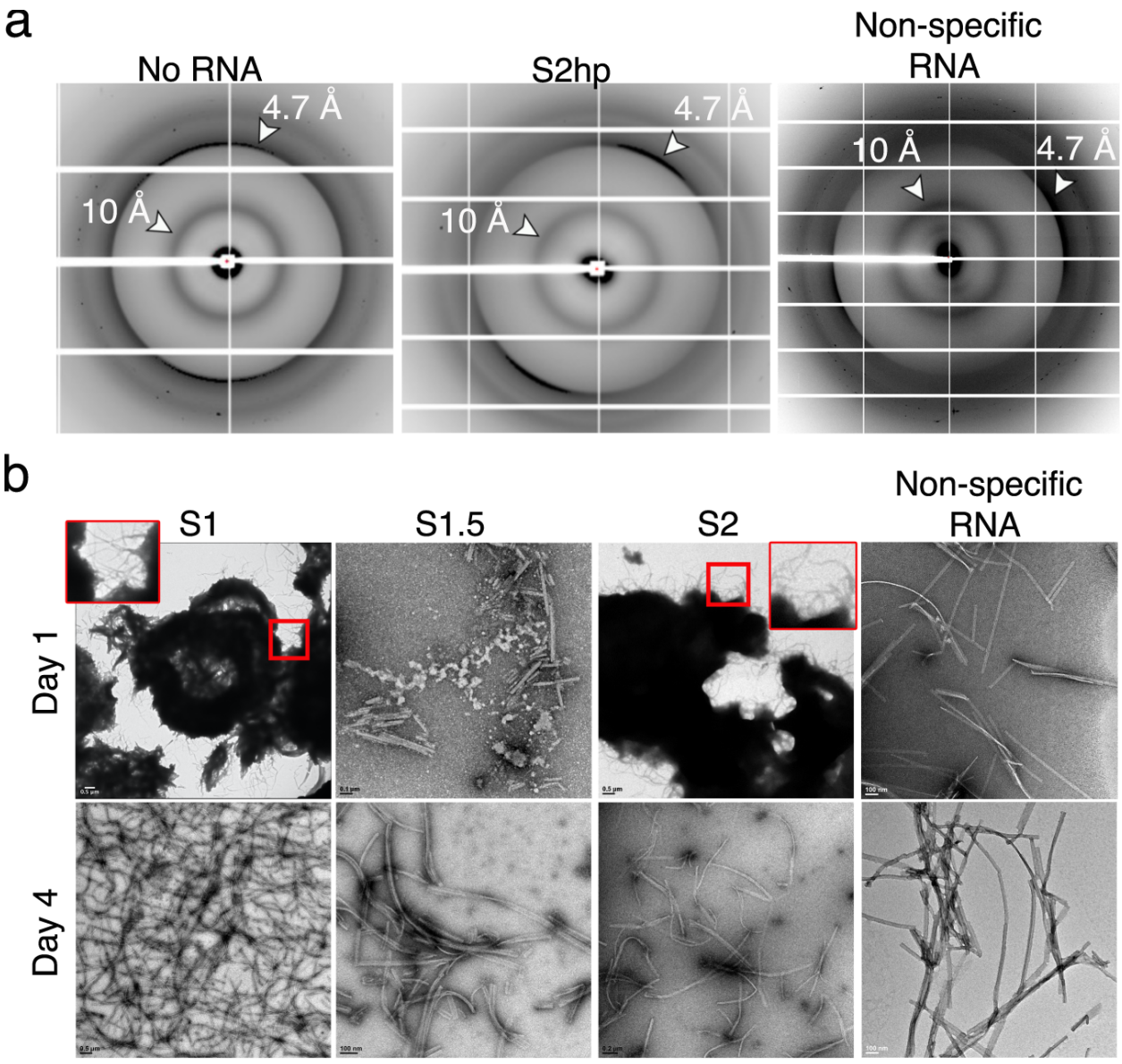

Figure 3-2: The central LCD segment of NCAP demonstrates amyloid-like characteristics.

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.

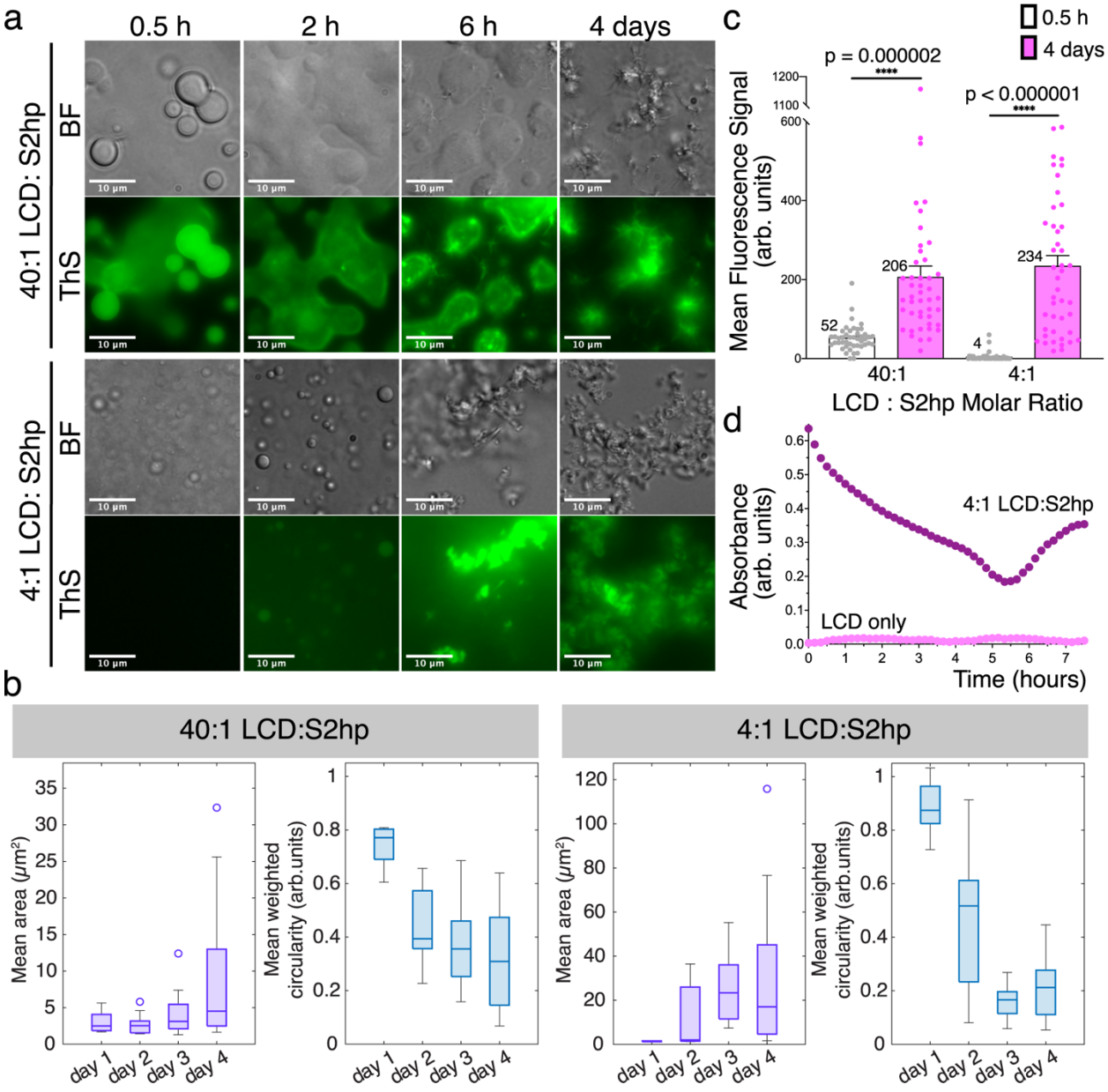

Figure 3-3: The central LCD segment of NCAP forms ThS fluorescent PS droplet that transition into amorphous and fibrous solid-like particles. a Brightfield (BF) and Thioflavin-

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 ± 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.

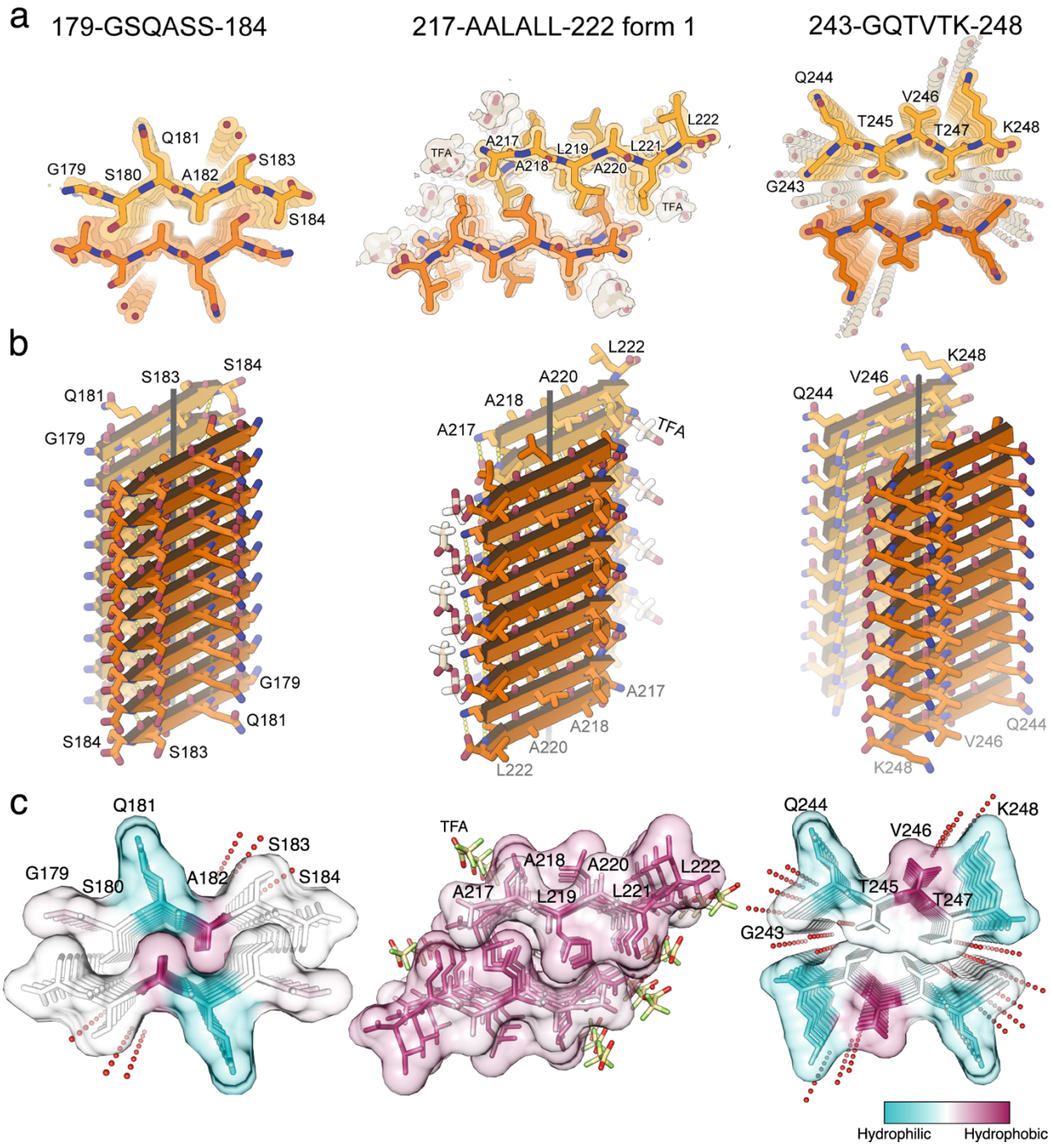

Figure 3-4: Atomic structures of amyloid-like association of NCAP segments revealed by crystallography. a Quality of the fit of each atomic model to its corresponding simulated annealing composite omit maps92. The maps are contoured at the 1.0 sigma level. All structural features are well defined by 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.

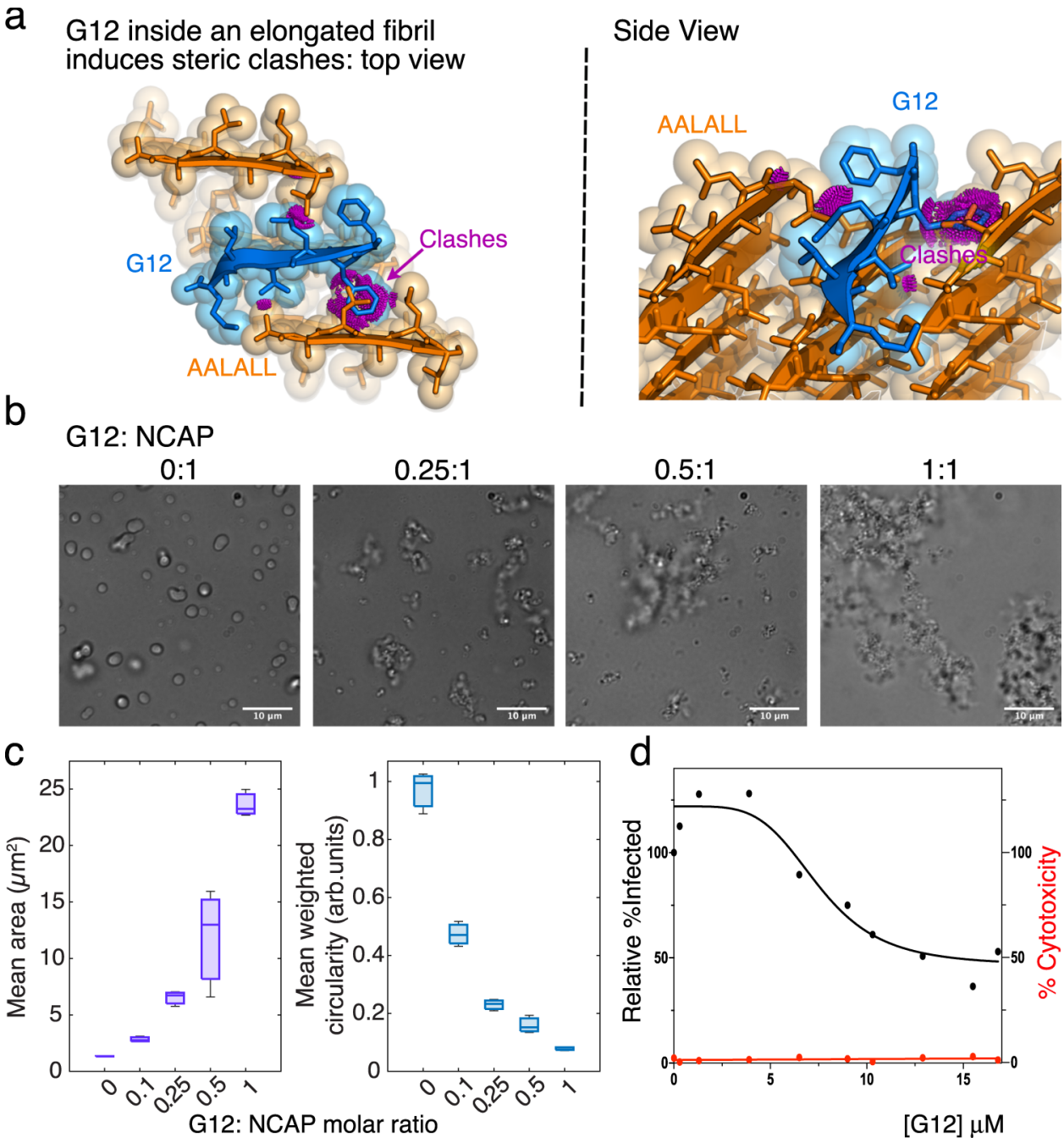

Figure 3-5: Design and evaluation of NCAP's self-assembly disruptor, G12. a The Rosetta-based design of G12 templated by the AALALL X-ray crystal structure form 1 (Fig. 4; Table 1). Model of the G12 (blue) capping an AALALL fibril (orange). The top view is down the fibril axis and the side view is tilted 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 μM. 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.

Segment	^{179}GSQASS^{184}	^{217}AALALL^{222} Form 1	^{217}AALALL^{222} Form 2	^{243}GQTVTK^{248}
Data collection				
Beamline	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E
Space group	$P2_12_12_1$	P1	$P2_12_12$	$P2_1$
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)	0.085 (0.446)
$CC_{1/2}$	99.7 (56.7)	98.5 (89.2)	99.7 (54.4)	99.5 (84.3)
Overall I/δ	3.5 (0.7)	5.9 (2.0)	5.9 (1.8)	6.0 (1.4)
Refinement				
R_{work}/R_{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]

Table 3-1: Crystallographic data collection and refinement statistics from SARS-CoV-2 NCAP segments. [a]Numbers in parentheses report statistics in the highest resolution shell. [b]Count includes hydrogen atoms.

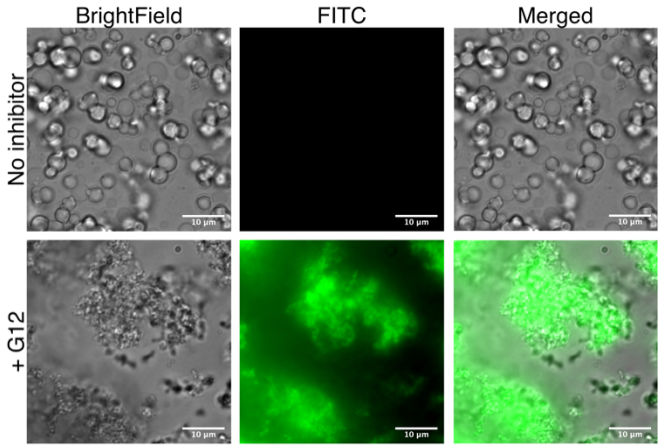

Figure 3-S6: Localization of FITC-tagged G12 in NCAP assemblies. PS droplets of NCAP were formed with S2hp vRNA and ZnCl2. FITC-tagged G12 was mixed with untagged peptide at a 1: 9 ratio, and was then added to the samples at 1: 1 final molar ratio with NCAP. NCAP PS without G12 shows no FITC fluorescence. Addition of G12 induces formation of aggregates that contain the G12 peptide (green).

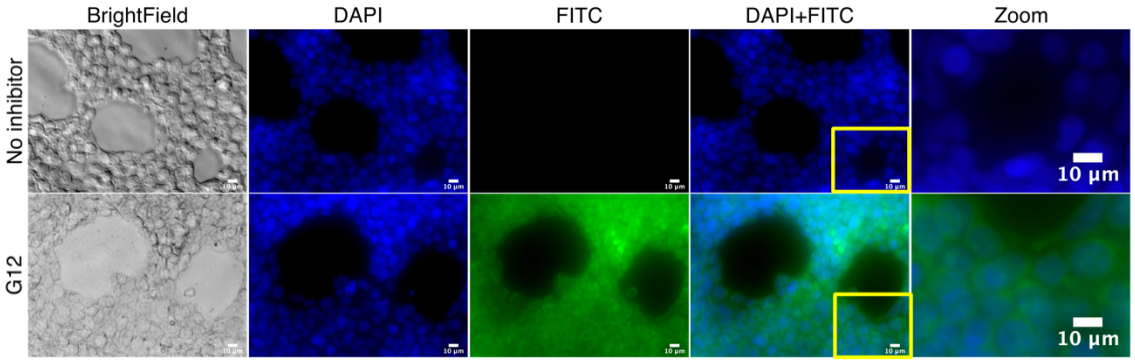

Figure 3-S7: FITC-labeled G12 is diffused in transfected HEK293-ACE2 cells as visualized using fluorescence microscopy. FITC-tagged G12 (green) was transfected into HEK293-ACE2 cells, which were then incubated for 24 hours at 37 oC, 5% CO2, then fixed and stained with DAPI (blue). The "zoom" inset on the right is an enlarged view of the yellow boxes in the composite DAPI+ FITC images. A 10 µm scale bar is shown on the right bottom side of each image.

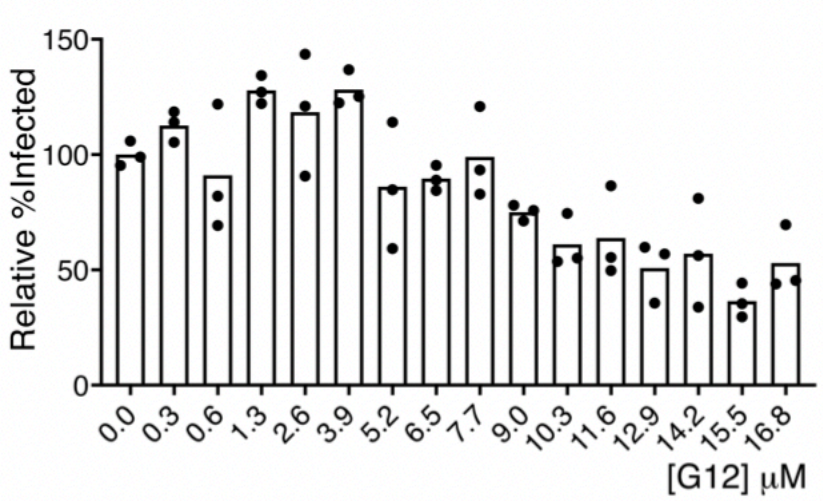

Figure 3-S8: (Extended version of Figure 3-5d). A full dose-dependence analysis of G12 inhibition of SARS-CoV-2 infection in HEK293-ACE2 cells. Cells were transfected with indicated concentrations of G12, infected 3-4 hours later with the virus and fixed at 24 hours post infection. The overall percentage of cells positive for infection in each sample were calculated via quantitative immunofluorescence labeling of the spike protein relative to the number of nuclei in each sample. The relative % infected was then achieved by normalizing the percentage of positive cells to the vehicle control (0 μM G12). Bars and dots indicate the mean and measured values of individual replicates, respectively.

Inhibitor Name	**Sequence**	**Target Structure**	**Design Approach**	**Rosetta Score[(a)]**
G12	d-(rrffmvlm)	AALALL	Rosetta-based	-34.5 (fibril top) -35.7 (fibril bottom)*

Table 3-S3: Sequence and Rosetta scores of G12 binding to an AALALL fibril. [(a)]An arbitrary energetic score calculated for the binding of the designed peptide inhibitor to the fibril tip6. These scores were calculated without the terminal arginine chain that was added to increase

the solubility of G12. *The Rosetta score for the binding of additional AALALL strand for comparison is -30.78 (fibril top) and -34.57 (fibril bottom).

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-κB 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 amylome, 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 & and the Center for Viral Systems Biology outbreak.info.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. Interimestimates 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 fromstaphylococcus 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 RHIMto 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.265

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. ActaCrystallogr. 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. ActaCrystallogr. 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).

CHAPTER 4

Structures of aggregation-driving segments of glaucoma-associated myocilin

Abstract

Primary open-angle glaucoma is caused by elevated intraocular pressure leading to damage of the optic nerve, however its precise molecular mechanisms are not known. Myocilin proteins are normally expressed in ocular cells and secreted into the aqueous humor, however disease-associated mutations cause their accumulation inside cells which may build up the intraocular pressure. Myocilin has been shown to form amyloid fibrils *in vitro*, leading to its intracellular retention in a cell model. Here, we investigated the structural basis of myocilin aggregation. We predicted nine aggregation-driving segments in the C-terminal olfactomedin domain, and validated six of them to fibrillize experimentally. We determined crystal structures of two segments, GAVVYS and LYTVSSY, which form steric zippers and serve as models for how these segments may contribute to the amyloid state.

Introduction

Glaucoma is the leading cause of irreversible blindness, affecting over 70 million people worldwide, and presently does not have a cure[1]. Glaucoma is a degenerative optic neuropathy in which the death of retinal ganglion cells and progressive damage to the optic nerve results in vision loss[2]. The most prevalent subtype of the disease, primary open-angle glaucoma (POAG), is characterized by the increase of intraocular pressure (IOP)[3]. The IOP is regulated by the trabecular meshwork (TM), a porous tissue responsible for drainage of the aqueous humor from the anterior chamber of the eye. Molecular aberrations within the TM lead to resistance of normal aqueous humor outflow, which raise the IOP and result in optic nerve damage[4]. The underlying mechanisms that contribute to the elevation of IOP are poorly understood.

In 1997, Stone at al. found that mutations in the *MYOC* gene, encoding for myocilin protein, cause autosomal-dominant forms of POAG[5]; however myocilin's physiological function and pathological role in glaucoma remain elusive more than 20 years later. Myocilin, expressed in ocular tissues, is synthesized in TM cells then secreted into the aqueous humor[6,7]. However myocilin containing glaucoma-associated mutations has been shown to aggregate in the secretory pathway and accumulate intracellularly[8,9]. A build up of myocilin may cause injury to TM cells, compromising the permeable tissue architecture[2]. The main clinical symptom of patients who harbor myocilin mutations is elevated IOP, suggesting that damage to TM cells leads to occlusion of aqueous humor outflow and increase of ocular pressure[10,11]. Recently, recombinant myocilin proteins were shown to form thioflavin T (ThT)-positive, proteinase K-resistant amyloid fibrils, and myocilin disease-mutants expressed in mammalian cells aggregated in ThT-positive puncta[12]. Amyloid formation is a possible explanation for myocilin intracellular accumulation and glaucoma pathogenesis, a hypothesis largely unexplored.

Results

Prediction and validation of aggregation-driving segments

Myocilin contains an N-terminal leucine zipper domain that facilitates multimer formation, and a C-terminal olfactomedin (OLF) domain whose native fold adopts a five-bladed β-propeller[13,14]. Nearly all disease-associated mutations in myocilin are found in the OLF domain[15,16]. Moreover, the OLF domain of myocilin is necessary and sufficient for the formation of amyloid fibrils *in vitro*[12]. However, the amyloid-driving segments within myocilin are not precisely known. We used ZipperDB, a computational algorithm that performs energetic evaluations to predict potential steric zippers in protein sequences[17], to identify segments of myocilin OLF that may

drive aggregation. We found that segments in myocilin with high amyloid propensity were concentrated in the OLF domain (Fig. 4-1a), consistent with results from other amyloid-predicting algorithms[18]. We chose nine of the most aggregation-prone segments ranging from 6 to 8 residues, to experimentally test their ability to aggregate: $_{313}$SKVHIL$_{318}$, $_{326}$GAVVYS$_{331}$, $_{331}$SGSLYF$_{336}$, $_{387}$GLWVIY$_{392}$, $_{399}$GAIVLS$_{404}$, $_{425}$SVANAFII$_{432}$, $_{436}$LYTVSSY$_{442}$, $_{460}$SKTLTI$_{465}$, $_{472}$KYSSMI$_{477}$. Six segments formed fibrils after shaking at 30 mg/mL concentration at 37 °C in phosphate buffered saline pH 7.4 for two weeks: GAVVYS, SGSLYF, GLWVIY, GAIVLS, SVANAFII, LYTVSSY. X-ray diffraction of fibrils produced cross-beta reflections characteristic of amyloids, at 4.7 Å and 8-10 Å[19,20], corresponding to the β-strand spacings parallel and perpendicular to the fibril direction (Fig. 4-1b). Fibrils were further confirmed by negative stain transmission electron microscopy (Fig 4-1c). Three of the tested segments did not form fibrils detectable by TEM: SKVHIL, SKTLTI, KYSSMI.

Crystal structures of steric zippers

Recombinant myocilin forms amyloid fibrils, and computational analyses predict segments in the protein to have high aggregation propensity[12,18]. To investigate the molecular basis of amyloid formation, we studied aggregation-driving segments using crystallography. We determined three structures from segments of myocilin's OLF domain: GAVVYS and LYTVSSY, the latter in two different conformations (Fig. 4-2). Crystals captured peptides as extended β-sheets, hydrogen bonding with layers above and below. These β-sheets interdigitate their sidechains like the teeth of a zipper, thus are known as "steric zippers". Steric zippers can be classified by symmetry, as described by Sawaya, Eisenberg et al[21,22]. The steric zipper formed by GAVVYS exhibits class 8 symmetry, with antiparallel, in-register β-sheets in a "face = back" arrangement. The zipper interface features the interdigitation of valines, creating a "dry interface" where solvent is

inaccessible. The LYTVSSY form 1 crystal is composed of antiparallel, out-of-register β-sheets. However unlike most antiparallel β-sheets observed in amyloid crystals, this β-sheet contains four unique orientations of the segment. Therefore the steric zipper symmetry is more complex than those in Sawaya et al.'s classification, which considers only two strand orientations at the most[21]. The β-sheets in the crystal are related to adjacent copies by 2_1 screw symmetry (Fig. 4-2e). GAVVYS exhibits better shape complementarity (Sc) of 0.79 compared to 0.66 for LYTVSSY. However LYTVSSY buries a larger surface area (A_b) of 158 Å^2, compared to 172 Å^2 for GAVVYS, due to its larger sidechains (Fig. 4-2c,d). These crystal structures of aggregation-prone segments, although studied in isolation, may be predictive of their conformation in the amyloid fibril. LYTVSSY form 2 does not form a steric zipper; the interfaces between peptide sidechains are separated by 1,4-dioxane molecules in the solvent. Thus, this structure is not a model for amyloid interactions of myocilin. If the sidechains of adjacent β-sheets did pack close enough to contact, the structure would resemble a class 6 steric zipper with its out-of-register β-sheets in a "face-to-back" arrangement.

Discussion

Elevated IOP is a major causative factor in primary open angle glaucoma, crucial for onset and progression of the disease. In healthy eyes, aqueous humor circulates in the anterior chamber to nourish avascular tissues. The porous tissue known as the trabecular meshwork regulates aqueous humor outflow and controls the IOP[4]. In glaucoma, higher IOP develops as the result of occluded aqueous humor drainage, however the molecular changes that compromise TM cellularity are not well understood. Wildtype myocilin is expressed in TM cells, then released into the aqueous humor through the secretory pathway[6]. However, glaucoma patients harboring

myocilin mutations have diminished levels of myocilin secreted into the aqueous humor[8]. In cultured primary human TM cells, myocilin mutants accumulate intracellularly and reduce secretion of endogenous myocilin[9]. Myocilin proteins containing glaucoma-associated mutations are prone to misfolding and aggregation, which suggests a mechanism for their retention in the ER of TM cells[23].

Recent reports of myocilin's amyloid properties exhibited *in vitro* have raised the possibility that conversion of myocilin to amyloid fibrils, potentially induced by disease-associated mutations, can lead to its intracellular accumulation in TM cells and contribute to glaucoma pathogenesis. To investigate this hypothesis, we performed computational predictions which identified nine candidate aggregation-driving segments from the C-terminal OLF domain. We found that six of those segments form amyloid-like fibrils *in vitro*, confirmed by x-ray diffraction and negative stain transmission electron microscopy, demonstrating their ability to aggregate under the right conditions. Mapping the aggregation-driving segments to the globular structure of myocilin OLF domain, we found they correspond well to β-strands in the native fold[14]. The six experimentally-validated segments are buried as adjacent β-strands within the second, third and fourth blades of the β-propeller (Fig. 4-S1a). Shielding of these segments by the native propeller fold helps rationalize why fibrillization could only occur when mutations lead to partial unfolding of the propeller and exposure of the aggregation-prone sequences. For example, the Gln368STOP accounts for ~50% of the myocilin mutation in POAG patients (which is ~2% of glaucoma patients). This nonsense mutation results in premature termination of translation and generates truncated proteins, which could expose segments susceptible to aggregation (Fig. 4-S1b).

We crystallized and determined structures from two of the segments, $_{326}$GAVVYS$_{331}$ and $_{436}$LYTVSSY$_{442}$; the latter was solved in two crystal forms. GAVVYS and LYTVSSY form 1

are steric zippers, with antiparallel β-sheets and sidechains mated in the zipper interface. Atomic structures of short peptides can be indicative of the conformations adopted by those sequences in the amyloid fibril, such as VQIVYK[21] and VQIINK[24] from tau in Alzheimer's disease and segments from the human prion protein[25]. The mated β-sheets offer plausible models for residue packing in the amyloid core, which can serve as structural templates for inhibitor designs. Glaucoma is the second leading cause of blindness worldwide, behind cataracts[26]. While vision loss caused by cataracts can be restored with surgery, current glaucoma therapies can only delay but not prevent retinal ganglion cell death and optic nerve damage. Much work is required to demonstrate that the amyloidogenicity of myocilin plays a role in disease, but if this is the case, targeting myocilin aggregation could be a new avenue for glaucoma intervention. Inhibition of amyloid fibrils, associated with numerous diseases, has potential for therapeutic development[24,27–29]. Using myocilin amyloid structures as templates, capping inhibitors can be designed to bind onto the ends of amyloid fibrils and block the addition of monomer copies of the protein, thereby halting fibril elongation[30,31] (Fig. 4-S2). Structure-based design can produce highly specific inhibitors; they have the ability to interact potently with the target but not interfere with normal cellular processes. Promising inhibitors of myocilin aggregation can be tagged with cell-penetrating peptide tags, and easily administered to glaucoma patients in the form of eye drops[32].

Methods

Computational prediction of aggregation-driving segments

The protein sequence of myocilin (UniProt accession number A0A0H3W5P3) was analyzed by the ZipperDB algorithm (https://services.mbi.ucla.edu/zipperdb/), which threads hexapeptide

segments onto the NNQQNY crystal structure and assesses their propensity to form self-complementary steric zippers using RosettaDesign[17].

Peptide aggregation

Segments from the myocilin OLF domain were synthesized and purified to >98% (GenScript). Lyophilized peptides were dissolved at 30 mg/mL concentration in phosphate buffered saline pH 7.4, then incubated at 37°C with ~900rpm shaking (Torrey Pines Scientific) for 2 weeks. Peptide fibrils were pelleted by centrifugation at 18,400 g for 10 minutes, resuspended and washed three times in water, then inspected by x-ray diffraction and negative stain electron microscopy.

X-ray diffraction of fibrils

Peptide fibrils were aligned by pipetting 3 μL of the resuspended pellet in a 3 mm gap between two fire-polished glass rods. After drying at room temperature for a few hours, another 3 μL of the fibril solution was applied and dried again. Diffraction data of the aligned fibrils was collected on a FR-E+ rotating anode x-ray generator (Rigaku) equipped with a R-AXIS HTC imaging plate detector (Rigaku). Cu K-α x-ray beam with 1.5406 Å wavelength was used, and the detector was placed 78 mm from the sample. Diffraction images are visualized using ADXV (The Scripps Research Institute).

Negative stain transmission electron microscopy

Carbon-coated formvar support films mounted on copper grids with 400-mesh (Ted Pella) were glow-discharged for 30 seconds. 5 μL of pelleted fibrils was applied to the grids and incubated for 10 minutes, then excess solution was blotted off using filter paper. Grids were washed with water, then stained with 5 μl of 2% uranyl acetate (Electron Microscopy Sciences) for 2 minutes

and air-dried. The grids were imaged using a Tecnai T12 transmission electron microscope (FEI).

Peptide crystallization

Lyophilized peptides were dissolved to 25 mg/mL concentration in water and ammonium hydroxide, to neutralize residual trifluoroacetic acid from the lyophilized powder. Peptide crystals were grown by hanging drop vapor diffusion. 96-well crystallization trays were set up by the UCLA-DOE Crystallization Facility using a Mosquito robot (SPT Labtech) dispensing 200 nL drops. GAVVYS formed needle crystals at 20 °C in a reservoir solution composed of 20% (w/v) PEG 1000, 100mM Tris base/HCl pH 7.0 (Rigaku). $_{436}$LYTVSSY$_{442}$ grew needle crystals at 20 °C in reservoir solutions composed of 10% (v/v) 2-propanol, 100mM Tris base/HCl pH 8.5 (form 1, Rigaku) and 35% 1,4-Dioxane (form 2, Molecular Dimensions).

X-ray data collection, processing and structure determination

Crystals were mounted on a loop and stored in liquid nitrogen. X-ray datasets were collected at the Advanced Photon Source Synchrotron on the microfocus beamline 24-ID-E, equipped with a Dectris Eiger 16M pixel detector. Crystals were cryo-cooled to 100 K. Data was collected in 5° wedges using a 5 μm beam at a wavelength of 0.9792 Å. Data was indexed, integrated, scaled and merged using XDS and XSCALE[33] and Denzo/Scalepack from the HKL suite of programs[34]. Initial phases for GAVVYS and LYTVSSY form 1 were obtained by molecular replacement with antiparallel poly-alanine β-strands as the search model using Phaser[35]. Phases for LYTVSSY form 2 were obtained by direct methods using ShelxD[36]. Refinement was performed using Refmac[37] and Phenix[38]. Model building was performed using Coot[39]. Protein structures and models were illustrated using PyMOL[40]. X-ray data collection and refinement statistics are reported in Table 1.

Data availability

Crystal structures have not been deposited into the Protein Data Bank. They are available on the local UCLA Molecular Biology Institute servers.

Author contributions

Yi Xiao Jiang. performed computational prediction, peptide aggregation assays and negative stain TEM. Y.X.J. and Michael Collazo set up crystallization trays. Michael R. Sawaya and Duilio Cascio trained Y.X.J., and together, performed x-ray diffraction, data collection, data processing and structure determination. Y.X.J. and David S. Eisenberg designed the project.

Acknowledgements

We acknowledge the use of Northeastern Collaborative Access Team beamline 24-ID-E for measuring X-ray diffraction data from our microcrystals. The beamline is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector 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.

Figures and Tables

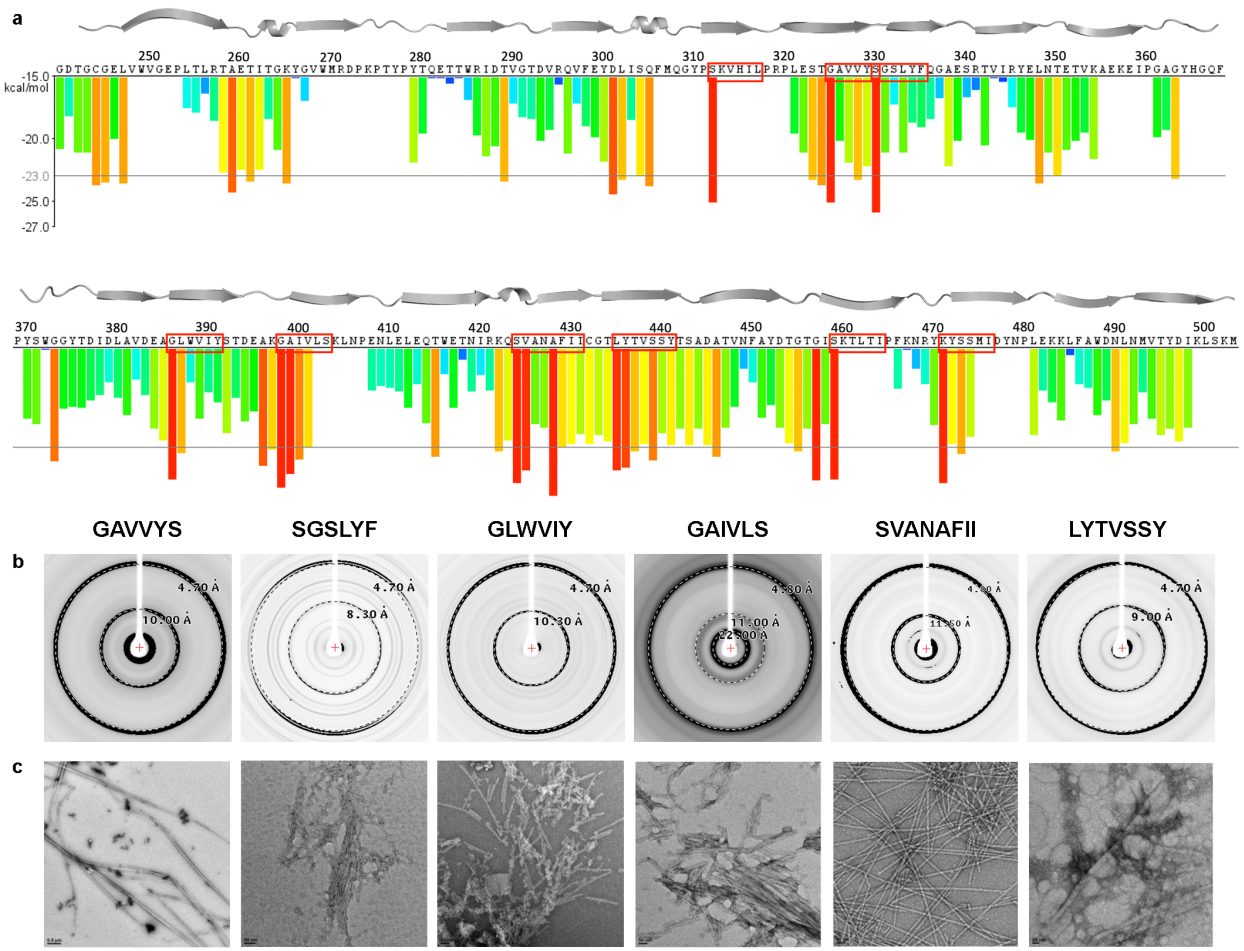

Figure 4-1: Computational prediction of aggregation-driving segments in myocilin olfactomedin domain and experimental validation. a, The sequence of the myocilin olfactomedin domain, residues 241-504, was analyzed by ZipperDB. Hexapeptide segments were scored for their propensity to form steric zippers and arbitrary energy values calculated by RosettaDesign are shown as a bar chart. The native secondary structure of the sequence is shown above. Nine of the most aggregation-prone segments (red boxes), ranging from 6 to 8 residues, were chosen for experimental validation. **b**, X-ray diffraction of aligned peptide fibrils exhibit cross-beta reflections at 4.7 Å and 8-10 Å. **c**, Negative stain transmission electron microscopy shows peptide fibril morphology.

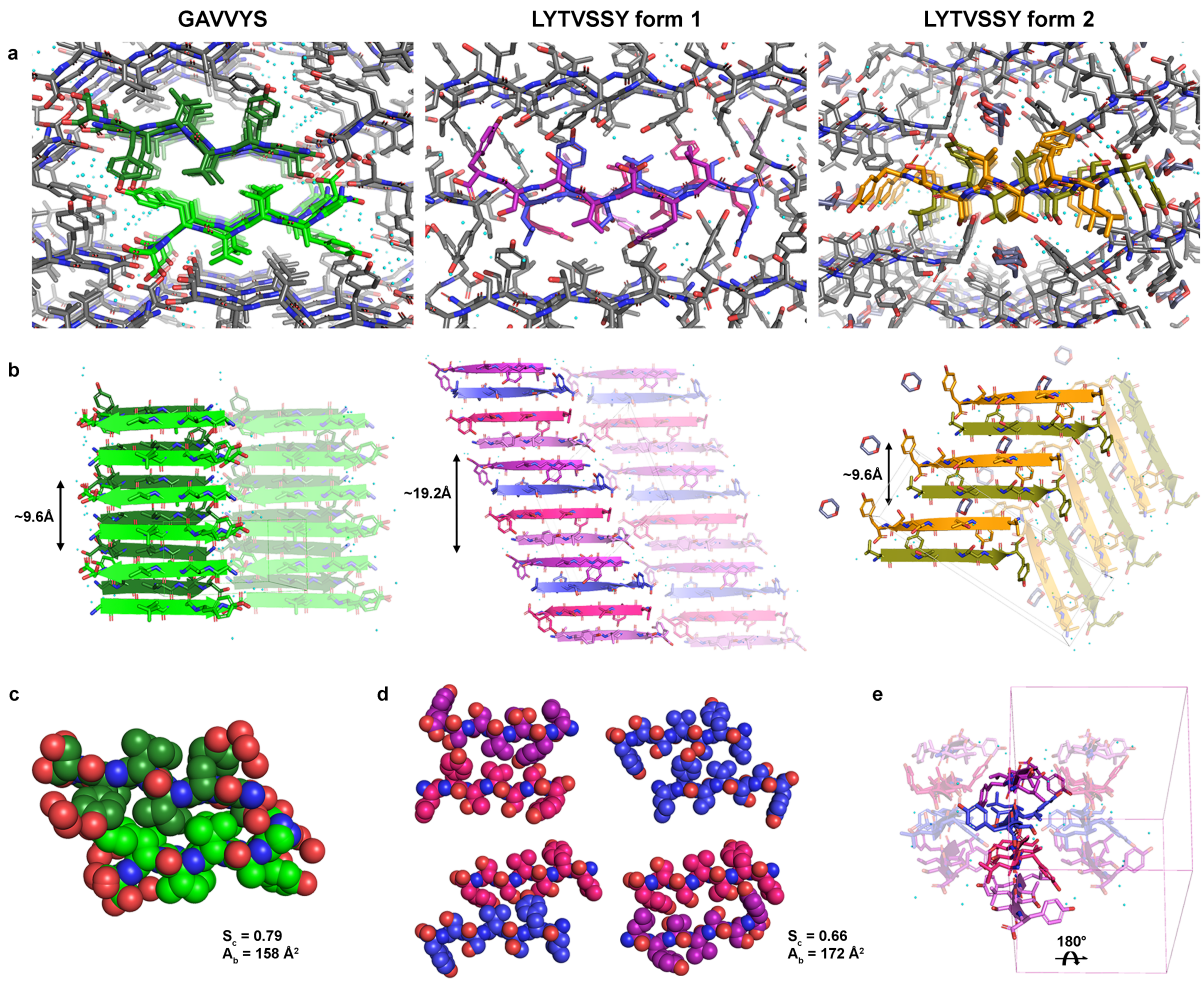

Figure 4-2: Crystal structures of GAVVYS and LYTVSSY segments from myocilin olfactomedin domain. a, Crystal packing of GAVVYS, LYTVSSY form 1 and form 2 shown by a view down the fibril axis. **b**, GAVVYS packs as antiparallel, in-register β-sheets while LYTVSSY form 1 and form 2 adopt antiparallel, out-of-register β-sheets, shown by a view perpendicular to the fibril axis. Arrows measure the asymmetric unit of the crystals: two β-strands ~9.6 Å for GAVVYS and LYTVSSY form 2, and four β-strands ~19.2 Å for LYTVSSY form 1. **c**,**d**, Steric zippers atoms shown as spheres to illustrate the shape complementarity (S_c) and buried surface area (A_b) of the interface, for GAVVYS (**c**) and LYTVSSY form 1 (**d**). **e**,

Four uniquely-oriented segments in the asymmetric unit of LTYVSSY form 1 crystals are related to adjacent copies by a 180° rotation and a ~10 Å translation.

Table 4-1: X-ray crystallography data collection and refinement statistics for myocilin segments. r.m.s.d. root mean square deviation. Values in parentheses refer to last shell.

Peptide	$_{326}$**GAVVYS**$_{331}$	$_{436}$**LYTVSSY**$_{442}$ **(form 1)**	$_{436}$**LYTVSSY**$_{442}$ **(form 2)**
DATA COLLECTION			
Space group	P1	$P2_1$	$P2_1$
Resolution (Å)	1.3 (1.33-1.30)	1.3 (1.34-1.30)	1.1 (1.13-1.10)
Unit cell dimensions a, b, c (Å)	9.59, 17.52, 19.57	20.18, 19.21, 25.35	11.50, 33.97, 13.51
Unit cell angles α, β, γ (°)	89.33, 89.97, 89.94	90.00, 112.13, 90.00	90.00, 92.16, 90.00
Measured reflections	8368	18621	17842
Unique reflections (%)	2764	4261	3702
Completeness (%)	88.3 (81.4)	93.2 (88.0)	87.2 (42.6)
R_{sym}	15.5 (71.8)	14.2 (77.8)	10.7 (73.2)
$CC_{1/2}$	98.4	99.5	99.5
I/δ	4.05 (1.80)	6.02 (1.32)	6.55 (1.29)
REFINEMENT			
R_{work}	0.193	0.125	0.149
R_{free}	0.232	0.150	0.154
r.m.s.d bond length (Å)	0.012	0.014	0.018
r.m.s.d bond angle (°)	1.70	1.80	2.09
Number of protein atoms	195	240	232
Number of solvent atoms	13	13	39

Average B factor of protein atoms (Å^2)	11.2	6.8	11.8
Average B factor of solvent atoms (Å^2)	20.3	21.4	28.9

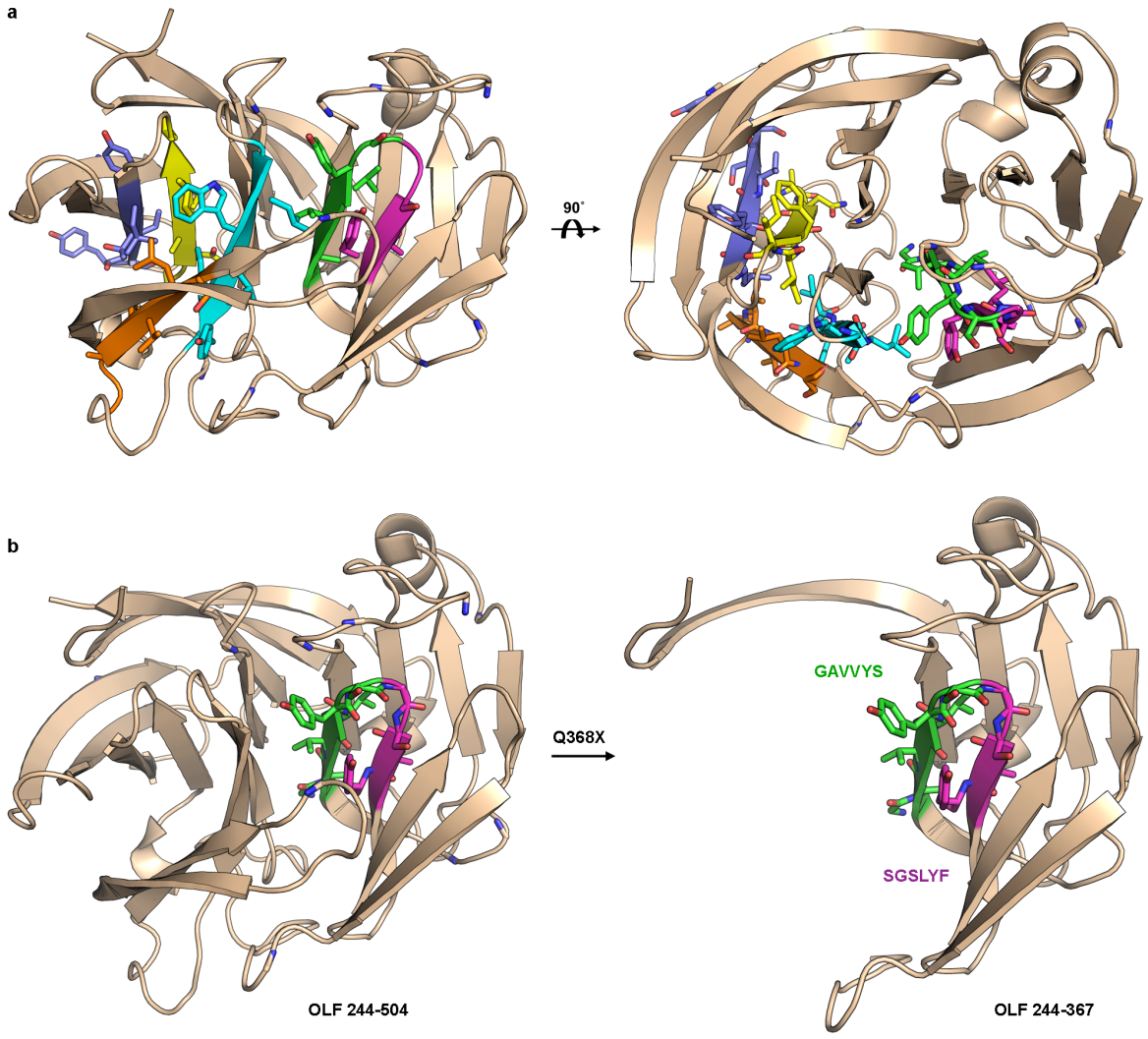

Figure 4-S1: Aggregation-driving segments protected in the stable five-bladed β-propeller fold of myocilin olfactomedin domain. a, Aggregation-driving segments identified by ZipperDB are mapped onto the native fold of the myocilin olfactomedin domain (PDB 4wxq)[14]: $_{326}$GAVVYS$_{331}$ (green), $_{331}$SGSLYF$_{336}$ (magenta), $_{387}$GLWVIY$_{392}$ (blue), $_{399}$GAIVLS$_{404}$ (orange), $_{427}$ANAFII$_{432}$ (yellow), $_{436}$LYTVSSY$_{442}$ (purple). **b**, The most common POAG-

associated mutation Gln368STOP causes truncation of the C-terminal end, potentially exposing a pair of aggregation-prone segments GAVVYS and SGSLYF.

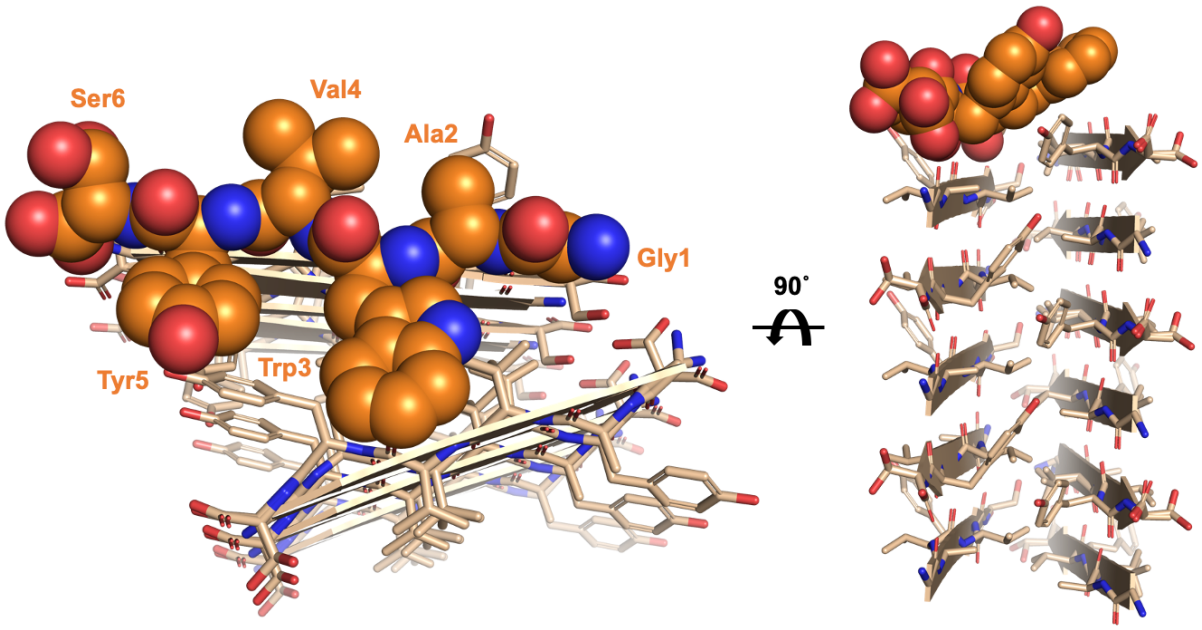

Figure 4-S2: An example of a structure-based capping inhibitor designed against the GAVVYS steric zipper structure. The GAVVYS crystal structure serves as a model for the tip of an elongating myocilin amyloid fibril, which can be the template for rational or computationally-aided designs of inhibitor peptides. In this rational design example, a sequence homologous to the steric zipper (GAVVYS) is used to ensure strong affinity for the fibril tip. Then, a residue in the steric zipper interface (Val3) is substituted with a with bulky side chain (Trp3) that results in a geometric clash with the opposite strand. The resulting inhibitor candidate (GAWVYS) should bind strongly to the fibril tip and prevent the addition of the next layer.

References

1. Quigley, H. A. & Broman, A. T. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* **90**, 262–267 (2006).
2. Kwon, Y. H., Fingert, J. H., Kuehn, M. H. & Alward, W. L. M. Primary open-angle glaucoma. *N Engl J Med* **360**, 1113–1124 (2009).
3. Bahrami, H. Causal inference in primary open angle glaucoma: specific discussion on intraocular pressure. *Ophthalmic Epidemiol* **13**, 283–289 (2006).
4. Alvarado, J., Murphy, C. & Juster, R. Trabecular meshwork cellularity in primary open-angle glaucoma and nonglaucomatous normals. *Ophthalmology* **91**, 564–579 (1984).
5. Stone, E. M. *et al.* Identification of a Gene That Causes Primary Open Angle Glaucoma. *Science (1979)* **275**, 668 LP – 670 (1997).
6. Polansky, J. R., Kurtz, R. M., Alvarado, J. A., Weinreb, R. N. & Mitchell, M. D. Eicosanoid production and glucocorticoid regulatory mechanisms in cultured human trabecular meshwork cells. *Prog Clin Biol Res* **312**, 113–138 (1989).
7. Escribano, J., Ortego, J. & Coca-Prados, M. Isolation and characterization of cell-specific cDNA clones from a subtractive library of the ocular ciliary body of a single normal human donor: transcription and synthesis of plasma proteins. *J Biochem* **118**, 921–931 (1995).
8. Jacobson, N. *et al.* Non-secretion of mutant proteins of the glaucoma gene myocilin in cultured trabecular meshwork cells and in aqueous humor. *Hum Mol Genet* **10**, 117–125 (2001).
9. Caballero, M., Rowlette, L. L. & Borras, T. Altered secretion of a TIGR/MYOC mutant lacking the olfactomedin domain. *Biochim Biophys Acta* **1502**, 447–460 (2000).

10. Alward, W. L. *et al.* Clinical features associated with mutations in the chromosome 1 open-angle glaucoma gene (GLC1A). *N Engl J Med* **338**, 1022–1027 (1998).

11. Fingert, J. H., Stone, E. M., Sheffield, V. C. & Alward, W. L. M. Myocilin glaucoma. *Surv Ophthalmol* **47**, 547–561 (2002).

12. Orwig, S. D. *et al.* Amyloid fibril formation by the glaucoma-associated olfactomedin domain of myocilin. *J Mol Biol* **421**, 242–255 (2012).

13. Fautsch, M. P. & Johnson, D. H. Characterization of myocilin-myocilin interactions. *Invest Ophthalmol Vis Sci* **42**, 2324–2331 (2001).

14. Donegan, R. K. *et al.* Structural basis for misfolding in myocilin-associated glaucoma. *Hum Mol Genet* **24**, 2111–2124 (2015).

15. Fingert, J. H. *et al.* Analysis of myocilin mutations in 1703 glaucoma patients from five different populations. *Hum Mol Genet* **8**, 899–905 (1999).

16. Wang, H. *et al.* Physiological function of myocilin and its role in the pathogenesis of glaucoma in the trabecular meshwork (Review). *Int J Mol Med* **43**, 671–681 (2019).

17. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A* **107**, 3487–3492 (2010).

18. Hill, S. E., Donegan, R. K. & Lieberman, R. L. The glaucoma-associated olfactomedin domain of myocilin forms polymorphic fibrils that are constrained by partial unfolding and peptide sequence. *J Mol Biol* **426**, 921–935 (2014).

19. Eanes, E. D. & Glenner, G. G. X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem* **16**, 673–677 (1968).

20. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729–739 (1997).

21. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).

22. Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *https://doi.org/10.1146/annurev-biochem-061516-045104* **86**, 69–95 (2017).

23. Zhou, Z. & Vollrath, D. A cellular assay distinguishes normal and mutant TIGR/myocilin protein. *Hum Mol Genet* **8**, 2221–2228 (1999).

24. Seidler, P. M. *et al.* Structure-based inhibitors of tau aggregation. *Nat Chem* **10**, 170–176 (2018).

25. Glynn, C. *et al.* Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. *Nat Struct Mol Biol* **27**, 417–423 (2020).

26. Pascolini, D. & Mariotti, S. P. Global estimates of visual impairment: 2010. *Br J Ophthalmol* **96**, 614–618 (2012).

27. Sievers, S. A. *et al.* Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100 (2011).

28. Griner, S. L. *et al.* Structure-based inhibitors of amyloid beta core suggest a common interface with tau. *Elife* **8**, (2019).

29. Sangwan, S. *et al.* Inhibition of synucleinopathic seeding by rationally designed inhibitors. *Elife* **9**, (2020).

30. Tjernberg, L. O. *et al.* Arrest of beta-amyloid fibril formation by a pentapeptide ligand. *J Biol Chem* **271**, 8545–8548 (1996).

31. Sciarretta, K. L., Gordon, D. J. & Meredith, S. C. Peptide-based inhibitors of amyloid assembly. *Methods Enzymol* **413**, 273–312 (2006).

32. Patel, S. G. *et al.* Cell-penetrating peptide sequence and modification dependent uptake and subcellular distribution of green florescent protein in different cell lines. *Sci Rep* **9**, 6298 (2019).

33. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).

34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326 (1997).

35. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).

36. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479–485 (2010).

37. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367 (2011).

38. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *urn:issn:0907-4449* **68**, 352–367 (2012).

39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).

40. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version~1.8*. (2015).