Title
Computational analysis of amyloid protein structure to identify novel pathologies and therapeutics

Author
Murray, Kevin Alexander

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational analysis of amyloid protein structure to identify novel pathologies and therapeutics

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Kevin Alexander Murray

2020

ABSTRACT OF THE DISSERTATION

Computational analysis of amyloid protein structure

to identify novel pathologies and therapeutics

by

Kevin Alexander Murray

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2020

Professor David S. Eisenberg, Chair

Many diseases are characterized by the pathologic accumulation of aggregated proteins. Known as amyloid, these fibrillar aggregates are present in many neurodegenerative diseases, including Alzheimer's and Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS). The development and spread of amyloid fibrils within the brain correlates with disease onset and progression, and inhibiting their formation is a possible route towards therapeutic development. Advances in structural biology, namely micro-crystal x-ray diffraction, micro-electron diffraction (MicroED), cryo-electron microscopy (cryoEM) and solid-state NMR spectroscopy (ssNMR) have enabled the determination of amyloid fibril structures to atomic-level resolutions, improving the possibility of structure-based inhibitor design. In Chapter 1, we use these amyloid structures to design inhibitors which bind to the ends of fibrils, "capping" them so as to prevent further growth. Applying recent breakthroughs in *de novo* protein design, we describe a

computational approach to develop mini-protein inhibitors of 35-48 residues which target the amyloid structures of tau, Aβ (found in Alzheimer's disease) and αSyn (found in Parkinson's disease). Biophysical characterization of the *in silico* designed inhibitors shows they form stable folds, with no sequence homology to naturally occurring proteins, and specifically prevent the aggregation of their targeted amyloid-prone proteins *in vitro*. The inhibitors also prevent the seeded aggregation and toxicity of fibrils in cells. *In vivo* evaluation reveals their ability to reduce aggregation and rescue motor deficits in *C. elegans* models of PD and AD.

In Chapter 2, we apply a similar design strategy to generate inhibitors of a different form of protein aggregation. Proteins with low complexity segments engage in liquid-liquid phase separation (LLPS) during normal cell processes, but aberrant LLPS leads to the eventual aggregation of such proteins into amyloid fibrils. This presents inhibition of LLPS as another potential therapeutic target. As LLPS appears to be a precursor state to fibrillization, we test if inhibitors targeting an amyloid fibril structure can also reduce LLPS of the same protein. To accomplish this, we design *de novo* miniprotein inhibitors which target the fibril structure of the FUS low-complexity domain, a protein known to undergo LLPS and eventually aggregate in ALS, frontotemporal dementia (FTD) and other dementias. Several designs are able to reduce FUS LLPS in vitro, as well as FUS fibrilization. Additionally, the same inhibitors reduce stress granule formation in cells, a form of LLPS. Mutations to improve binding to the FUS fibril interface improve stress granule reduction, while steric clashes introduced into the interface abolish the inhibitor effects. The top inhibitor construct iFUS-G specifically inhibits FUS LLPS and has no effect on phase separation of low complexity segments from TDP43 or hnRNPA2, aggregation prone proteins similar in sequence composition to FUS. These findings present a

rational design strategy to specifically inhibit the phase separation of low complexity proteins and have implications for the structural underpinnings of protein LLPS.

Next, we transition from therapeutically targeting known amyloid proteins to identifying new ones. In addition to FUS, many proteins including hnRNPA1, hnRNPA2, and TDP-43 have been established to undergo aggregation into amyloid-like fibrils through interactions of their low-complexity domains. Mutations in the low-complexity domains of the same proteins can lead to irreversible amyloid aggregation and disease. In Chapter 3 we introduce a computational procedure to identify mutations in low-complexity domains of disease-related proteins that are predicted to increase their propensity for amyloid aggregation. This procedure found several disease-related mutations in a low complexity region of the intermediate filament protein Keratin-8 (KRT8). Atomic structures of wild-type and mutant KRT8 segments confirm the transition of a highly extended strand to a pleated strand capable of amyloid formation. Biochemical analysis of KRT8 reveals the protein forms amyloid aggregates and that the identified mutations promote aggregation. Aggregated KRT8 is found in Mallory-Denk bodies, often observed in the hepatocytes of livers with alcoholic steatohepatitis (ASH). We demonstrate that ethanol promotes KRT8 aggregation, and KRT8 amyloid structures co-crystallize with alcohol. We also observe that KRT8 aggregation can be seeded with ASH patient liver extract, consistent with the amyloid nature of KRT8 aggregates.

Lastly, in Chapter 4 we explore structural characteristics that distinguish amyloid proteins known to undergo reversible versus irreversible aggregation. While all amyloid fibrils are primarily composed of repeating layers of beta-sheets, we observe that fibril structures of proteins known to reversibly aggregate have an enrichment of highly extended non-ideal beta-sheets. Quantum calculations of pleated- and extended-beta sheet amyloid structures show that

extended backbones decrease the energy required to separate strand pairs. Non-covalent interaction analysis shows that the extended beta-sheets may be stabilized by interactions between the amide proton and carbonyl oxygen of the same residue, known as C5 hydrogen-bonding. These findings identify a key structural element that may regulate reversible amyloid assembly.

This body of work offers insight into new ways to therapeutically target protein aggregation, identifies novel amyloid pathologies, and explores the structural underpinnings that may distinguish different forms of amyloid aggregation.

The dissertation of Kevin Alexander Murray is approved

Lin Jiang

Jose A. Rodriguez

Todd O. Yeates

Feng Guo

David S. Eisenberg, Committee Chair

University of California, Los Angeles

2020

To MFM and DRM – Thank you for your endless support, kindness, teaching me the value of hard work, and providing for me what you did not have yourself. This would not exist without your encouragement to pursue my own passions.

To KDM, TLB/AJB, and JMM/MB – Thank you for being role models, your generosity, and keeping me humble. Long live the Murray's.

To CNJ – For making sure I do things other than science from time to time and aiding in my pursuit of large quantities of frequent flier miles. You are my joy and best friend.

To my friends – May this document serve as an explanation for why I fell asleep at dinner or never seemed to make it to surprise parties on time. Each of you have been a unique source of support and the occasional ride home from campus.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKOWLEDGEMENTS

My pathway through education is largely a testament to the support I have received along the way. This support has come in many ways. It has come from family, friends, mentors, and colleagues. It has taken the form of two parents whose kindness and love have shaped the person I am today. Who were always out front of the school in a Ford waiting to pick me up from practice or drive me to school when I missed the bus and it was -10 degrees outside. Parents that were there to help me walk after spine surgery, or there to watch me run at a track meet. Parents that saw all four of their children through college, now three through PhD's, even though it was not an opportunity they themselves had. It has taken the form of siblings to look up to. Siblings that let me live in their houses for free or bought me a computer when I couldn't afford a new one. Family, loved ones, and friends that kept my feet grounded and my spirits high.

I have had the privilege of meeting and learning from many mentors both in and out of academia. My undergraduate research mentors Robert Szilagyi and Matt Queen at MSU were the first to teach me how scientific research works, and that it can actually be fun. The experiences I had in Robert's lab were a large factor in what drove me to continue on to a PhD. I found equal levels of mentorship during my Fulbright year, with Clemence Corminboeuf and Matt Wodrich at EPFL showing hospitality to the lost-looking American kid that showed up in their lab. That year I also met David Kastner, who has since been a friend and invaluable resource for helping me navigate the combined MD/PhD world.

My luck in finding outstanding mentors only increased as I entered my PhD program. David Eisenberg's excellence as a scientist is only matched by ability to teach. His teaching has manifested itself in many ways—either directly, through comments after presentations or on manuscript drafts, or indirectly, by observing how he fosters an atmosphere of community and

collaboration within his research group. He is proof that being an accomplished scientist does not have to come at the expense of patience or respect for those who work under him. David has also granted me a level of autonomy in my research that has helped me grow as an independent scientist. It will likely take me most of my career to fully unpack the wisdom he has displayed while I've been in his lab.

I have had the great fortune of learning from many others at UCLA, both in and out of the Eisenberg lab. Lin Jiang was instrumental early in my graduate studies in teaching me computation and continued to serve as fantastic mentor in the subsequent years, as we have collaborated on many projects. My other committee members, Todd Yeates, Jose Rodriguez, and Feng Guo have all provided guidance for my research both individually and during our formal committee meetings. If not for the patient teaching of Michael Sawaya, the protein structures reported in this document would not have been solved. This extends to Duilio Cascio as well. Alex Lisker was essential in the success of all computational projects. Likewise, Dan Anderson was an indispensable resource in the wet lab, not only for managing the lab so well, but also for his experienced input on many experiments.

My other colleagues in the lab, fellow graduate students or postdocs, were also critical to my success. Chiefly among them is Paul Seidler, who has been an excellent sounding board for ideas and a great teacher of experimental techniques, always acting gracious when I'd harass him with questions the moment he stepped into the office in the morning. David Boyer and I also had many helpful discussions, primarily on our way to lunch at the hospital cafeteria. Lukasz Salwinski was critical in teaching me many of the molecular biology techniques I now rely on. Undergraduate student Carolyn Hu has also been an excellent help in tackling large volumes of experimental work. Much of the phase separation/low complexity work was a product of

discussions with Michael Hughes. The rest of the Eisenberg group also provided useful input and suggestions along the way. I also would like to thank the UCLA-Caltech MSTP for their support and directors Carlos Portera-Cailliau and David Dawson. I also thank the Chemistry-Biology Interface Training Program and Heather Maynard for making it such an excellent program. I finally thank Audree Fowler for her generous donation to the Fowler Fellowship.

Chapter 1 is a draft manuscript by Kevin A. Murray, Carolyn J. Hu, Paul M. Seidler, Sarah L. Griner, Jeannette T. Bowler, Romany Abskharon, Gregory M. Rosenberg, Megan Bentzel, and David S. Eisenberg. The project was conceived and designed by KAM and DSE. KAM performed all computational designs and simulations. KAM and CJH expressed and purified MP inhibitors. KAM expressed and purified alpha-synuclein protein. SLG expressed and purified Aβ(1-42). KAM, CJH, GMR, and MB performed cellular seeding assays. KAM executed all ThT aggregation assays, performed fluorescence microscopy, ELISAs, and circular dichroism. JTB performed cellular toxicity assays. KAM performed C. elegans assays with assistance from RA. MB performed extractions from patient tissue samples. PMS performed electron microscopy. Manuscript was written by KAM and DSE with contributions from all other authors. In this work I was able to learn and apply techniques in computational protein design, protein expression/purification, microscopy, human cell culture, *C. elegans* culture.

Chapter 2 is a draft manuscript by Kevin A. Murray, Michael P. Hughes, Carolyn J. Hu, Paul M. Seidler, Qin Cao, Jiahui Liu, and David S. Eisenberg. The project was conceived and designed by KAM and DSE with assistance from MPH. KAM designed miniprotein inhibitors and performed accompanying calculations. Protein purification was carried out by KAM (miniproteins/FUS/aSyn), CJH (miniproteins), QC (TDP43), and JL (hnNRPA2). PMS performed electron microscopy. All experimental assays were done by KAM. Manuscript was

written by KAM and DSE with contributions from all other authors. In this work I was able to learn and apply techniques in computational protein design, protein expression/purification, microscopy, human cell culture, and phase separation biochemistry.

Chapter 3 is a draft manuscript by Kevin A. Murray, Michael P. Hughes, Carolyn J. Hu, Michael R. Sawaya, Lukasz Salwinski, Paul M. Seidler, Samuel W. French, and David S. Eisenberg. The project was conceived and designed by KAM and DSE. Identification of LARKS in proteome was done by MPH, identification of disease mutations within LARKS was done by LS, and assessment of steric zipper propensity for each mutation was measured by KAM. Crystallization and determination of KRT8 segment structures was performed by KAM with assistance from MRS. Protein purification was performed by KAM and CJH. Aggregation experiments, x-ray diffraction, SDS denaturation, liver tissue extraction was performed by KAM. Liver disease tissue samples were acquired and characterized with assistance from SWF. Electron microscopy was performed by PMS and KAM. Manuscript was written by KAM and DSE with contributions from all other authors. In this work I was able to learn and apply techniques in bioinformatics and protein structure analysis, protein crystallography and X-ray diffraction, protein expression/purification, microscopy, and human tissue extraction.

Chapter 4 is a draft manuscript by Kevin A. Murray, Declan Evans, Michael P. Hughes, Michael R. Sawaya, Kendall N. Houk, and David S. Eisenberg. Project was conceived and designed by KAM and DSE. KAM performed structure analysis. DE carried out all quantum calculations, with guidance from KH. MPH wrote script to calculate atomic distances. MPH and MRS crystallized and determined the Nup54 structure. Manuscript was written by KAM and DSE with contributions from all other authors. In this work I was able to learn and apply techniques in protein structure analysis.

VITA

2009-2012	B.S. in Chemistry, Montana State University
2014-2016; 2020-2022	M.D., University of California, Los Angeles
2012-2013	Health Policy staff member, U.S. Senate Finance Committee
2013-2014	U.S. Fulbright Scholar, Ecole Polytechnique Federal de Lausanne
2018-2019	Audree Fowler Fellowship in Protein Science
2017-2020	UCLA-NIH Chemistry Biology Interface Training Grant

Select Publications

K.A. Murray, C.J. Hu, P.M. Seidler, S.L. Griner, J.T. Bowler, M. Bentzel, M.R. Sawaya, G.M. Rosenberg, D.S. Eisenberg. *De novo* designed inhibitors of amyloid aggregation and seeding. *Manuscript in preparation*. (*Intended submission June 2020*).

K.A. Murray, M.P. Hughes, C.J. Hu, P.M. Seidler, J. Liu, Q. Cao, D.S. Eisenberg. Structure-based inhibition of FUS liquid-liquid phase separation using *de novo* protein design**.** *Manuscript in preparation*. (*Intended submission July 2020*).

K.A. Murray, M.P. Hughes, C.J. Hu, P.M. Seidler, L. Salwinski, M.R. Sawaya, D.S. Eisenberg. Discovering amyloid-related diseases by mapping mutations in low-complexity protein domains to known pathologies: the case of Keratin-8 in alcoholic liver disease**.** *Manuscript in preparation*. (*Intended submission June 2020*).

K.A. Murray, D. Evans, M.P. Hughes, C.J. Hu, M. Sawaya, K. Houk, D.S. Eisenberg. Extended beta-sheets contribute to reversible amyloid formation. *Manuscript in preparation*.

R. Abskharon, P.M. Seidler, **K.A. Murray**, M.R. Sawaya, M. Bentzel, D.R. Boyer, D. Cascio, P. Sieminski, E.L.S. Pardon, J. Steyaert, C.G. Glabe, D.S. Eisenberg. De novo design of antibodies that block tau seeding by Alzheimer's disease brain tissue**.** *Manuscript in preparation*

A. Hatami, B. Li, **K.A. Murray**, W.S. Shin, V. John, L. Jiang. Alternative protofilamental packing arrangements of α-synuclein fibril polymorphs contribute to seeding differences in cells and in brain. *Submitted*.

W.S. Shin, C. Sun, P. Lin, J. Xu, **K.A. Murray**, T. Olafsen, J.T. Lee, W. Hong, L. Chen, L. Jiang. Computational design of a fluorescent acetylcholine receptor binding protein that enters the brain and penetrates neurons. *In Review*.

S. Sangwan, S. Sahay, **K.A. Murray**, S. Morgan, E. Guenther, L. Jiang, C.K. Williams, M. Goedert, D.S. Eisenberg. Inhibition of synucleinopathic seeding by rationally designed inhibitors. *eLife*. **2020**.

W.S. Shin, **K.A. Murray**, C. Sun, G. Bitan, L. Jiang. Different amyloid-beta self-assemblies have distinct effects on intracellular tau aggregation. *Frontiers in Molecular Neuroscience*. **2019**.

W.S. Shin. C. Sun, **K.A. Murray**, B. Li, G. Bitan, L. Jiang. Amyloid-beta protein oligomers promote the uptake of tau fibril seeds potentiating intracellular tau aggregation. *Alzheimer's Research & Therapy*. **2019**.

P.M. Seidler, D.R. Boyer, **K.A. Murray**, T.P. Yang, M. Bentzel, M.R. Sawaya, G. Rosenberg, D. Cascio, C.K. Williams, K.L. Newell, B. Ghetti, M.A. DeTure, D.W. Dickson, H.V. Vinters, D.S. Eisenberg. Structure-based inhibitors halt prion-like seeding by Alzheimer's disease- and tauopathy-derived brain tissue samples. *Journal of Biological Chemistry*. **2019**.

S.L. Griner, P. Seidler, J. Bowler, **K.A. Murray**, T.P. Yang, S. Sahay, M.R. Sawaya, D. Cascio, J.A. Rodriguez, S. Phillip, S. Sonsa, C.G. Glabe, T. Gonen, D.S. Eisenberg. Structure based inhibitors of amyloid beta core suggest a common interface with tau. *eLife*. **2019**.

Q. Cao, W.S. Shin, H. Chan, C.K. Vuong, B. Dubois, B. Li, **K.A. Murray**, M.R. Sawaya, J. Feigon, D.L. Black, D.S. Eisenberg, L. Jiang. Inhibiting amyloid-β cytotoxicity through its interaction with the cell surface receptor LilrB2 by structure-based design. *Nat. Chemistry*. **2018**.

B. Li*, P. Ge*, **K.A. Murray***, P. Sheth, M. Zheng, G. Nair, M.R. Sawaya, W.S. Shin, D.R. Boyer, S. Ye, D.S. Eisenberg, Z. H. Zhou, L. Jiang. Cryo-EM of full-length alpha-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications*. **2018**. *(*authors contributed equally*)

S. Sangwan, M.R. Sawaya, **K.A. Murray**, D.S. Eisenberg. Atomic structures of corkscrew-forming segments of SOD1 reveal varied oligomer conformations. *Protein Science*. **2018**.

P.M. Seidler, D.R. Boyer, JA Rodriguez, M.R. Sawaya, D. Cascio, **K.A. Murray**, T. Gonen, D.S. Eisenberg. Structure-based inhibitors of tau aggregation. *Nature Chemistry*. **2018**.

P. Krotee, S. Griner, M. R. Sawaya, D. Cascio, J. A. Rodriguez, D. Shi, S. Philipp, **K. A. Murray**, J. Lee, P. Seidler, C. G. Glabe, L. Jiang, T. Gonen, D. S. Eisenberg. Common Fibrillar Spines of Amyloid-β and hIAPP Revealed by MicroED and Inhibitors Developed Using Structure-Based Design. *Journal of Biological Chemistry*. **2017**.

K.A. Murray, M.D. Wodrich, X.L. Hu, C. Corminboeuf. Toward Functional Type III [Fe]-Hydrogenase Biomimics for H2 Activation: Insights from Computation. *Chem. Eur*. **2015**.

M.S. Queen, B.D. Towey, **K.A. Murray**, B.S. Veldkamp, H.J. Byker, R.K. Szilagyi. Electronic structure of [Ni(II)S4] complexes from S K-edge X-ray absorption spectroscopy. *Coordination Chemistry Reviews*. **2013.**

CHAPTER 1

De novo designed inhibitors of amyloid aggregation and seeding

INTRODUCTION

The aberrant aggregation of proteins into amyloid fibrils is a hallmark of many neurodegenerative diseases, including Alzheimer's disease (AD) and Parkinson's disease (PD)[1]. In AD, amyloid-β (Aβ) and tau amyloid fibrils comprise the extracellular amyloid plaques and intracellular neurofibrillary tangles, respectively, characteristic of disease progression[2]. Likewise, intracellular Lewy bodies found in the neurons of PD patients are primarily made up of αSyn fibrils[3]. There are currently no therapies capable of slowing or stopping the progression of either of these diseases, and inhibition of fibril formation has become a major target for therapeutic development[4,5]. Amyloid fibrils are composed of repeating layers of β-sheet rich protein monomers stacked upon each other. The β-sheets interdigitate to form a stable fibril core known as a steric zipper[6]. Anti-amyloid therapies have typically focused on small molecules that prevent aggregation or dissociate already formed aggregates and antibodies that promote fibril clearance. An alternative approach is the design of molecules that bind to the ends of the growing fibrils, capping their growth and preventing the further addition of more protein monomers[7]. This approach has been successfully used to design peptide-based inhibitors of tau, Aβ, and αSyn aggregation[8–11]. This design strategy takes into account the atomic structures of fibrils, employing rational and computational design techniques to derive a peptide sequence complementary to the growing fibril surface.

Since the initial work to design structure-based capping inhibitor peptides, many advances have been made in both the determination of amyloid protein structure of as well as the field of protein design. The first atomic-resolution structures of amyloid fibrils determined by x-ray crystallography were restricted to small peptide segments ~6-11 amino acids in length[12]. The recent advent of cryo-electron microscopy (cryo-EM), micro electron diffraction (MicroED), and solid-state nuclear magnetic resonance (ssNMR) spectroscopy have enabled the determination of amyloid protein structures that were previously unsolvable[13–15]. These techniques have been used to solve an ever-growing list of structures of both recombinantly-derived fibrils[16–18] as well as fibrils directly extracted from patient tissue[19–21]. These structures have provided key insights into fibril architecture and polymorphism in relation to disease.

Like the structural knowledge of amyloid fibrils, the toolbox of protein structure prediction and design has been rapidly expanding in recent years[22]. Significant advances in algorithms and computing power have facilitated the *de novo* design of proteins with a variety of properties and functions, ranging from stability, pH sensitivity, to even logic operations, with vast potential for use in therapeutics, diagnostics, etc[23–25]. While the underlying design principles of *de novo* generated proteins are becoming well established, only a few examples of their direct application into biological systems have been demonstrated thus far. In this work, we use *de novo* protein design to create 35-50 residue miniproteins which bind to the growing ends of tau, αSyn, and Aβ fibrils. We target recently determined full-length atomic structures of each amyloid protein in our designs to generate inhibitors capable of preventing aggregation, seeding, and toxicity both *in vitro* and *in vivo*.

RESULTS

High-throughput computational design pipeline

Three amyloid proteins were chosen for inhibitor design: tau, αSyn, and Aβ (Figure 1.1a). Atomic-level fibril structures of all of these proteins have been determined, including short peptide segments determined through crystallographic methods and full-length structures solved by cryo-EM/ssNMR. For tau, the paired helical filament (PHF) structure derived from AD patient brains was selected[19]. The PHF structure is the dominant tau fibril morphology in AD brains and this structure is conserved in fibrils extracted from the brains of different patients[26]. For αSyn, the rod polymorph from Li et al. was selected[17]. Cryo-EM structures of αSyn reported from multiple groups have demonstrated the rod polymorph to be the dominant fibril morphology[27–29]. A recent study of brain-derived αSyn fibrils shows that the core structural elements of the αSyn rod polymorph are still maintained in patients with the synucleinopathy multiple system atrophy (MSA)[30]. For Aβ, two structures were used for inhibitor design. One is the disease-relevant full-length structure from Riek and colleagues[31], the other is a segment from amyloid β (residues 16-23) with the D23N Iowa mutation which has been successfully used to design peptide-based inhibitors[11].

The design of the inhibitor library was performed using the software suite Rosetta. Backbone topologies were generated using the blueprint format of RosettaRemodel to guide Rosetta's Monte Carlo-based fragment assembly, followed by the FastDesign algorithm to generate the amino acid sequence[32,33]. For the inhibitor scaffolds, miniproteins, 35-50 amino acids in length, were designed into seven unique classes, with each class differing by the arrangement of secondary structural elements (Figure 1.1b). Each inhibitor topology contains one α-helix (H) and two or three β-strands (E), yielding the seven classes: HEE, EHE, EEH, HEEE, EHEE, EEHE,

and EEEH. To mimic the natural interactions found in amyloid fibrils, the primary interaction between the inhibitors and fibrils is the stacking of a β-strand of the inhibitor onto the β-strand of the growing fibril end (Figure 1.1c). Because of this, only inhibitor topologies with at least one β-strand were selected. Several classes of inhibitors composed exclusively of β-sheets (EEE-EEE) were tested, but did not yield consistently stable designs, highlighting the importance of the stabilizing α-helix found in each of the selected classes. In addition, as the final aim of these inhibitors is to therapeutically target largely intracellular protein aggregates found, disulfide bonds were not incorporated into the designs, as has been done in previous studies, because of the reducing conditions of the cytoplasm[34]. 5000 unique scaffolds were generated for each inhibitor class. Inhibitors were docked onto the fibrils using Rosetta's MotifGraft protocol, creating a backbone alignment of a selected portion of the native fibril strand with the β-strand of the inhibitor[35]. Once docked to the fibril, the inhibitor sequence along the binding interface was optimized to improve binding energy.

The tips of each amyloid fibril structure used in this study are nearly flat, open β-sheet rich surfaces. Because of this, no obvious binding cavity exists, and it is unclear whether any particular segment of the surface is the most important for fibril aggregation. Because of this, we chose to systemically select each possible segment of the fibril ends as an inhibitor binding site (Figure 1.2a). Nine, nine, and five unique sites were selected for the tau, αSyn, and Aβ structures, respectively. Inhibitors from each of the seven classes were docked and sequence optimized to each site in an all-vs-all fashion, yielding ~1 million unique inhibitor sequences. Docked poses were ranked by several scoring metrics pertaining to both the stability of the inhibitor/fibril interaction (ddg, number of unsaturated hydrogen bonds, number of atoms in interface) as well as

the stability of the inhibitor alone (total score, p_aa_p). These metrics were demonstrated to be predictive of successful designs in work by Chevalier et al. to create miniprotein binders[34].

To further validate the fold and stability of each inhibitor, several more rigorous computational steps were incorporated into the screening pipeline. First, long-range molecular dynamics (MD) simulations of unbound inhibitors were performed to measure stability in a dynamic system (Figure 1.1d). Simulations in an explicit cubic water box were carried out in GROMACS 2018 for 200 ns[36]. Those inhibitors that showed stability after the initial MD run, as measured by a low root-mean-square deviation (RMSD) from the starting configuration, were tested for an additional 400 ns. The folds of the most stable inhibitor designs were then subjected to Rosetta's fragment-based *ab initio* folding algorithm, which predicts protein structure based on primary sequence[37]. The sequences of top-ranking inhibitor designs were provided to the prediction algorithm, and 50,000 trajectories were calculated per sequence. The lowest energy trajectories converging on a conformation close in backbone RMSD to the original design indicates a stable fold (Figure 1.1e). Inhibitors whose 50 lowest energy trajectories had the lowest RMSD values were then selected for experimental characterization.

Biophysical characterization of inhibitor proteins

From the top-ranking inhibitor sequences selected for experimental testing, 46 soluble miniproteins were expressed and purified (Supplementary Table 1.1). The inhibitors were expressed in *E. coli* using a pET28b(+) vector with a thrombin-cleavable N-terminal His-tag. Miniproteins were purified using a Ni-NTA affinity column followed by size-exclusion chromatography (see Methods). Beyond computational prediction, circular dichroism (CD) spectroscopy was used to verify that each miniprotein adopts a stable fold (Figure 1.2b-d). One selected tau inhibitor, iTau-N, was treated with increasing concentrations of the denaturant

guanidinium hydrochloride (GdnHCl) to assess its stability (Figure 1.2e). No significant changes in its CD spectrum were observed at 1M GdnHCl, and the CD signal gradually diminishes with as GdnHCl concentration increases. Lacking any disulfide bonds, the stability of each inhibitor fold is likely derived from its tightly packed hydrophobic core region, with charged and polar residues decorating the miniprotein exterior (Figure 1.2f). The sequences of each tested inhibitor are truly *de novo*, lacking apparent homology to natural protein sequences. BLAST E-values, the statistical term indicating sequence homology, for each inhibitor are well below the significance threshold, except for iAβ-F, which holds some happenstantial sequence similarity to a universal stress family protein from *P. claussenii* (Figure 1.2g)[38].

We next sought to assess the effects of the inhibitors on primary amyloid aggregation. Thioflavin T (ThT) kinetics assays were performed for tau, αSyn, and Aβ in the presence of inhibitors. For the tau assays, monomeric tau k18+ (Q244-E380), which contains the core observed in the PHF structure, was used at 50 μM concentration. For αSyn, 50 μM of full-length αSyn was used, and for Aβ, 10 μM of Aβ(1-42) was used. Screening of each inhibitor with its designed target (Supplementary Figures 1.1-1.3) identified constructs capable of completely abolishing or significantly delaying aggregation. At a 2:1 stoichiometric ratio (inhibitor:tau), the inhibitor iTauP leads to a 4-fold increase in the lag time needed for tau k18+ to aggregate (Figure 1.3a). The αSyn inhibitor iαSyn-F prevents αSyn aggregation at even sub-stoichiometric ratios, with only minimal aggregation observed at 1:5 inhibitor to α-syn ratio (Figure 1.3b). Likewise, at equimolar ratios, the Aβ inhibitor iAβ-H prevents the aggregation of Aβ(1-42) (Figure 1.3c). Negative stain transmission electron micrographs of αSyn and Aβ(1-42) illustrate that the iαSyn-F and iAβ-H fully prevent fibril formation *in vitro* (Figure 1.3d-e). The effects of each inhibitor are also specific

to the amyloid protein they are designed to target. iαSyn-F and iAβ-H have little effect on tau k18+ aggregation, while iTau-P and iAβ-H show little effect on αSyn aggregation (Figure 1.3g).

To gain an approximate measurement of binding affinity each inhibitor has for their target fibrils, we performed ELISA binding assays. Fibrils of tau, αSyn, or Aβ were coated onto the bottom of a plate then incubated with inhibitors with their N-terminal His-tag uncleaved. Following this, an anti-His antibody conjugated with an Alexafluor647 fluorescent dye was added. Fluorescent measurement of the samples shows the inhibitors reach a binding saturation in the low-hundreds nM range (Figure 1.3g; Supplementary Figure 1.4). To validate if the binding to the inhibitors truly occurs at the fibril tips, we performed a nanogold binding assay in conjunction with electron microscopy. Similar to the ELISA assay, an EM grid was coated with tau fibrils seeded with AD patient brain tissue (see Methods), then treated with the inhibitor iTau-N with its His-tag intact. The sample was then treated with an anti-His primary antibody followed by secondary antibody conjugated to a gold nanoparticle, which is highly visible by EM. As shown in Figure 1.3h, gold nanoparticles complexed to iTau-N can be seen binding to the tips of tau fibrils (Additional images in Supplementary Figure 1.5), highlighting the binding to the fibrils is consistent with the intended design.

Effects of inhibitors on amyloid seeding and toxicity

Amyloid pathology is believed to spread throughout the brain via a process known as templated seeding. Fibrillar aggregates that form in one cell migrate to adjacent cells and seed the subsequent aggregation of additional soluble protein. Having demonstrated the effects of the designed inhibitors on primary amyloid aggregation *in vitro*, we next aimed to assess their ability to prevent seeding in cells. HEK293T biosensor cells overexpressing either tau k18 or αSyn fused with green fluorescent protein (GFP) were used in the seeding assays (Figure 1.4a). In nontreated

cells, the endogenous GFP-fused amyloid proteins remain soluble and unaggregated, visible as diffuse fluorescence throughout the cell. Upon the addition of an exogenous fibril seed, either recombinant or brain-derived, the fluorescent endogenous protein becomes incorporated into the fibrillar form. This is visualized as bright fluorescent puncta forming within the cell. For tau seeding, AD patient brain extract was incubated with each inhibitor (10 μM final inhibitor concentration) overnight, then the mixture was applied to the tau biosensor cells. 48 hours later, the cells were imaged, and the number of fluorescent puncta were quantified (Supplementary Figure 1.6). Several inhibitors showed a significant reduction in the number of seeded puncta, including iTau-D and iTau-N (Figure 1.4b-c), with inhibitory effects observed in the nanomolar range. Similarly, seeding αSyn biosensor cells with recombinant αSyn fibrils incubated with the panel of αSyn inhibitors identified many inhibitors with inhibitory effects (Supplementary Figure 1.7), including iαSyn-E and iαSyn-F (Figure 1.4d-e), both of which also show nanomolar range efficacy.

To further validate the binding mechanism of the designed inhibitors, we introduced negative control mutations into iTau-N at the binding site of iTau-N with the tau PHF (Supplementary Figure 1.8). These mutations create large steric clashes meant to disrupt binding, altering two key hydrophobic residues found within the iTau-N/PHF interface, Ala42 and Val44. Introducing mutations A42R and V44Y do not disrupt the overall computed fold of the inhibitor, however they greatly affect the ability of iTau-N to reduce biosensor cell seeding with AD brain extract. The single A42R mutation more than halves the inhibitory capacity of iTau-N, and the double mutant A42R/V44Y completely abolishes inhibition.

Aggregated Aβ(1-42), particularly oligomers, has been shown to be neurotoxic to neurons[39]. To test if the designed inhibitors mitigate the toxic effects of Aβ, inhibitors were

incubated with Aβ(1-42) overnight at 37°C, then applied to cultured N2a neuronal cells, to a final Aβ(1-42) concentration of 1 μM. A 3-(4,5-dimethylthiazol-2-yl)−2,5-diphenyltetrazolium bromide (MTT) dye reduction cell viability assay of the treated N2a cells reveals that inhibitors iAβ-H and iAβ-D can rescue Aβ(1-42) cytotoxicity at equimolar inhibitor:Aβ ratios. αSyn aggregates are also toxic to neurons, and an MTT assays of N2a cells treated with 1 μM αSyn fibrils reveals that iαSyn-F can effectively rescue cytotoxicity (Supplementary Figure 1.9)[40].

***In vivo* evaluation of inhibitors**

To evaluate the effects of the designed miniprotein *in vivo*, inhibitors were tested in *C. elegans* model strains of tau and αSyn aggregation. *C. elegans* strain DDP1 overexpresses αSyn fused with either yellow or cyan fluorescent protein (YFP, CFP). Extensive amounts of aggregation can be visualized in the adult worms. Day 4 adults of synchronized DDP1 cultures of DDP1 worms were administered iαSyn-F for 8 hours using cationic lipids (Figure 1.5a). The amount of aggregated fluorescent αSyn was then quantified in day 6 adults by fluorescent microscopy (Figure 1.5b-c). iαSyn-F treatment largely reduces then number of visible αSyn aggregates in the worm head regions. *C. elegans* strain BR5706 co-expresses tau with the pro-aggregation V337M mutation and tau F3ΔK280, an aggregation prone fragment from the tau repeat domain with K280 deleted, leading to increased levels of insoluble tau species and locomotion deficits. Synchronized cultures were treated with iTau-N at the L4 larval stage using cationic lipids for 8 hours. The following day, worm locomotion was tracked and analyzed, and insoluble tau was extracted from lysed worms and analyzed by Western blot. iTau-N treated worms showed a recovery in locomotion speed compared to vehicle treated control (Figure 1.5f). Anti-tau antibody staining reveals a significant reduction in aggregated tau species in the detergent soluble (RIPA buffer) fraction.

DISCUSSION

Structure-based drug design has generated effective therapeutics for a variety of diseases, from HIV to cancer[41]. However, lack of complete structural knowledge has hampered the application of this approach to amyloid diseases. In this work, we have used atomic structures of full-length amyloid proteins, determined through cryoEM, ssNMR, and MicroED, to generate targeted inhibitor molecules capable of preventing amyloid aggregation *in vitro*, in cells, and *in vivo*. Each of these amyloid structures is a product of significant advancements that have occurred recently in the field of protein structure determination. This study represents a systematic endeavor to utilize these next-generation amyloid structures for the design of protein drugs.

This work is also a product of the simultaneous advancement in the field of computational protein design occurring in recent years. Proteins designed *de novo* in a computer are beginning to reliably resemble their computed folds and functions more and more frequently, allowing us to leverage their unique properties into biological systems[22]. Here we have chosen miniproteins as the scaffold for inhibitor design because of their extremely high stabilities and amenability to high-throughput screening. A unique challenge with designing amyloid capping inhibitors is how to screen for successful designs. Traditional techniques to screen for protein binders typically involve display assays, such as yeast or phage display, to identify sequences with enriched binding. However, applying such a technique to screen for capping proteins would likely only identify sequences that tightly bind to the fibrils but not necessarily prevent their growth. Because of this, we primarily relied on computational techniques to funnel down the ~1 million unique inhibitor sequences to a pool which could be feasibly expressed, purified, and tested experimentally. Extensive 200-400 ns MD simulations were useful in identifying scaffolds with good computed binding energies but were unstable for extended periods of time, as has also been described by

Buchko et. al. Likewise, the *ab initio* structure prediction from primary sequences proved useful for finding designs which appear stable but whose sequence is more likely to adopt a completely different fold. Additionally, our design process employed a shotgun-like approach in terms of inhibitor binding sites. Inhibitors were docked in many different positions along the fibril chain. The ranking and selection of inhibitors was done completely blind to their binding position, allowing the computed metrics to identify the optimal inhibitor/binding site pair in an unbiased manner. The binding sites for the successful inhibitors may prove useful starting points for future structure-based design approaches. Successful designs iTau-N and i αSyn-F both bind near the beta arch region on their targeted structures, suggesting this motif may be an important target.

We demonstrate the ability of the designed miniprotein inhibitors to prevent primary amyloid aggregation, as measured through the ThT kinetics assays, as well as secondary seeding in the biosensor cell assays. There is increasing evidence that templated seeding is a primary driving force for the spread and progression of amyloid pathology in the brain, both in the case of tau and αSyn. Interestingly, comparison of the effects the tau and αSyn inhibitors have on primary aggregation versus seeding shows that the tau inhibitors that best reduce primary aggregation do not reduce seeding and vis versa (Supplementary Figure 1.10). On the other hand, αSyn inhibitors that reduce αSyn primary aggregation are also effective at reducing cellular seeding. This could be a result of some underlying phenomenon that distinguishes the aggregation pathways between tau and αSyn. However, this effect may also be attributed to a difference in the protein material used, as tau ThT assays used recombinant protein while the seeding assays used patient-derived material, and the αSyn assays used recombinant material for both.

De novo designed proteins, particularly miniproteins, represent a new forum for therapeutic development. Their small size, stability, and ease of expression present some benefits over

established antibody therapeutics, while their adaptability and designability leverages advantages over small molecules. Further improvements to the inhibitor designs presented in this work could focus on improved binding, perhaps through protein evolution techniques, as well as methods of delivery. The eventual goal of these designed miniprotein inhibitors is to enter the brain, where amyloid fibrils are found in AD and PD. This warrants future work on delivery strategies, including fusion of successful inhibitors with cell-penetrating peptide tags or conjugation to larger delivery constructs such as brain-penetrating nanoparticles. The treatment of neurodegenerative disease is a problem that has yet to be solved by conventional therapeutic approaches, opening a wide field for discovery and adaptation of modern techniques. Computational protein design may prove to be a useful method for addressing these challenges.

MATERIALS AND METHODS

Protein expression and purification

Miniprotein inhibitors were expressed using a pET28b(+) plasmid in BL21(DE3)gold *E. coli*. Cultures were grown at 37ºC to an OD600 = 0.4-0.8, then induced with 0.5 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG), then grown overnight at 18oC. Cells were centrifuged and pellets collected for purification. For inhibitors with solubility issues, constructs were transformed into Lemo21(DE3) *E. coli*., in which protein expression levels could be more finely tuned using optimized concentrations of L-rhamnose; transformation and induction were carried out according to the manufacturer's protocol (New England Biolabs).

Cells were lysed by sonication, and the clarified lysate ran over a Ni-NTA affinity column (lysis buffer: 50 mM Tris pH 7.5, 300 mM NaCl, 15 mM imidazole, 0.1 mg/mL lysozyme, HALT protease inhibitor cocktail, 20 mM BME). Column was washed with 10 column volumes of Wash

1 buffer (50 mM Tris pH 7, 1 M NaCl, 15 mM imidazole, 20 mM BME), followed by 20 column volumes of Wash 2 buffer (50 mM Tris pH 6, 1 M NaCl, 15 mM imidazole, 20 mM BME). Proteins were eluted and fractions collected (50 mM Tris pH 7.5, 150 mM NaCl, 500 mM imidazole, 20 mM BME). The N-terminal His-tag was then cleaved using bovine thrombin, and proteins were further purified using size-exclusion chromatography (SEC) and dialyzed into 1x PBS.

Tau k18+ (residues Q244-E380), αSyn, and Aβ(1-42) were expressed and purified as previously described.

Preparation of AD brain tissue extract

Brain tissue from the inferior temporal gyrus of a histopathologically confirmed case of AD (69 y.o. female; post-mortem interval time: 12.9 hours). A section of ~250 mg of tissue was cut on dry ice then homogenized in a 15 mL disposable tube using 50 mM Tris (pH 7.4), 150 mM NaCl, 1X HALT protease inhibitor cocktail. Sample was sonicated in a cuphorn water bath at 4^{o}C, for 2 hours at 30% power. Samples were frozen at -80^{o}C until experimental use.

Computational design pipeline

Design of the miniprotein scaffold library follows the approach previously described by Chevalier et. al[34]. Briefly, scaffolds with mixed α and β topologies (containing at least two β-strands) were designed using the blueprint format of RosettaRemodel. No disulfide bonds were allowed. Sequences for each scaffold were designed with Rosetta's FastDesign. 15,000 unique designs for each of the 7 topologies (EEEH, EEHE, EHEE, HEEE, EEH, EHE, HEE) were generated, and the top 5,000 designs from each topology were kept, filtering by overall score, β-sheet backbone dihedral angles, and number of unsaturated hydrogen bonds.

Scaffolds were docked onto the different binding sites of each amyloid structure using Rosetta MotifGraft. Scaffolds were aligned to a truncated strand of the native fibril using a backbone

RMSD < 1 Å, filtering by clash score. Residues at the binding interface were then optimized, favoring the native residue sequence for each scaffold. HotSpot residues were not enabled. All Rosetta calculations were performed using the REF2015 score function with the -β option enabled. Stability of each scaffold after Rosetta design and optimization was assessed using molecular dynamics simulations. MD simulations were performed using GROMACS version 2018 and the CHARMM27 all-atom forcefield. Proteins were solvated in a cubic water box using periodic boundary conditions with counter ions added. Systems were energy minimized then temperature and pressure equilibrated for 100 ps. Production runs were carried out for 200 ns or 400 ns. Calculations of non-bonded interactions were gpu accelerated.

Folding of inhibitor designs were assessed with Rosetta's AbinitioRelax application, using fragments generated from the Robetta server. 50,000 trajectories were calculated for each design tested.

***In vitro* aggregation assays**

Thioflavin T (ThT) based aggregation kinetics assays were performed in Nunc black 96 well optical bottom plates (Thermo Scientific) in a microplate reader taking fluorescence measurements every 5 minutes (FLUOstar OMEGA, BMG Labtech). All assays were carried out at 37 oC in 1x PBS buffer and 50 μM ThT at a final well volume of 100 uL. For tau aggregation, 50 μM of tau k18+ (residues Q244-E380), 1 mM DTT, and 0.225 mg/mL heparin sulfate were used. For αSyn aggregation, 50 μM full-length αSyn was used. For both the tau and αSyn assays, PTFE beads (0.125-inch diameter) were used to agitate the sample, and plates were shaken at 700 rpm with double orbital rotation. For Aβ, 10 μM Aβ(1-42) was used and samples were aggregated quiescently.

3-(4,5-dimethylthiazol-2-yl)−2,5-diphenyltetrazolium bromide (MTT) dye reduction cell viability assay

Clear 96-well plates (Costar) were plated with N2a cells in 90 uL culture media at 5000 cells/well and allowed to adhere to the plate for 24 hours. Aβ(1-42) samples were co-incubated in the presence or absence of inhibitor overnight at 37°C in 10 uL volume, then added to the N2a cells (final Aβ(1-42) concentration of 1 μM). Following an incubation of 24 hours, 20 μL of Thiazolyl Blue Tetrazolium Bromide MTT dye (Sigma; 5 mg/mL stock in DPBS) was added to each well, incubated at 37°C for 3.5 hours, . Assay was halted by replacement of all well media with 100% DMSO and removal from incubator. Next, a SpectraMax M5 reader was used to measure absorbance at 570 nm, and a background reading taken at 700 nm which was subtracted from the 570 nm reading. Vehicle alone treated cells were designated as 100% viable, while that treated with 100% DMSO were designated as 0% viable, and all other well readings were normalized to these values.

***C. elegans* experiments**

The following strains were acquired from the Caenorhabditis Genetics Center (CGC) and used for experiments: BR5706 (byIs193 [rab-3p::F3(delta)K280 + myo-2p::mCherry]; bkIs10 [aex-3p::hTau V337M + myo-2p::GFP]), and DDP1 (uonEx1 [unc-54::αSyn::CFP + unc-54::αSyn::YFP(Venus)]).

C. elegans strains were grown and maintained using standard conditions. Hypochlorite bleaching was used to synchronize the worms, and they were hatched overnight at 16°C in M9 media (5 g/l NaCl, 6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 1 μM $MgSO_4$), then cultured on plates with nematode growth medium (NGM; 17 g/l agar, 7.5 g/l casein, 3 g/l NaCl, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 250 μM KH_2PO_4 pH 6, 5 μg/ml cholesterol) seeded with OP50 *E. coli*. Strains were

maintained at 16ºC. Inhibitor proteins were delivered using the cationic lipid Lipofectamine 2k. 50 uL of Lipofectamine 2k was incubated with 20 μM of inhibitor in 1x PBS to a final volume of 1 mL overnight. Worms were then added to the solution and incubated for 8 hours. Following incubation, worms were transferred onto NGM plates containing 1 μM 5-fluoro-2′deoxy-uridine (FUDR) to prevent growth of progeny.

For the αSyn strain (DDP1), inhibitors were added to day 4 adults and analysis was performed at adult day 6. Worms were mounted onto glass slides, immobilized with 100 nm polystyrene beads, and with imaged by fluorescent microscopy (GFP channel) using a ZEISS Axio Observer D1 fluorescence microscope. ASyn aggregates in the head region were counted. For the tau strain (BR5706), inhibitors were added at larval stage 4, and worms were analyzed at adult day 1. Locomotion was tracked using a Leica M205 C light microscope in 30 second intervals. Tracking data was processed using WormLab software (MBF Bioscience).

Insoluble tau levels between treated and untreated BR5706 groups were analyzed using the protocol established by Fatouros et. al. Briefly, worms were removed from plates with M9 buffer, and a 30% sucrose solution was used to remove dead animals and bacteria by flotation. Worms were then pelleted and resuspended in RAB buffer (100 mM MES, 20 mM NaF, 1 mM EGTA, 0.5 mM MgSO4). Worms were lysed in a cuphorn bath sonicator for 5 minutes at 30% power. Lysate was centrifuged at 40,000g for 30 minutes. The pellet was extracted with RAB buffer containing 1 M sucrose, then centrifuged for 20 minutes at 40,000g. The pellet was subsequently extracted with RIPA buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 1% Nonidet P-40, 0.5% deoxycholate, and 0.1% SDS) and centrifuged at 40 000g for 20 min. The supernatant was then loaded onto a NuPAGE 12% Bis-Tris pre-cast protein gel and ran at 200V for 35 minutes. For Western blot analysis, iBLOT2 dry blotting system was used to transfer protein from the gel to a

nitrocellulose membrane. Membrane was blocked with 5% milk in TBST for 1 hour, then washed three times with TBST. The membrane was incubated with the primary antibody (anti-tau A0024 (Dako) 1:1000 dilution in 2% milk/TBST solution) for 3 hours, washed three times with TBST, incubated with the horseradish peroxidase-conjugated secondary antibody (goat anti-mouse IgG H and L (HRP); 1:1000 Dilution in 2% milk/TBST), and washed three times in TBST. The signal was detected with Pierce ECL Plus Western Blotting Substrate (Cat # 32132), and imaging was performed with a Pharos FX Plus Molecular Imager. Actin was subsequently measured using the same protocol with a B-Actin (C4) primary antibody (Santa Cruz Biotechnology) (1:250 dilution, 2% milk/TBST).

Transmission electron microscopy

6 uL of fibril samples (taken from *in vitro* aggregation experiments) were spotted onto Formvar Carbon film 400 mesh copper grids (Electron Microscopy Sciences) and incubated for 4 minutes. Grids were stained with 6 uL uranyl acetate solution (2% w/v in water) for 2 minutes. Excess solution blotted off, and grids were dried for 30 minutes. TEM images were taken using a JEOL 100CX TEM electron microscope at 100 kV.

Nanogold particle binding

Tau k19+ fibrils seeded with AD patient extract were generated by boiling 350 μM tau k19+ in 100 mM 2-mercaptoethanol for 5 minutes in 1x PBS (pH 7.4). Sonicated AD tissue extract was added to the samples and aggregation was induced by shaking at 350 rpm at 37 ºC overnight. 6 uL of the seeded fibrils were spotted onto Formvar Carbon film 400 mesh copper grids (Electron Microscopy Sciences) and incubated for 3 minutes. Grids were blotted by quickly dabbing onto filter paper, then blocked for 15 minutes with a 1% gelatin solution in PBS. Grids were blotted and 10 uL of the primary antibody (anti-His tag, HIS.H8; 1:100 dilution in 1% gelatin-PBS) was

applied and incubated for 20 minutes. Grids were washed and blotted five times. 10 uL of secondary antibody (goat anti-mouse IgG H&L (20nm gold) preadsorbed antibody in PBS) was applied for 10 minutes then blotted. Grids were then washed five times with PBS, then five time with deionized water. Grids were stained with 6 uL uranyl acetate solution (2% w/v in water) for 2 minutes, blotted, then air dried for 30 minutes. TEM images were acquired with a JEOL 100CX TEM electron microscope at 100 kV.

Circular dichroism and denaturation assay

Circular dichroism measurements were performed with a JASCO J-715 spectrophotometer in a crystal cuvette (path-length of 1mm) at a concentration of 0.20 mg/mL in 1x PBS buffer (pH 7.4). For denaturation experiments, guanidinium hydrochloride was added to protein samples with final concentrations from 0-6 M and incubated for 20 minutes before measurement.

Cell seeding assays

HEK293T biosensor cells stably expressing tau-k18 or αSyn fused with YFP, developed by Diamond and colleagues at UTSW were utilized without additional authentication or characterization. Cells were grown in a humidified incubator at 37 °C, 5% CO_2 in DMEM (Life Technologies, cat. 11965092) with FBS (10% vol/vol; Life Technologies, cat. A3160401), penicillin/streptomycin (1%; Life Technologies, cat. 15140122), and Glutamax (1%; Life Technologies, cat. 35050061). Inhibitors were incubated with recombinant fibrils or patient tissue overnight in OptiMEM media before being applied to ~70% confluent biosensor cells. To seed cells, the coincubated inhibitor/fibrils were sonicated for 3 minutes in a cuphorn water bath and mixed with a 1:20 dilution of Lipofectamine 2000 in OptiMEM for 20 minutes. 10 uL of the inhibitor/fibril + Lipofectamine mixture was added to 90 uL of cells plated in black 96-well tissue culture plates in triplicate for each concentration of inhibitor tested. Quantification of seeded

aggregates was performed using a Celigo Image Cytometer (Nexcelom) in the YFP channel, imaging the entire 96-well plate. Images were processed in ImageJ, subtracting background fluorescence from unseeded cells, and using the Particle Analyzer function to count the number of particles above background. Total number of aggregates in each well was normalized by cell confluence. Standard deviation between triplicates and IC50 values for dose response curves were calculated using a nonlinear regression curve. To obtain high quality fluorescent images, a ZEISS Axio Observer D1 fluorescence microscope was used, imaging in the YFP fluorescence channel.

ELISA binding assay

100 uL of fibril samples (AD seeded tau-k19+, αSyn, Aβ(1-42); 600 ng monomer equivalent protein) was added in triplicate to Nunc black 96 well plates and allowed to adsorb overnight at 4^{o}C. At room temperature, sample solution was discarded by quickly inverting the entire plate, then 200 uL of SuperBlock T20 blocking buffer was added for 30 min. Blocking buffer was discarded and wells were washed with 100 uL TBST three times, then solution was discarded and replaced with various concentrations of miniprotein inhibitors in TBST for 2 hours. Plate was washed three time with TBST. 100 uL of Alexafluor 647 labelled anti-His tag antibody (1:500 dilution, in TBST) was added for 1 hour, and plate was washed three times with TBST. Fluorescence was measured using a SpectraMax M5 plate reader (ex: 650 nm; em: 671).

ACKNOWLEDGEMENTS

We thank Mark Arbing of the UCLA-DOE Protein Expression Technology Core facility for assistance in protein purification and Duilio Cascio and Alex Lisker of the UCLA DOE Biocomputing Technology Core. We thank Michael Sawaya for assistance in structural analysis. We thank David Baker for helpful suggestions regarding protein design. KAM is supported by the

UCLA-Caltech Medical Scientist Training Program (GM08042) and the UCLA Chemistry-Biology Interface training grant (USPHS National Research Service Award 5T32GM008496).

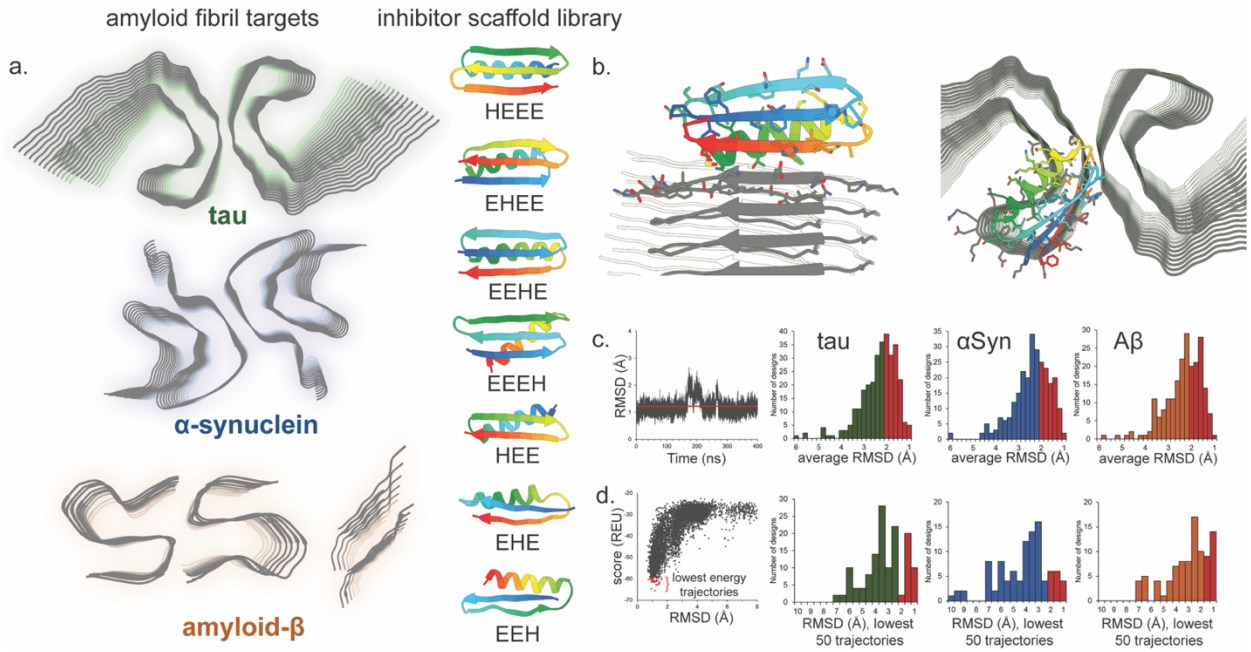


Figure 1.1: Computational design of amyloid inhibiting miniproteins. a. Amyloid fibrils of three different proteins were used as scaffolds for inhibitor design: tau, αSyn, and Aβ. Atomic structures of the tau paired helical filament derived from Alzheimer's disease patient brain (top)[19], the αSyn rod polymorph (middle)[17], and two forms of Aβ (bottom) were targeted[11,31]. A library of *de novo* designed miniproteins was used as inhibitor scaffolds. Seven unique classes of inhibitors were used, each class differing from the arrangement of secondary structural elements (H = alpha-helix; E = beta-sheet). **b**. Inhibitor scaffolds were docked to the ends of the fibril structures, capping their growth by preventing further addition of protein monomers. Binding of the miniprotein scaffolds to the fibrils was primarily driven by interacting beta-sheets, mimicking the fibril native stacking. **c**. The stabilities of top-ranking hits from the docking calculations were assessed by long-range molecular dynamics simulations. Those inhibitors with lowest average RMSD's over time were selected for further testing (red bars). **d**. Final screening of inhibitors was performed using Rosetta's ab initio structure prediction algorithm. The structures of each inhibitor were predicted based on primary sequence alone. The energies of each prediction trajectory were

plotted against their RMSD to the original design. Those inhibitors whose lowest energy predictions were the smallest RMSD from the original design were then selected for experimental characterization (red bars).

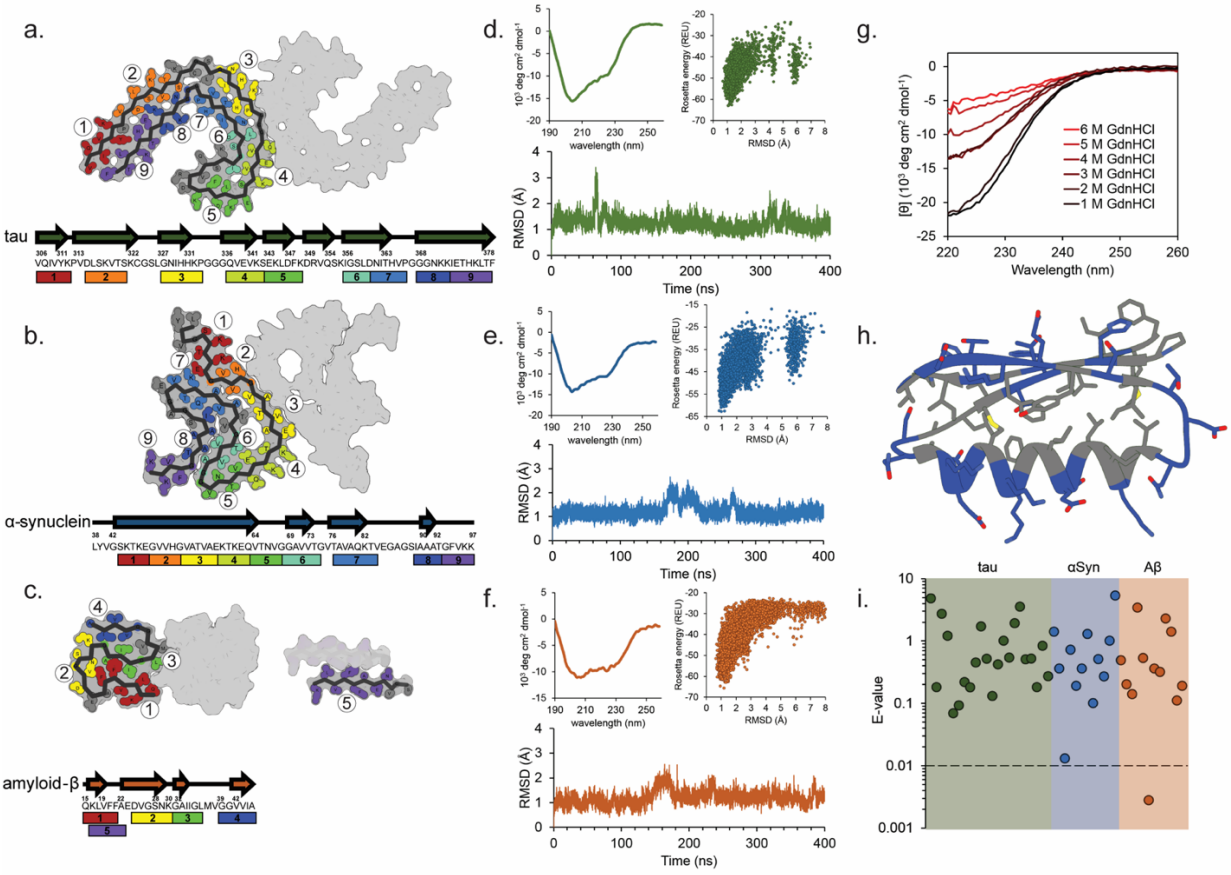


Figure 1.2: Biophysical characterization of designed inhibitors. a-c. Multiple binding sites on each amyloid fibril structure were selected for targeted inhibitor design. The inhibitor scaffolds were systematically docked to different sites along the fibril ends (each number/color corresponding to a unique binding site). The chosen binding sites correspond to particular beta-strand segments (shown in arrows) occurring along the protein chain for tau (**a**) αSyn (**b**) and Aβ (**c**). **d-f**. Initial biophysical characterization of each design consisted of circular dichroism (CD) spectroscopy (top left), ab initio structure prediction (top right) and long-range molecular dynamics simulations (bottom). Inhibitors iTau-N (**d**), i αSyn-F (**e**), and iAβ-H (**f**) are shown to maintain stable folds both computationally and experimentally. **g**. To assess the stability of the designed inhibitors, CD measurements were taken after 20-minute incubation with increasing

concentrations of the denaturant guanidinium hydrochloride (GdnHCl). iTau-N (shown) remains completely stable in 1 M GdnHCl. **h**. The fold of each designed inhibitor is driven by a hydrophobic core region (gray residues) surrounded by an exterior of charged and polar residues (blue) (inhibitor iTau-N shown). **i**. Each inhibitor was generated *de novo*, with no apparent homology to known naturally occurring proteins. BLAST E-values, a metric indicating protein homology, the designs are well above the significance threshold of 0.01, for all inhibitors targeting tau (green), αSyn (blue) and Aβ (orange), with the exception of iAβ-L.

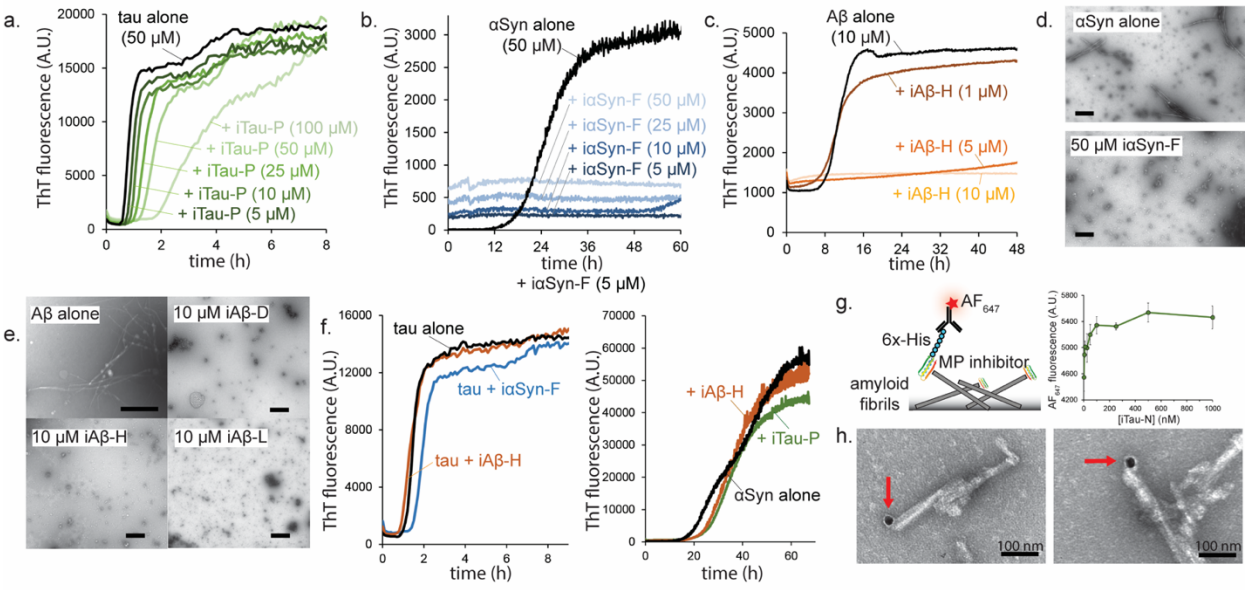


Figure 1.3: Inhibitors prevent amyloid aggregation by capping fibril ends. a-c. Thioflavin T aggregation kinetics assays for tau (**a**), αSyn (**b**), and Aβ (**c**)**.** Each amyloid protein was aggregated alone and in the presence of increasing concentrations of inhibitor, resulting in reduction in the rate of aggregation or complete abolishment of aggregation. **a**. The aggregation of tau k18+ (50 μM monomer) with the inhibitor iTau-P (100 μM) produces a 4-fold increase in aggregation lag time. **b**. Aggregation of αSyn (50 μM monomer) with inhibitor iαSyn-F nearly eliminates measured aggregation, even at sub-stoichiometric ratios. **c**. Similarly, Aβ(1-42) (10 μM monomer) does not aggregate with equimolar ratios of inhibitor iAβ-H. **d**. Transmission electron micrographs (TEM) of αSyn (50 μM) aggregated in the absence (top) or presence (bottom) of iαSyn-F show the inhibitor prevents the formation of fibrillar aggregates. (scale bar indicates 50 nm) **e**. TEM images of Aβ(1-42) alone (10 μM) reveal numerous fibrils, whose growth is prevented by the addition of equimolar amounts of iAβ-D, iAβ-H, and iAβ-L. (scale bar indicates 100 nm) **f**. Both iαSyn-F and iAβ-H show little effect on tau k18+ aggregation (50 μM k18+ monomer, 50 μM inhibitors). Likewise, iTau-N and iAβ-H have minimal influence on αSyn aggregation (50 μM αSyn monomer, 50 μM inhibitors). **g**. ELISA binding assays were used to measure inhibitor

binding to fibril ends. Inhibitors with uncleaved His-tags were incubated with fibril coated plates, then labeled with an AF_{647} conjugated anti-His primary antibody. Fluorescence measurements at different concentrations of inhibitor (iTau-N shown) illustrate binding saturation in the high nanomolar regime. **h**. To visualize the designed miniproteins bound to the fibril ends, inhibitors were incubated with fibrils then labelled with 20 nm gold nanoparticles (indicated by red arrows), demonstrating a binding mode consistent with their intended design.

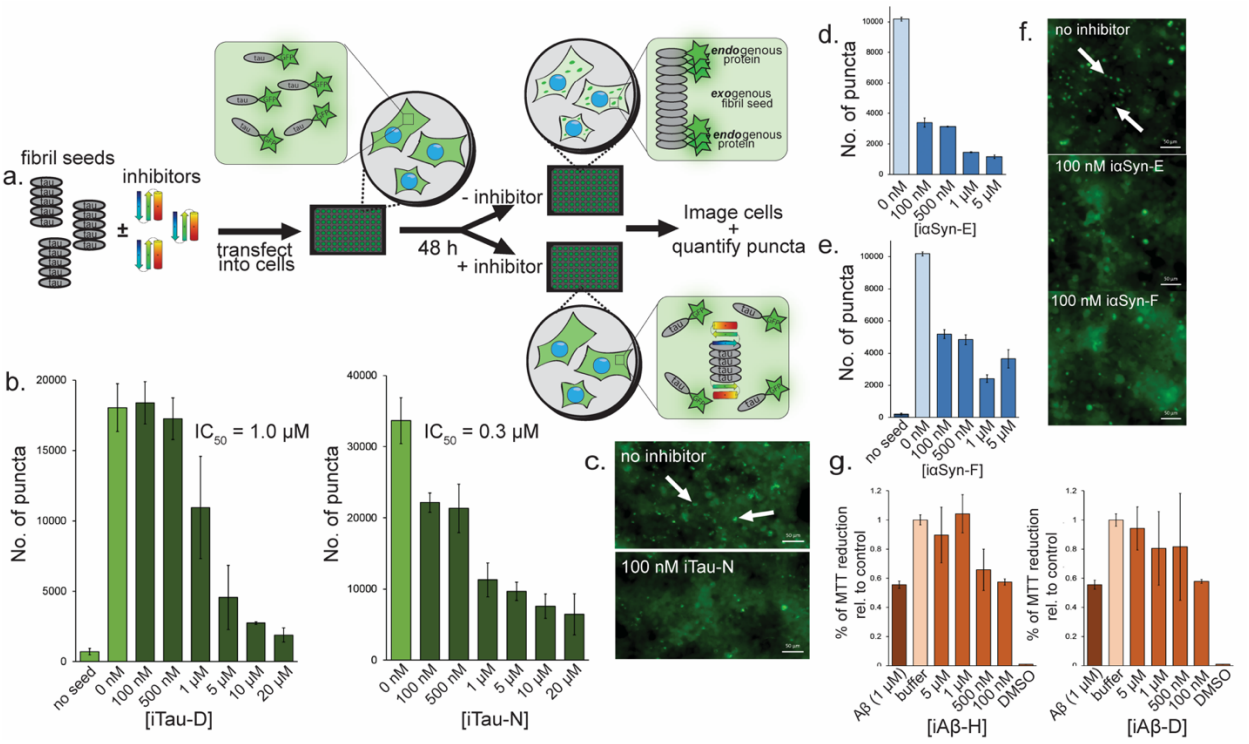


Figure 1.4: Inhibition of amyloid seeding and toxicity in cells. a. To assess the effects of the designed inhibitors on amyloid seeding within cells, a HEK293T biosensor cell assay was used for both tau and αSyn. Fibril seeds (tau shown) are incubated overnight with inhibitors, then transfected into the biosensor cells. The cells overexpress amyloid protein (either tau or αSyn) fused to green fluorescent protein (GFP). At baseline the cells show diffuse fluorescence, but the endogenous fluorescent amyloid protein can be incorporated into the transfected exogenous fibril seeds, resulting in visible fluorescent puncta. Addition of inhibitor caps the fibrils, preventing the incorporation of the GFP-labeled amyloid protein and the subsequent formation of puncta. **b**. Inhibitors iTau-D and i-Tau-N cause a significant reduction in the number of fluorescent puncta in cells transfected with AD patient brain extract containing tau fibrils. Minimal aggregation occurs in the absence of transfected fibril seeds. **c**. Fluorescent microscopy images of biosensor cells with and without 100 nM iTau-N. Seeded aggregates can be visualized by discrete bright puncta (indicated by white arrows). **d-f**. Inhibitors iαSyn-E (**d**) and iαSyn-F (**e**) greatly reduce

aggregated puncta in biosensor cells expressing fluorescently labelled αSyn. **f**. Similar to tau aggregates, αSyn aggregates can be visualized as intracellular fluorescent puncta and quantified (white arrows). **g**. MTT dye reduction assays were used to assess the capacity of inhibitors to mitigate Aβ aggregate-induced cytotoxicity in N2a neuronal cells. Aβ aggregates alone (1 μM) resulted in ~40% cell death compared to buffer control. This toxicity can be rescued by addition of either iAβ-H or iAβ-D. Scale bars indicate 50 μm. All error bars represent standard deviation.

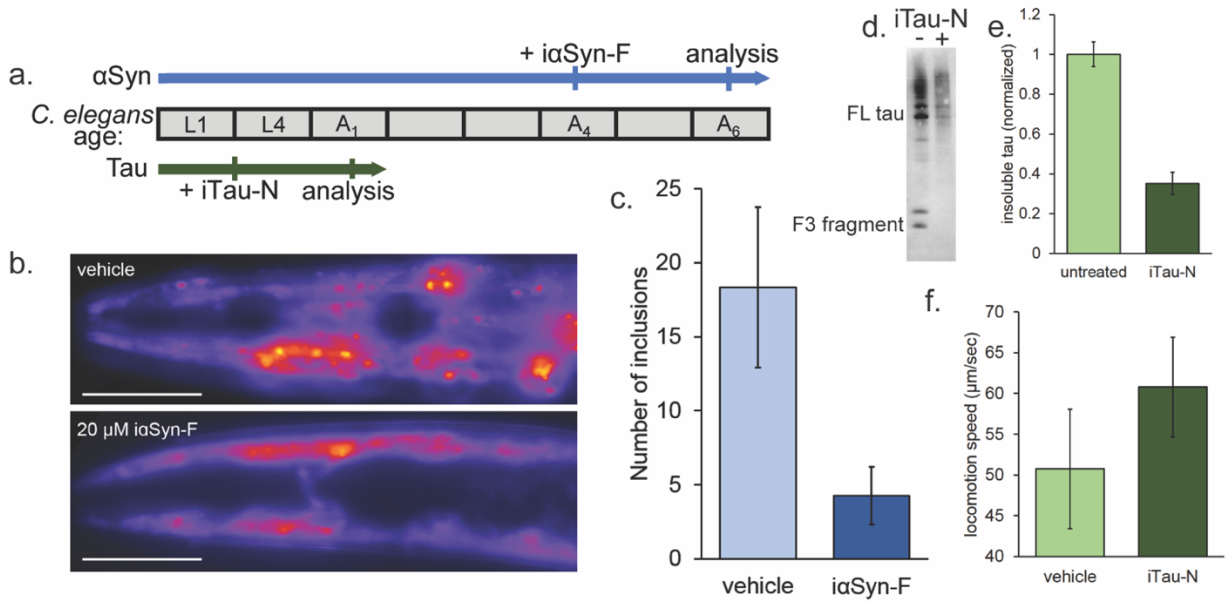


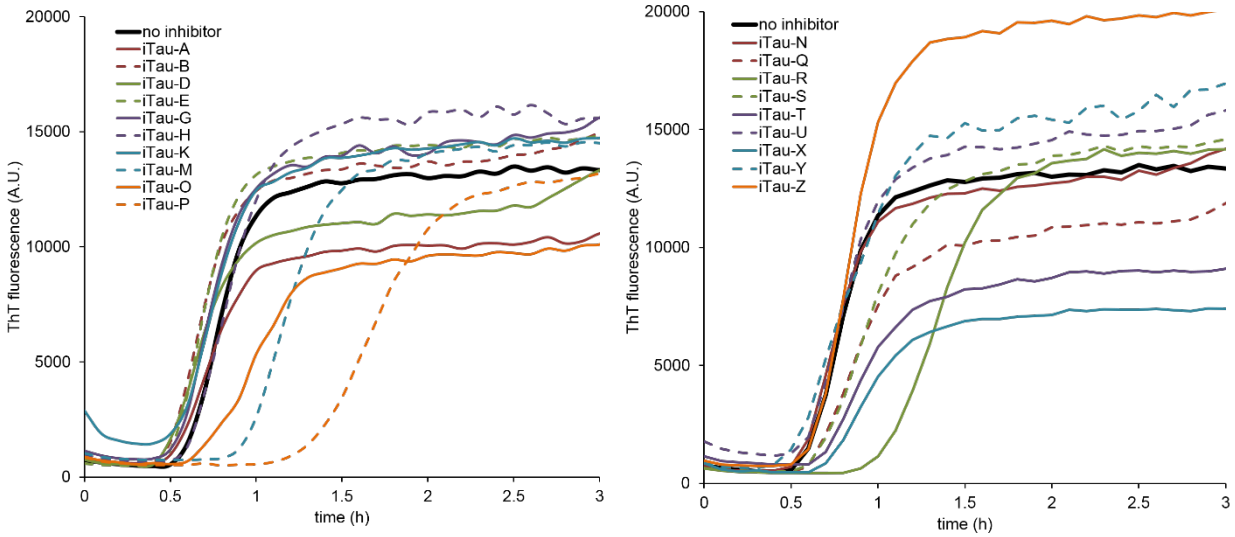
Figure 1.5: Inhibitors prevent amyloid aggregation in *C. elegans* model strains. a. Two strains of *C. elegans* were used to test the *in vivo* efficacy of the designed miniprotein inhibitors (αSyn: DDP1, tau: BR5706). A timeline of worm lifespan through larval stages (L) and adult days (A) is shown, indicating age of worms during inhibitor treatment and subsequent analysis. **b-c**. Alpha-synuclein strain DDP1 overexpresses fluorescently labelled αSyn, which aggregates in the worm head region as adults. **b**. Vehicle treated worms display numerous fluorescent puncta as day 6 adults, which are diminished with the addition of iαSyn-F. (Scale bars indicate 100 μm). **c**. Quantification of head region αSyn aggregates reveals a large reduction in inhibitor treated worms. **d-f**. Tau strain BR5706 co-expresses full-length tau with the pro-aggregation V337M mutation and the F3 fragment of the tau microtubule binding domain with the pro-aggregation K280 deletion. d. Tau Western blot of RIPA-soluble worm extracts with and without inhibitor iTau-N treatment. Bands of full-length (FL) and the F3 fragment are indicated. **e**. Western blot quantification shows a significant reduction of insoluble tau species in the iTau-N treated *C. elegans* compared to vehicle control. **f**. The movement of treated tau worms was

tracked, and average locomotion speed across 30 second intervals was measured. Treatment iTau-N leads to an increase in speed, indicating a recovery of locomotion deficits.

inhibitor	sequence	inhibitor class
iTau-A	PTRSNEYNANGDALKAAELVEKAAKELSTNADTFFFVVY	EHE
iTau-B	DKAQETADRIQTELERKNARNVTITDTVDNTFYISVHLGEFIKISVAAK	HEEE
iTau-C	PRVEQEMRGIDPQQAMTIALKLAKKLGGAMLVEVHGDTVRWTVQV	EHEE
iTau-D	DPTEDDLARKALQWAKDAPPGSTWSDTYDNRYKVTVHVGEFILVRIETE	HEEE
iTau-E	PTQHYDYRGTDIAEVARQAAELADKYNGIVTVVTRGDTIHVTVHI	EHEE
iTau-G	PRTYQTTSDNRVTVTVHIGEFILIHIDVKVPDPTATDKAAKLAEELAKT	EEEH
iTau-H	PTTYHYNLDNKVTIQVHISEFLRIRIDFDPRDMELLIRASLIAADMAAK	EEEH
iTau-K	DKAQETADRIQTELERKNARNVTITDTVDNTFYISVHLGEFIKISVAAK	HEEE
iTau-M	DPRTEEASKIATQLAKEAGPNSDVKVTYERGITIKITNGEFILIIIDIK	HEEE
iTau-N	PVHVTFHIPGVDTKQYTITDGNVAKILTELFKKMQEKANGSALVTHVIMF	EEHE
iTau-O	PTEKYKRRGDNASELHADLARKAAERGASVKLVVFGDDVEVEVHP	EHEE
iTau-P	PRRQTYNVDNNVQVHVKDNGAVHVIIITKRPDTEAIKKAAKIAETAMKT	EEEH
iTau-Q	PKTTYDRRGVDSNTATYEAAKIAAENNGTAAVYVRGDRVRIDIQ	EHEE
iTau-R	PTSTYHFKGYDVAQAAKWATDVATKRNGTTVVVLDGDTVTVHIKD	EHEE
iTau-S	PRVTIASVGGDEKKLKELAEFVAKKLNGTVETRRHGETFTFHVQG	EHEE
iTau-T	PTYHDHQRGLDAKDAVDIARKLLETYNGSVAIHIFGDTLTVRVKV	EHEE
iTau-U	APEMQKAQTTATKLAKKAQPGQTLQSNVDNKVNVTVHVGESLVIVIDAK	HEEE
iTau-V	PRTVRIERPGRFKVEVKLGESIVVYVYVESGNKEDAEEAAKLALTAIKT	EEEH
iTau-W	PRITMHFTNVDAEQIAWEAARAAAENNTAVSIFVDGNHVEVRVQP	EHEE
iTau-X	PVAKMVFYGLSEEQARKIAQKAANLSNGTVSIDSNGDTIDVTVHV	EHEE
iTau-Y	YVQVHIHFANGRTKTIKFEDGDPDEVAKKATEQAKKEANGIPVAVLVIVY	EEHE
iTau-Z	AVIIAIYIGDDETKVKKIAEDAAKALNGEYRVRTDGDTIQLTVTT	EHEE
iαSyn-A	PRKTYTSRGTQAQELATEVTKEAAKLNASIIVLVDGDTVTVHLQV	EHEE
iαSyn-B	GQHHQTYQNTDTRQVIFEAALEAAKKNADVVITQDGDTIHVDIQV	EHEE
iαSyn-C	PTERSEYKGVDATDAIRDALRIALEQNGVVVVSVHGDTVHVTFHN	EHEE
iαSyn-D	PSQRYTYKGISLDQVITIATEIAQKMGGALAIVVYGDTLSFEVHV	EHEE
iαSyn-E	PRVDIEFKNVDPFEARDVALKVAQKLGGAMAVIIDGDTIHIHIQV	EHEE
iαSyn-F	PVYHYRYKGRAAAEAAKEAAKIAQKLGGALVVRVDGDTIRITIAV	EHEE
iαSyn-G	PRSAAQYRGVDEDEVKKLAQTTAEVLNIDYKIHSDGDTITVHFQK	EHEE

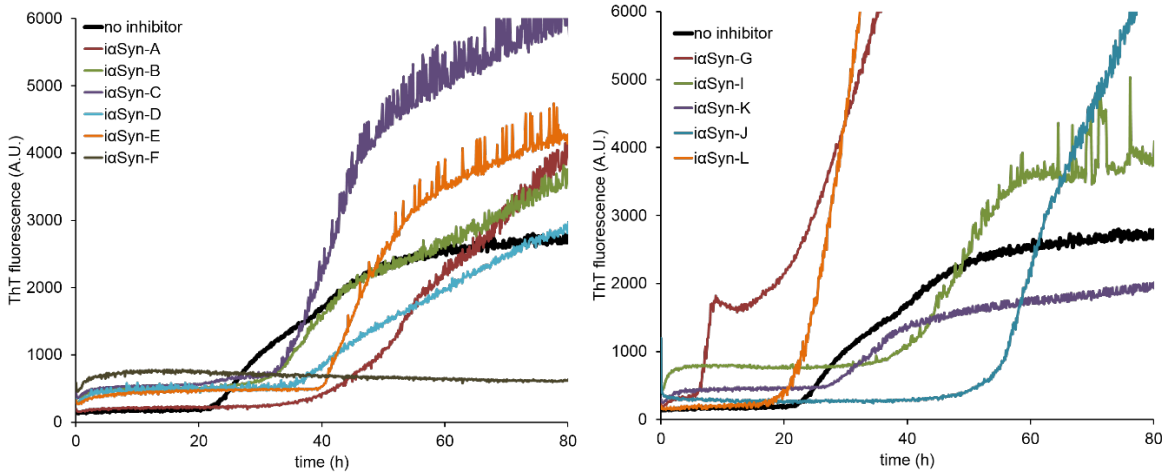
iαSyn-H	DVAVTVVAGSDPREAEDWAKKVADKLGGEIRSRQDGDIVVVEVHA	EHEE
iαSyn-I	DVYVIVAKGVDEREVTKTAKKWADEANAEIRVTSDGDLVEVRVQA	EHEE
iαSyn-J	DVLVYVYNGQDPEEVEKKAKEIADKSNGKVEVDKQGDQVHVTIKV	EHEE
iαSyn-K	DVIVLVFAGYSEEDAKKIAQDTANKKGASYQTRRDGDIIVFHIHV	EHEE
iαSyn-L	DVFVYVVAGDDSTEVATKLKDEAKKQNAKFDLKRDGDIVHITIHR	EHEE
iAβ-A	PSAIYIAQGLDEDQARKIAETMSKTAGGHVEMRTDGDVIEIKLQV	EHEE
iAβ-B	PSTQETYKNTDWREVAETAEKKAREMNGTVYVEARGSTIHVTIET	EHEE
iAβ-C	PTSVVQSRGDDYDELKKKYEDWARKNNADIKVSVDGDTVRITIHM	EHEE
iAβ-D	PLEVTITIGKKIRAHAKSHQDPQLARDLKTVAEKAAKTQNLPLEVHVNNI	EEHE
iAβ-E	PVLVYKARGVDEKEARTAAKEASKALNAEVKIESDGDTFRFQVHQ	EHEE
iAβ-F	MPEAKKAIEKATEIAKKNNARIVYVEVIPGGEIHIDVKP	HEE
iAβ-G	PDITLRYRGTNAAEVIATAEDLSDKLGAIAETYSDGDTITLHLKR	EHEE
iAβ-H	PKRVTYTLNNRVHVQITHTDQKIVYVESSTGDKDAAMTAVKIADELAKK	EEEH
iAβ-I	PYTVENHKGDQSSTKITDSEDPKKAWELAMKILTEALKKNLPIFVILDQP	EEHE
iAβ-J	PTVNIAMAGSDEKTATDIAKKVADELGGELRVTTDGNAVHVHVHV	EHEE
iAβ-K	GTTFNLTTRNRIKYKFHVETDSAEDLTRVADEADKAAHNNAPFETVMVAY	EEHE
iAβ-L	PTIHETYKGYDAREVAKTAATRATKLGAEVYVELDGDTVHVRIQL	EHEE

Supplementary Table 1.1: Sequences and class type of designed inhibitors.



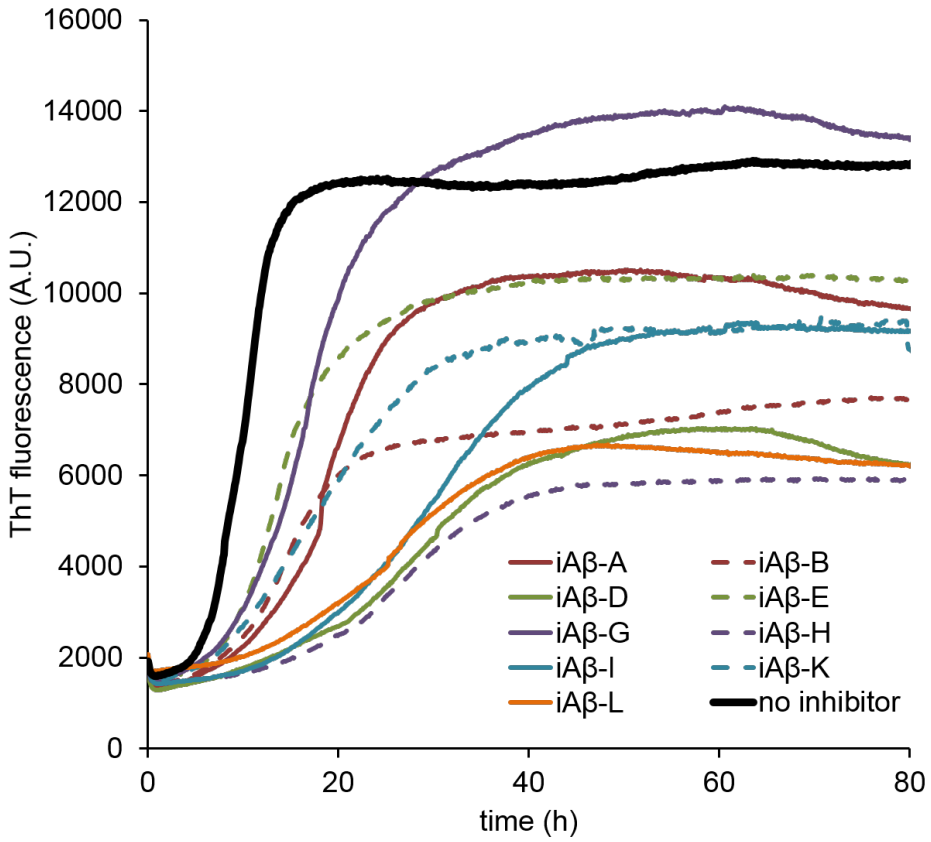
Supplementary Figure 1.1: Effects of designed tau inhibitors on primary aggregation.

Thioflavin T kinetics assays of tau k18+ aggregation (50 μM) in the presence of 50 μM of each designed inhibitor.

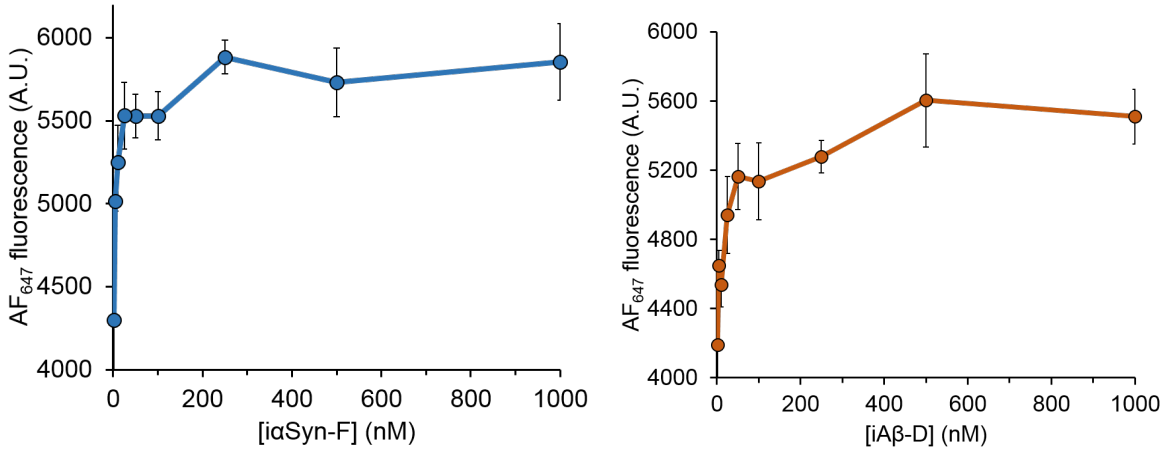


Supplementary Figure 1.2: Effects of designed αSyn inhibitors on primary aggregation.

Thioflavin T kinetics assays of αSyn aggregation (50 μM) in the presence of 50 μM of each designed inhibitor.

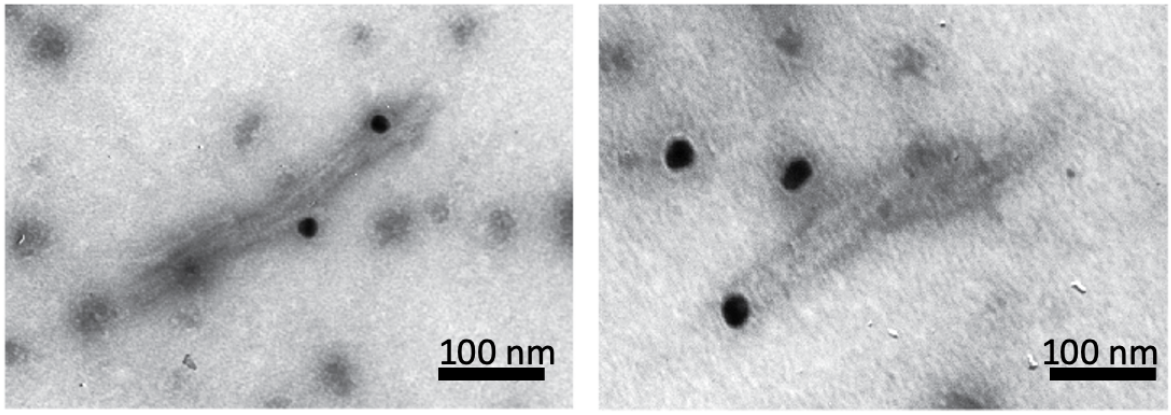


Supplementary Figure 1.3: Effects of designed amyloid-beta inhibitors on primary aggregation. Thioflavin T kinetics assays of amyloid-beta aggregation (10 μM) in the presence of 5 μM of each designed inhibitor.

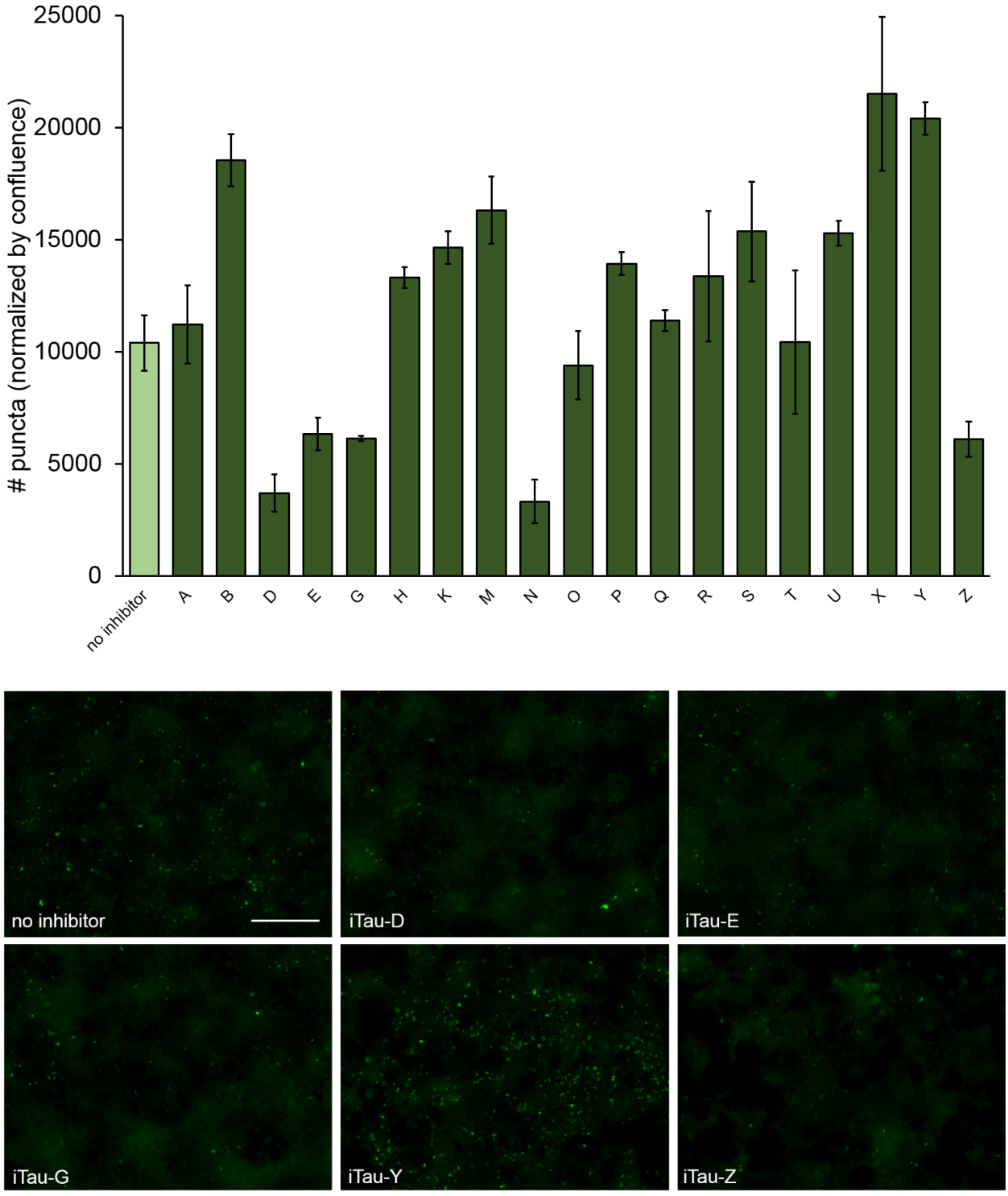


Supplementary Figure 1.4: ELISA binding affinity assays of iαSyn-F and iAβ-D to fibrils.

The binding of the inhibitor iαSyn-F to αSyn fibrils as well as the binding of inhibitor iAβ-D to Aβ(1-42) fibrils saturates in the high nanomolar regime. Both inhibitors have apparent K_D values of <100 nM.

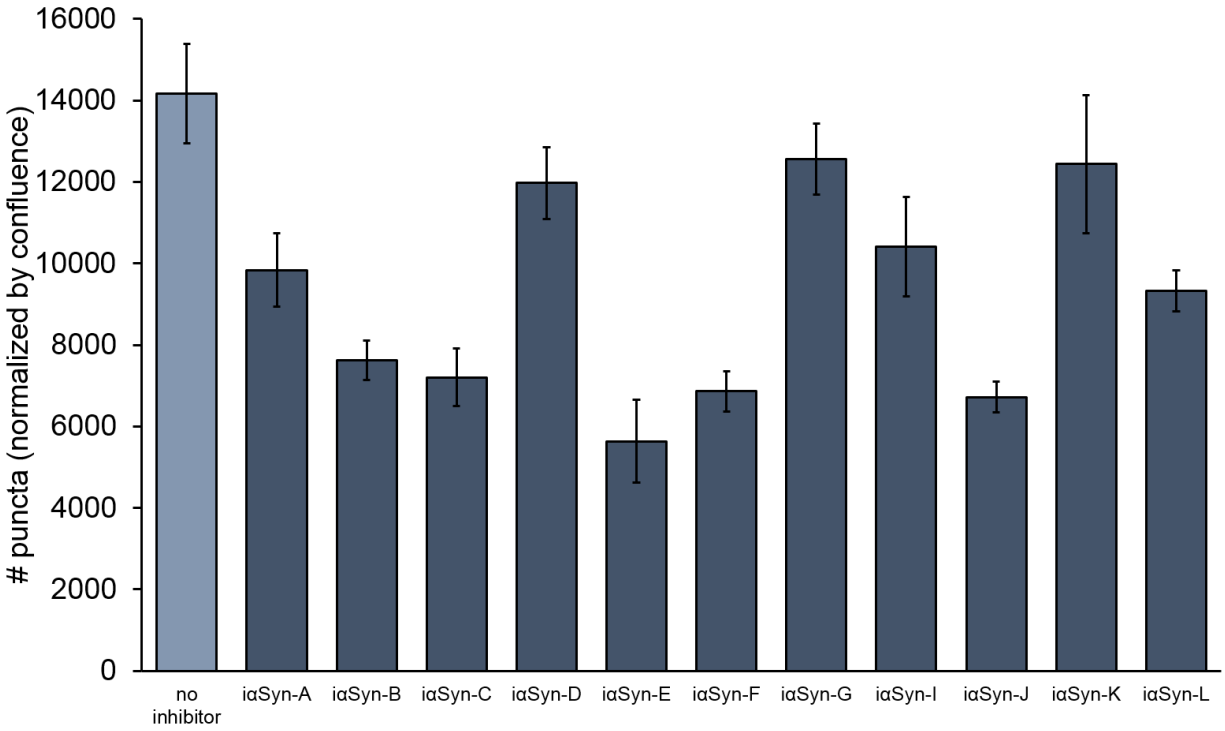


Supplementary Figure 1.5: Additional images of nanogold binding experiments. Inhibitors labelled with gold nanoparticles bind to the growing end of AD tissue seeded tau fibrils.

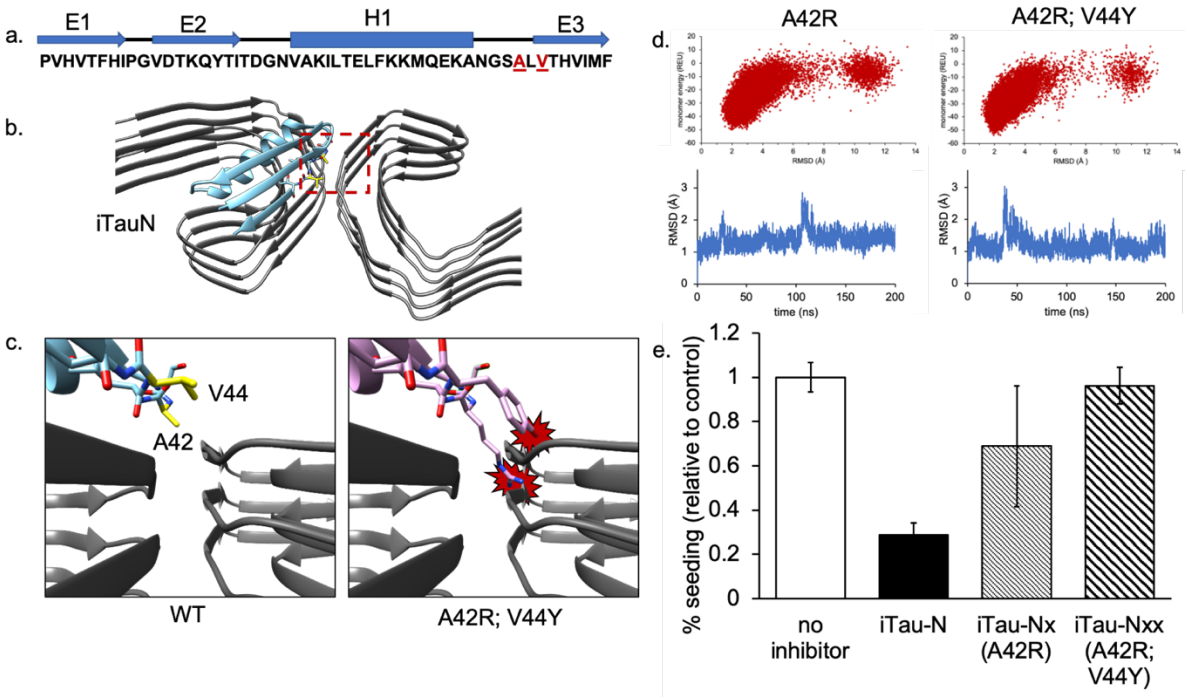


Supplementary Figure 1.6: Inhibition of cellular seeding by designed tau inhibitors. AD patient brain extract was incubated with tau inhibitors (final concentration 10 μM) and added to

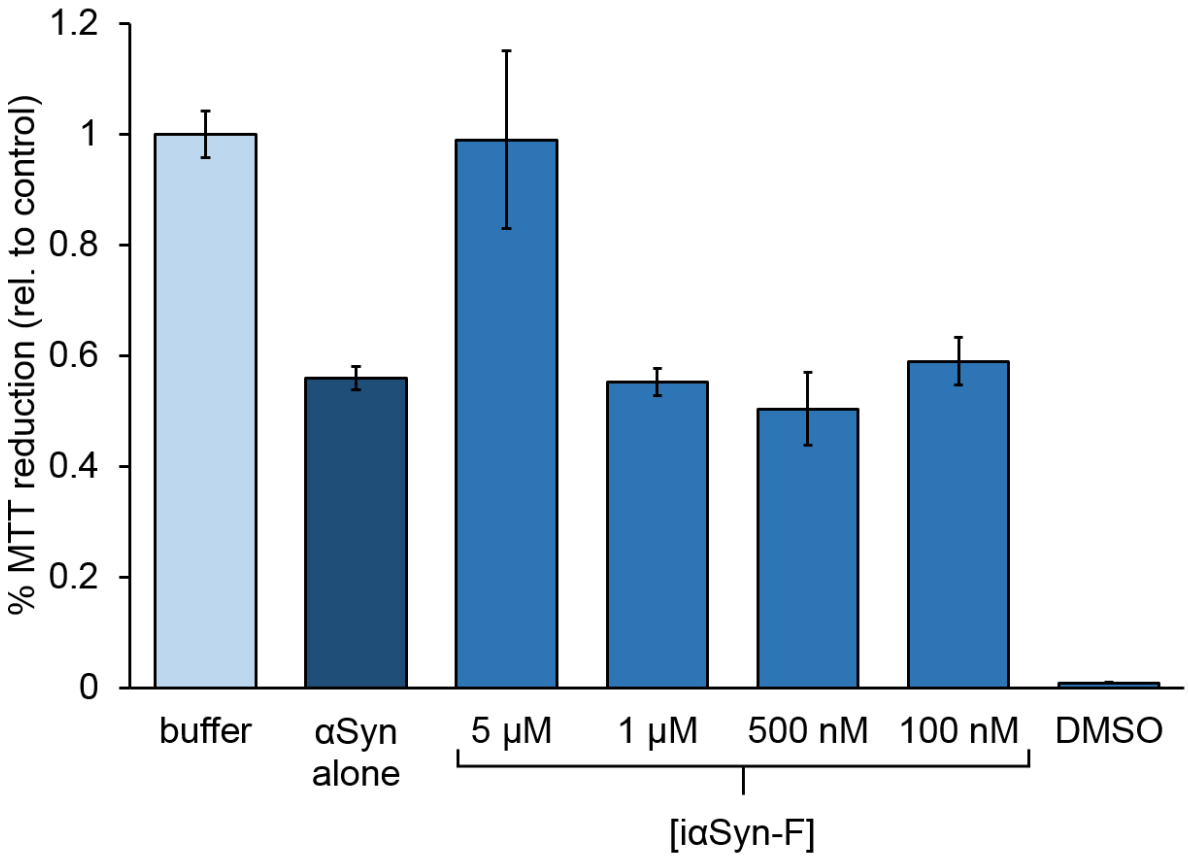
HEK293T tau biosensor cells. Effects of the inhibitors ranged from significant inhibition (iTau-N, iTau-D) to enhancement of seeding (iTau-Y), as shown by quantification of fluorescent aggregates (top) and example fluorescent images (bottom). Scale bar represents 10 μm.



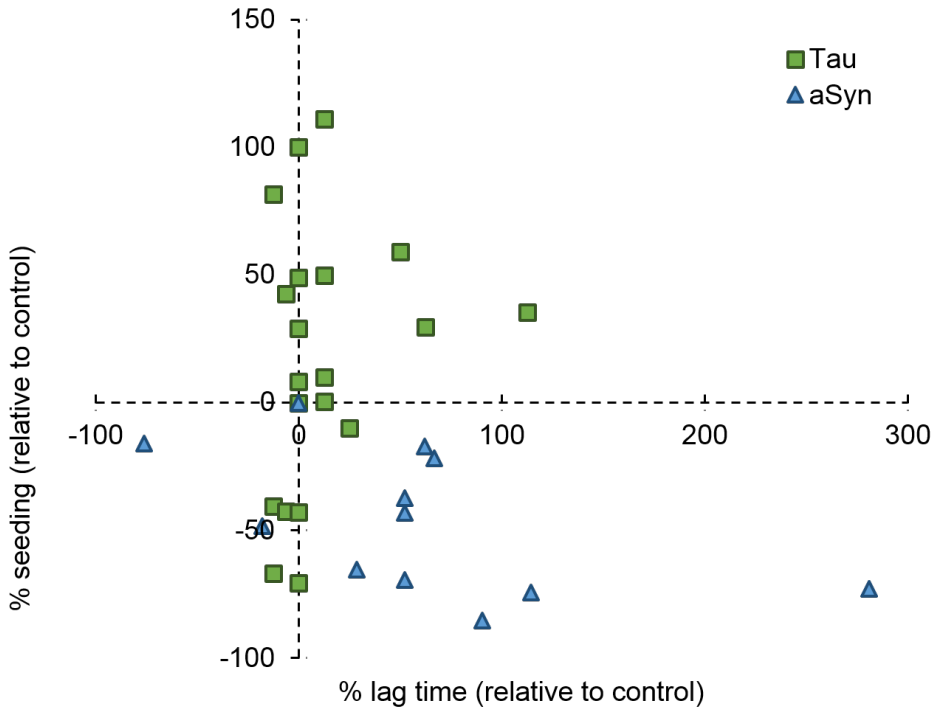
Supplementary Figure 1.7: Inhibition of cellular seeding by designed αSyn inhibitors. αSyn fibrils were incubated with designed αSyn inhibitors (final concentration 10 μM) and added to HEK293T αSyn biosensor cells. Multiple inhibitors displayed significant inhibition.



Supplementary Figure 1.8: Steric clashes introduced into the iTau-N/tau fibril binding site reduce seeding inhibition. a. Sequence of iTau-N, highlighting two key residues found in the binding interface with the tau PHF fibril, A42 and V44. **b.** iTau-N bound to tau PHF fibril. **c.** Mutations A42R and V44Y both introduce bulky steric clashes into the binding interface. d. Both the single and double mutant forms of iTau-N maintain stable folds, as computed by ab initio structure prediction calculations (top) and long-range MD simulation (bottom). **e.** Comparison of biosensor cell seeding between iTau-N and the two control mutants. Addition of the single A42R mutation greatly reduces the ability of iTau-N to inhibit tau seeding and is further reduced by the addition of V44Y.



Supplementary Figure 1.9: iαSyn-F rescues αSyn fibril toxicity. N2a neuronal cells were treated with αSyn fibrils and increasing concentrations of iαSyn-F. 5 μM iαSyn-F completely prevents cellular toxicity.



Supplementary Figure 1.10: Effects of designed tau and αSyn on primary aggregation and cellular seeding. Tau inhibitors have significant effects on either primary aggregation *in vitro*, as measured by changes in aggregation lag time (x-axis), or seeded aggregation in biosensor cells (y-axis). αSyn inhibitors appear to affect both processes while tau inhibitors one affect one or the other.

inhibitor	Amyloid protein target	Amyloid Binding site (Sequence)	Amyloid binding site (ZipperDB score)	Inhibitor binding sequence
iTau-A	tau	QVEVKS	-23.3	FFVVYA
iTau-B		GNIHHKP	-23.1	IKISVAA
iTau-C		VQIVYK	-25.9	AMLVEVH
iTau-D		VQIVYK	-25.9	ILVRIET
iTau-E		GNIHHKP	-23.1	IVTVVTR
iTau-G		VQIVYK	-25.9	ILIHID
iTau-H		IGSLD	-20.8	SEFLRI
iTau-K		GNIHHKP	-23.1	KISVAAK
iTau-M		VQIVYK	-25.9	FILIII
iTau-N		QVEVKS	-23.3	VTHVIMF
iTau-O		EKLDFKD	-19.5	SVKLVVS
iTau-P		QVEVKS	-23.3	HVIIITK
iTau-Q		GNIHHKP	-23.1	TAAVYVR
iTau-R		QVEVKS	-23.3	GTTVVVL
iTau-S		GNIHHKP	-23.1	PRVTIAS
iTau-T		EKLDFKD	-19.5	SVAIHIF
iTau-U		QVEVKS	-23.3	LVIVID
iTau-V		VQIVYK	-25.9	SIVVYV
iTau-W		LDNITHV	-23.8	AVSIFVD
iTau-X		LDNITHV	-23.8	AKMVFY
iTau-Y		VQIVYK	-25.9	VLVIVY
iTau-Z		VQIVYK	-25.9	VIIAIYI
iαSyn-A	αSyn	KTKEG	-22.6	ASIIVL
iαSyn-B		GVVHG	-24.5	ADVVIT
iαSyn-C		GVVHG	-24.5	GVVVVS
iαSyn-D		GVATVAE	-26.9	ALAIVVY
iαSyn-E		GVATVAE	-26.9	AMAVIID
iαSyn-F		GAVVTG	-26.1	GALVVR

iαSyn-G		AVAQKTV	-25.4	RSAAQYR
iαSyn-H		VVTGVTA	-23.9	VAVTVVA
iαSyn-I		VVTGVTA	-23.9	VYVIVAK
iαSyn-J		VVTGVTA	-23.9	VLVYVYN
iαSyn-K		VVTGVTA	-23.9	VIVLVFA
iαSyn-L		VVTGVTA	-23.9	VYVYVVA
iAβ-A	Aβ	KLVFFAE	-22.2	SAIYIAQ
iAβ-B		KLVFFAE	-22.2	TVYVEAR
iAβ-C		GVVIA	-27.8	TSVVQ
iAβ-D		KLVFFAE	-22.2	IRAHAKS
iAβ-E		KLVFFAE	-22.2	VLVYKAR
iAβ-F		KLVFFAE	-22.2	RIVYVEV
iAβ-G		VGSNK	-19.9	IAETYS
iAβ-H		KLVFFAE	-22.2	KIVYVESS
iAβ-I		GVVIA	-27.8	PIFVIL
iAβ-J		QKLVF	-24.0	TVNIAM
iAβ-K		QKLVF	-24.0	TVMVAY
iAβ-L		KLVFFAE	-22.2	EVYEVLD

Supplementary Table 1.2: Designed inhibitor targets and sequences. The amyloid protein targets for each designed inhibitor, with the sequence and ZipperDB score of each segment listed. The sequence of the inhibitor at the amyloid binding site is also listed, highlighting that the inhibitors lack any sequence similarity to the native amyloid protein.

REFERENCES

1. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
2. Bloom, G. S. Amyloid-β and Tau: The Trigger and Bullet in Alzheimer Disease Pathogenesis. *JAMA Neurol.* **71**, 505–508 (2014).
3. Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6469–6473 (1998).
4. Wang, Q., Yu, X., Li, L. & Zheng, J. Inhibition of amyloid-β aggregation in Alzheimer's disease. *Curr. Pharm. Des.* **20**, 1223–1243 (2014).
5. Bulic, B. *et al.* Development of tau aggregation inhibitors for Alzheimer's disease. *Angew. Chem. Int. Ed Engl.* **48**, 1740–1752 (2009).
6. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
7. Sievers, S. A. *et al.* Structure-Based Design of Non-Natural Amino Acid Inhibitors of Amyloid Fibrillation. *Nature* **475**, 96–100 (2011).
8. Seidler, P. *et al.* Structure-based inhibitors of tau aggregation. *Nat. Chem.* **10**, 170–176 (2018).
9. 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).
10. Sangwan, S. *et al.* Inhibition of synucleinopathic seeding by rationally designed inhibitors. *eLife* **9**,.

11. Griner, S. L. *et al.* Structure-based inhibitors of amyloid beta core suggest a common interface with tau. *eLife* **8**, e46924 (2019).

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

13. Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).

14. Fitzpatrick, A. W. & Saibil, H. R. Cryo-EM of amyloid fibrils and cellular aggregates. *Curr. Opin. Struct. Biol.* **58**, 34–42 (2019).

15. Rodriguez, J. A. *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).

16. 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).

17. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9**, 1–10 (2018).

18. 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).

19. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).

20. Zhang, W. *et al.* Novel tau filament fold in corticobasal degeneration. *Nature* (2020) doi:10.1038/s41586-020-2043-0.

21. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).

22. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

23. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016).

24. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).

25. Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).

26. Falcon, B. *et al.* Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol. (Berl.)* **136**, 699–708 (2018).

27. Tuttle, M. D. *et al.* Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-Synuclein. *Nat. Struct. Mol. Biol.* **23**, 409–415 (2016).

28. Guerrero-Ferreira, R. *et al.* Cryo-EM structure of alpha-synuclein fibrils. *eLife* **7**, e36402 (2018).

29. Li, Y. *et al.* Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Res.* **28**, 897–903 (2018).

30. Schweighauser, M. *et al.* Structures of α-synuclein filaments from multiple system atrophy. *bioRxiv* 2020.02.05.935619 (2020) doi:10.1101/2020.02.05.935619.

31. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4976–E4984 (2016).

32. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **6**, (2011).

33. Lin, Y.-R. *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5478–E5485 (2015).

34. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).

35. Silva, D.-A., Correia, B. E. & Procko, E. Motif-Driven Design of Protein-Protein Interfaces. *Methods Mol. Biol. Clifton NJ* **1414**, 285–304 (2016).

36. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

37. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871 (2005).

38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

39. Mucke, L. & Selkoe, D. J. Neurotoxicity of Amyloid β-Protein: Synaptic and Network Dysfunction. *Cold Spring Harb. Perspect. Med.* **2**, (2012).

40. Lashuel, H. A., Overk, C. R., Oueslati, A. & Masliah, E. The many faces of α-synuclein: from structure and toxicity to therapeutic target. *Nat. Rev. Neurosci.* **14**, 38–48 (2013).

41. Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **10**, 787–797 (2003).

CHAPTER 2

Structure-based inhibition of FUS liquid-liquid phase separation using *de novo* protein design

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are both neurodegenerative diseases which have poor prognoses and currently lack therapies capable of slowing of delaying disease progression beyond several months[1–3]. For the two drugs capable of delaying progression of ALS, their mechanism of action is largely unknown, and no FDA approved treatments for FTD exist. This presents a need for therapeutic strategies that directly target the pathological underpinnings of these diseases. ALS and FTD, like many other neurodegenerative diseases, are hallmarked by aberrant protein aggregation within neurons[4]. These aggregates are primarily composed of proteins which contain low-complexity domains (LCD) thought to be the primary drivers of aggregation. These proteins include FUS (fused in sarcoma), TDP43 (TAR DNA-binding protein 43), and hnRNPA1/hnRNPA2 (heterogeneous nuclear ribonucleoproteins)[5]. Pathogenic mutations can drive the mislocalization of these proteins from the nucleus into the cytoplasm, where they participate in the formation stress granules, membraneless bodies involved in cellular stress response[6,7]. Excessive stress granule formation is observed in both ALS and FTD[8,9]. FUS, TDP43, and hnRNPA1/hnRNPA2 are known to undergo liquid-liquid phase separation (LLPS) and stress granules are thought to form, at least in part, from this process[10,11]. There is also evidence suggesting that solid protein aggregates can arise from liquid stress granules in a disease state[6]. This highlights LLPS of these LCD-containing proteins, and their contribution to stress granule formation, as a possible target for mitigating the pathologic effects of protein aggregation in ALS or FTD.

Significant efforts have been made to target protein aggregation in neurodegenerative disease, but they have primarily focused on blocking fibrillar aggregates. Atomic structures of amyloid fibrils have been used to design inhibitors that prevent aggregation of beta-amyloid[12] and tau[13–15], found in the plaques and neurofibrillary tangles of Alzheimer's disease (AD) brains[16], or alpha-synuclein[17], found in Lewy bodies during Parkinson's disease (PD)[18]. However, lack of definitive structural information for proteins undergoing LLPS has hampered the efforts to create rationally designed therapeutics. Pathogenic amyloid structures, like those found in AD and PD, are typically composed of repeating layers of interdigitating pleated beta-sheets, which contribute to their highly stable and irreversible formation. Work by Hughes et al. demonstrates that amyloid-like fibrils with kinked beta-sheets, as opposed to pleated, contribute to the reversible formation of FUS LCD segments into hydrogels[19]. Furthermore, fibrillar structures have also been observed within reversible phase-separated droplets of hnRNPA1, which also contain a kinked beta-sheet conformation[20]. Solved by solid-state NMR (ssNMR), a recent fibril structure from the FUS LCD (residues 37-97) reveals extensively kinked beta-sheets throughout the fibril backbone stabilized by polar interactions and side-chain hydrogen bonding, versus the usual hydrophobic interactions stabilizing canonical amyloid fibrils[21,22]. Phosphorylation and mutation of core-forming residues found in the structure disrupts FUS phase separation *in vitro*, highlighting the potential role of such a structure in reversible LLPS.

In this study, we use the atomic structure of the FUS LCD amyloid-like fibril to design inhibitors capable of preventing FUS self-assembly. The designed inhibitors bind to the growing end of the fibril, capping its elongation by preventing the incorporation of additional protein monomers. We assess if this design strategy yields inhibitors that block both FUS LLPS and fibrilization *in vitro*, as well as intracellular stress granule formation. Lastly, we determine if

targeting the inhibitors specifically to the FUS structure prevents them from interfering with the phase separation of other LCD-containing proteins TDP43 and hnRNPA2. To generate the FUS inhibitors, we use *de novo* computational protein design to create a library of 35-50 residue miniproteins, from which we identify nine potential inhibitors. Computational protein design is a rapidly growing field and the number of biological applications from proteins designed *de novo* using this technology is continuously expanding[23]. Here, we have selected miniproteins as our chosen design scaffold due to their size and high degree of stability[24,25]. We identify miniprotein inhibitors capable of mitigating both solid and liquid aggregation of FUS, providing a potential route for treatment of diseases with FUS pathology like ALS and FTD. This approach has potential for translation to other neurodegenerative diseases involving aberrant phase separation.

RESULTS

Computational design of miniprotein inhibitors

The selected target of our inhibitor design pipeline is the fibril structure of FUS, determined using solid-state nuclear magnetic resonance (ssNMR) by Murray and colleagues[21]. The fibril core of this structure consists of residues 37-97 from the N-terminal low-complexity domain (LCD; residues 1-214) of FUS, which is rich in asparagine, serine, tyrosine, and glycine amino acids (Figure 2.1). As opposed to ssNMR or cryo-electron microscopy (cryoEM) structures of other amyloid proteins, the FUS LCD structure lacks stabilizing hydrophobic interactions within its core region and is instead stabilized by polar side-chain and backbone interactions[26–28]. Phosphorylation of core residues found within the structure prevents the formation of liquid-like droplets, the binding of soluble FUS, and hydrogel formation, indicating the role this ordered region plays in liquid-liquid phase separation (LLPS). Like other amyloid structures, the FUS LCD fibril is

composed protein monomers stacked into beta-sheets, yielding a cross-beta architecture. Interdigitation of residue sidechains between the stacked sheets and backbone hydrogen-bonding between layers stabilize the structure.

In order to prevent the growth of the FUS fibril architecture, and thus block its potential role in LLPS, we have designed inhibitors which bind to the exposed end of the fibril, capping its elongation by preventing the incorporation of additional FUS LCD monomers onto the ends. The inhibitors mimic the native interactions found in the fibril, as their primary binding mode is through beta-sheet interactions between the inhibitor and the top layer of the fibril. Residues 84-90 (QSSQSSY) were selected as a site for inhibitor docking onto the fibril based on several criteria (Figure 2.1b). First, several regions of the structure were determined to be dynamically disordered (res. 37-43, 53-64), making them difficult targets, while FUS_{84-90} remain stable. Second, while the structure is largely composed of beta-sheets, many of them have kinked backbones, which is a difficult conformation to mimic with small inhibitor proteins. In contrast, residues FUS_{84-90} exist as a stable pleated sheet ideal for docking. Third, computational prediction of which sequence elements in FUS are most amyloid-prone highlight FUS_{84-90} as a highly amyloidogenic segment which may be important to overall assembly of the fibril (Supplementary Figure 2.1). Lastly, Ser84 and Ser87 we determined to have important effects on fibril stability, hydrogel binding, and liquid-droplet formation through targeted mutagenesis and phosphorylation experiments[21].

The inhibitors targeting the FUS LCD fibril are 35-50 amino acid miniproteins, computationally designed using Rosetta. Inhibitor designs can be grouped into seven classes, based on the arrangement of secondary structural elements within the protein fold (Figure 2.1a). They consist of two or three beta-sheets (E) with a single stabilizing alpha-helix (H), producing the classes EEH, EHE, HEE, EEEH, EEHE, EHEE, HEEE. Inhibitor backbones were designed *de*

novo using Rosetta's fragment-based assembly approach, and sequences for the designed backbone topologies were created and optimized to accommodate the backbone using Rosetta FastDesign[29,30]. Previous work involving the design of similar miniproteins scaffolds incorporated disulfide bridges to aid in design stability[24,25]. However, since the intended target of our designs are FUS aggregates found in the reducing environment of the cytoplasm, disulfide bonds were omitted. For each inhibitor class, 5,000 unique inhibitor topologies were created, all with distinct primary sequences. The inhibitors were sequentially docked onto the end of the FUS LCD fibril using Rosetta MotifGraft[31]. In this application, residues 84-90 of the top layer of the fibril were selected as the binding motif, and a backbone alignment (RMSD < 1 Å) of beta-strands from the inhibitors to this motif was used to guide docking (Figure 2.1c). Once docked onto the fibril, inhibitor residues near the binding interface with the fibril were optimized to improve binding energy and shape complementarity. This includes residues of the inhibitor beta-strand, as well as residues of the stabilizing helix making secondary contacts with the fibril surface.

After docking, top inhibitors were selected based on several metrics related to both the stability of the designed inhibitor (p_aa_p, total score) and the inhibitor/fibril interaction (number of atoms in interface, number of unsatisfied hydrogen-bonds, ddg), based upon previous miniprotein design efforts[25]. As a secondary validation of the inhibitors, additional computational methods were used to evaluate inhibitor stability and overall folds (Figure 2.1d). Long-range molecular dynamics (MD) simulations of 200 ns were used to screen out unstable inhibitors, and those with lowest average backbone RMSDs were subjected to an additional 200 ns of simulation for further confirmation[32]. Stability during extended time scale MD simulations has been shown to be predictive of success in previous miniprotein designs[33]. After the MD calculations, the most stable designs were then assessed using Rosetta's *ab initio* structure prediction algorithm[34]. The

fold of each inhibitor was predicted based solely on its primary sequence. 50,000 prediction trajectories were calculated for each sequence and the energies of each predicted structure were then compared to their RMSD from the original design. The RMSD values for the 50 lowest energy trajectories were compared for each design, and those inhibitors whose primary sequence led to a predicted structure similar to the original design were selected, and top designs chosen for further testing (Supplementary Table 2.1). Expression and purification of top hits resulted in nine soluble and well-folded designs, as measured by circular dichroism (Supplementary Figure 2.2), which were used in subsequent experimental assays. BLAST analysis of the designs reveals they are truly *de novo* and have no significant sequence similarity to any known proteins (Supplementary Figure 2.3).

De novo designed proteins reduce FUS LLPS and fibrilization *in vitro*

We first aimed to assess the impacts of the designed miniprotein inhibitors on FUS LLPS *in vitro*. The FUS LCD (residues 1-214) was expressed and purified, fused with an N-terminal mCherry for solubility. At 37ºC in 1x PBS, mCh-FUS_{1-214} remains soluble and non-phase separated. When cooled to 4ºC, rapid phase separation occurs, producing visible phase separation droplets (Figure 2.2a). Thioflavin T, a dye which fluoresces when bound to cross-beta amyloid structures, was used to assess the presence of fibrillar structure within the phase separated droplets. As shown in Figure 2.2b, bright field and fluorescent microscope images of mCh-FUS_{1-214} at 37ºC shows little phase separation or ThT fluorescence (em: 480 nm). However, upon cooling to 4ºC, extensive droplet formation is observed as well as ThT fluorescence within the droplets. This process occurs reversibly, as returned to cooled solution back to 37ºC eliminates the fluorescence signal (Supplementary Figure 2.4). Importantly, ThT fluorescence is not observed in the 4ºC solution outside of the liquid droplets, and the fluorescence of ThT alone is not affected by

temperature changes (Supplementary Figure 2.5). ThT fluorescence has also been observed in the phase separation of hnRNPA1 liquid droplets[20]. Combined, this presents ThT fluorescence as a method for quantitatively measuring LLPS in solution.

To test the efficacy of the miniprotein inhibitors, mCh-FUS_{1-214} (50 μM in 1x PBS, 50 μM ThT) was incubated with equimolar ratios of each of the nine designs. Samples were maintained at 37ºC for 1 hour prior to measurement to eliminate any residual phase separation. At t=0, samples were cooled to 4ºC, and incremental readings of ThT fluorescence were taken (Figure 2.2c). An immediate increase in measured fluorescence is observed which plateaus after ~20 min. A range of effects were observed for the different designs. Several inhibitors significantly reduced ThT signal (iFUS-C, iFUS-G, iFUS-H), while others led do moderate reduction (iFUS-A, iFUS-B), little effect (iFUS-I, iFUS-J, iFUS-K), or even promotion (iFUS-L). A concentration dependent assessment of inhibitor iFUS-G shows it reduces mCh-FUS_{1-214} LLPS in a dose-dependent manner at substoichiometric ratios. Interestingly, at 100 μM, the effects of iFUS-G are less than at 50 μM concentration.

As a secondary assessment of the *in vitro* effects of the inhibitors on FUS LLPS, top inhibitors iFUS-G and iFUS-H were incubated with mCh-FUS_{1-214}, and similarly cooled from 37ºC to 4ºC to induce LLPS (Supplementary Figure 2.6). The mixtures were then centrifuged at 20,000 x g to separate the light and dense phases. Following this, the concentration of mCh-FUS_{1-214} in the sample supernatant was quantified by mCherry fluorescence. In the sample with no inhibitor added, a sizeable drop in mCh-FUS_{1-214} concentration is observed between pre- and post-centrifugation samples, indicating liquid droplets formed and were pelleted out of the supernatant. In samples containing iFUS-G and iFUS-H, a lesser drop in mCherry fluorescence between pre- and post- centrifugation was measured. This observed effect may be caused by the inhibitors

increasing the saturation concentration required for LLPS to occur, leading to more soluble protein being left in the supernatant.

Lastly, as the inhibitors were designed using the FUS fibril structure, we sought to determine if they in fact prevented FUS fibrilization, not just LLPS (Supplementary Figure 2.7). To test this, the top LLPS inhibitor iFUS-G was incubated quiescently at 37ºC with FUS_{1-214} in 1x PBS for 1 week. Fibril growth was monitored by solution turbidity (abs: 380 nm). Increasing concentrations of iFUS-G led to significant reductions in fibril formation, and transmission electron micrographs of the sample reveal an absence of observable fibrils at high inhibitor concentration. To assess if iFUS-G specifically inhibits the fibrillization of FUS, we also tested it against the aggregation of other known amyloid proteins, tau and alpha-synuclein (aSyn) (Supplementary Figure 8). The amyloidogenic microtubule binding domain of tau (k18) or aSyn were incubated with an equimolar ratio (50 μM) of iFUS-G, then shaken at 37ºC. Aggregation of the proteins was monitored by ThT fluorescence, revealing that iFUS-G has little influence on either proteins' amyloid aggregation. Additionally, we tested if iFUS-G influences the seeded aggregation of aSyn in cells. To measure of impact of iFUS-G on intracellular seeding of aSyn, HEK293T biosensor cells were used[35]. These cells endogenously express aSyn fused to either cyan or yellow fluorescent protein (CFP/YFP), which can be visualized as diffuse green fluorescence throughout the cell. However, upon the transduction of an exogenous seed of aSyn amyloid fibrils into the cells, the endogenous aSyn becomes incorporated into the exogenous fibrils through templated seeding. This produces visible fluorescent puncta throughout the cell which can be imaged and quantified. Incubation of aSyn fibril seeds with iFUS-G (10 μM final concentration in cells) had no impact on the number of quantifiable puncta, indicating that it does not influence seeded aSyn aggregation.

Designed inhibitors prevent FUS incorporation into cellular stress granules

We next assessed the impact of designed inhibitors on FUS phase separation within the cell. Previous studies have implicated FUS in the formation of stress-induced membraneless compartments known as stress granules which arise, at least in part, due to LLPS driven by the FUS low-complexity domain. FUS stress granules can be induced from a variety of conditions, including oxidative, heat, and osmotic stress. To measure the effect the miniproteins have on FUS stress granule formation, HEK293T cells were transiently transfected with mCh-FUS_{1-214} (Figure 2.3a). 24 hours post transfection, cells were transduced with either inhibitor protein or vehicle control. 24 hours after this, cells were osmotically stressed with 0.4 M sorbitol to induce granule formation. 1 hour later, cells were imaged, and granules were quantified. As visualized by fluorescent microscopy, without the addition of sorbitol, no stress granule formation was observed (Figure 2.3b). Cells that were stressed but received no inhibitor displayed extensive, discrete granules formed throughout the cell. The addition of inhibitor led to a reduction in the number of quantifiable granules. Inhibitors iFUS-C, iFUS-G, and iFUS-H all led to a reduction in granule formation. At 10 μM, iFUS-G caused the most pronounced effect, halving the number of granules formed, while inhibitor iFUS-L had little influence on granule formation (Figure 2.3c). The inhibitory effects of iFUS-G are concentration dependent, with a calculated IC_{50} of 12.0 μM (Figure 2.3d).

Given the success of iFUS-G to reduce FUS LLPS both *in vitro* and in cells, we aimed to further improve its design. iFUS-G was rerun through the computational design pipeline, again being docked to the fibril and sequence optimized at the binding interface and validated by MD simulation and refolding calculations. The end result yielded a mutation in Thr25 of iFUS-G to

alanine, located on the helix of inhibitor which forms contacts with the fibril surface near FUS residues Thr78/Gly79 (Figure 2.3e). The T25A mutation resolved a mild clash between the threonine side chain on iFUS-G and backbone carbonyls of the FUS fibril. The improved inhibitor, termed iFUS-G2, shows better inhibition of stress granule formation in cells over iFUS-G (Figure 2.3f). To further verify that the interaction of the helix region of iFUS-G is important for inhibiting FUS phase separation, steric clashes were introduced into the inhibitor with T25Y and A21R mutations (Supplementary Figure 2.9). MD and *ab initio* structure prediction calculations show these mutations have no effect on the overall fold of the miniprotein, and actually improved its solubility during purification (data not shown). Introduction of the single A21R mutation reduces stress granule inhibition, while the double A21R/T25Y mutation nearly completely abolishes its inhibitory capacity.

Inhibitors reduce LLPS of FUS but not TDP43 or hnRNPA2

Our next goal was to determine if the observed effects of the designed FUS inhibitors are specific to FUS LLPS or non-specifically affect the LLPS of other low-complexity protein domains. LCDs of hnRNPA2 (res. 181-342) and TDP43 (res. 286-331) are similar in sequence composition to FUS_{1-214} (Figure 2.4a). All three protein segments are derived from the glycine-rich domains of their parent proteins, with enrichment of not only glycine but also serine, glutamine, asparagine, tyrosine, and alanine. Both the low complexity domains of hnRNPA2 and TDP43 are capable of LLPS both *in vitro* and within the cell. In addition, the atomic-level fibril structures of both protein domains are known and do not resemble that of the FUS LCD fold[36]. As previously discussed, inhibitors iFUS-G, iFUS-C, and iFUS-H all reduce FUS_{1-214} LLPS *in vitro* when the sample is cooled to from 4°C to 37°C, as measured by ThT fluorescence (Figure 2.4b).

Both hnRNPA2$_{181-342}$ and TDP43$_{286-331}$ shown similar phase separation behavior in this assay, with ThT fluorescence spiking shortly after the sample is cooled. All three FUS inhibitors have no effect on hnRNPA2$_{181-342}$ phase separation, and only iFUS-H appears to significantly affect TDP43$_{286-331}$.

It has been previously demonstrated that osmotic stress can lead to the incorporation of TDP43 and hnRNPA2 into stress granules[37,38]. As was done for FUS, we used a HEK293T-based cellular assay in which mCh-TDP43$_{288-331}$ or mCh-hnRNPA2$_{181-342}$ were transiently expressed in cells transduced with the designed FUS inhibitors. Fluorescent microscopy of 0.4 M sorbitol stressed cells shows that both TDP43 and hnRNPA2 form discrete and quantifiable puncta in response to the stress (Figure 2.4c). Addition of the FUS inhibitor iFUS-G (10 μM) leads to no significant change in the number of quantified granules for both TDP43 and hnRNPA2 (Figure 2.4d-e).

DISCUSSION

The number of known cellular processes, both functional and pathologic, driven by membraneless organelles formed through LLPS is rapidly increasing. Since the initial observation of liquid-like assemblies in the cell by Wilson in 1899, dozens of native cellular functions have been attributed to intracellular condensates[39]. These include ribosome or histone assembly, RNA transport, nuclear pore formation, stress response, regulation of transcription factors, and many more[40]. Similarly, associations of LLPS to disease continue to grow. ALS-linked mutations in the LCD of hnRNPA1 and hnRNPA2 have been demonstrated to potentiate their phase separation into liquid droplets and the eventual formation of solid aggregates[41,42]. Similarly, mutations in FUS associated with increased LLPS and stress granule formation have been tied to ALS[6]. Recent

evidence even points to tau, a primary amyloid component of AD, being capable of LLPS[43]. Phase separation pathology also extends beyond the brain, and is linked to several cancers, such as Ewing sarcoma[44]. Because of all these associations with disease, there is an ever-increasing need of drugs capable of modulating LLPS. However, due to the number of fundamental cell processes that rely on phase separation, pharmacologically modulating pathologic LLPS runs the risk of interfering with these functions. Therapies targeting aberrant stress granule formation in the brain of an ALS or FTD patient could potentially lead to significant off target effects in DNA synthesis or gene expression, for example, if not carefully considered.

One path forward to guide therapeutics to exclusively target pathological LLPS is through the use of protein structural knowledge. In this work, we develop inhibitors which specifically prevent the LLPS of the FUS LCD but have little to no effect on phase separation of hnRNPA2 and TDP43. Similarly, the inhibitors prevent fibrillization of FUS but not of other amyloid proteins tau and alpha-synuclein. We believe this specificity is achieved through directly targeting the fibrillar structure of FUS, which has a distinct fibril structure from both hnRNPA2 and TDP43 LCDs, as well as tau or aSyn. Despite the wide prevalence of biomolecular condensates in human health and disease, much remains unknown about the molecular forces driving their formation. Evidence points to multi-valent interactions between segments of protein low-complexity domains driving LLPS, but the degree of structural order present within those interactions is debated[45]. It is argued that LLPS of low-complexity proteins does not contain any ordered structure but is instead a result of unstructured multivalent interactions. Because TDP43 and hnRNPA2 have similar low complexity sequence compositions to FUS, it is unlikely that we could achieve such specific inhibition if their phase separation was governed exclusively by non-specific interactions. A more likely explanation for these findings is that LLPS is driven by transient assembly of fibril-like

structural elements, which are precisely targeted by the inhibitors. Our structure-based designs, such as inhibitor iFUS-G, are capable of inhibiting both FUS LLPS and fibrilization (Supplementary Figure 2.10). This indirectly implies that both of these processes share some sort of structural similarities, at least transiently.

Unlike pathologic amyloid fibrils such as tau or alpha synuclein[26,28], FUS fibrils are not polymorphic in nature, similar to functional yeast prion fibrils. This may imply an evolved function of FUS fibril assemblies[46]. Additionally, FUS fibrils are much less energetically stable compared to other pathologic amyloid proteins, which may aid to their reversibility during native functions[36]. However, further investigation into the precise molecular structure of FUS and other biomolecular condensates is needed to make a definitive conclusion. It is also unknown how well the molecular structure of phase separation droplets *in vitro* matches that of intracellular phase separations such as stress granules. To that end, we do observe a correlation of inhibitory efficacy across phase separation experiments. Inhibitors iFUS-G, iFUS-C, and iFUS-H all reduce *in vitro* phase separation as well as stress granule formation. Likewise, iFUS-L is not effective at reducing either process.

The inhibitors described in this work were created using *de novo* protein design. Computational design of proteins with no sequence similarity to naturally occurring proteins is positioning itself as a new and exciting methodology for therapeutic and diagnostic development[25,47–49]. A major benefit of *de novo* design compared to a traditional small molecule approach is its modularity and ease of modification. Improvements to inhibitor iFUS-G to produce iFUS-G2 were made by a simple mutation at the protein binding interface, as were the negative binding controls iFUS-Gx/iFUS-Gxx. Neurodegenerative disease is an area of medicine with a

dearth of treatment options. Novel solutions such as structure-based *de novo* protein design may provide new routes for addressing this lack of available therapies.

Many outstanding questions remain regarding the inhibition of biomolecular condensates. Nucleic acids, particularly RNA, have play an important role in the formation and regulation of many membraneless organelles[50]. Disrupting or modifying their interactions with the protein components within LLPS droplets could prove another useful target for therapeutic design. Additionally, post-translational modification and upstream regulation of proteins involved in condensate formation also plays a large part in steering their function[51]. In this work, we describe an approach to target LLPS by directly targeting the protein components involved. Therapeutic development for diseases featuring phase separation pathology like ALS or FTD may need to consider all of these biological components to yield effective treatments.

MATERIALS AND METHODS

Computational design of miniprotein inhibitors

Initial design of the miniprotein inhibitor library was performed with Rosetta, following the general approach taken by Baker and colleauges[25]. Backbone topologies for the inhibitor scaffolds were selected using the blueprint functionality in RosettaRemodel[29], in which total miniprotein size, number and type of secondary structural elements, location and length of loops was designated. Final designs ranged between 35-50 amino acids, contained at least one alpha-helix and two or three beta sheets, yielding 7 unique classes (HEE, EHE, EEH, HEEE, EHEE, EEHE, EEEH). Once the backbone was generated, sequences for each design were added using the FastDesign algorithm. For each class, 15,000 inhibitors were designed, and the 5,000 lowest scoring designs were used for docking. Scoring of designs was based on total score, dihedral angles

of beta-sheet conformation (to ensure the orientation of beta-sheets in the inhibitors matched those of the FUS binding site), total number of unsatisfied hydrogen-bonds, and compactness of fold. Docking to the FUS fibril structure was done with the MotifGraft mover[31]. Each inhibitor was aligned to a native strand of the FUS fibril using a backbone RMSD alignment < 1 Å, allowing a beta-strand of the inhibitor to properly orient to the FUS fibril geometry. Once docked, residues near the fibril binding interface were sequence optimized to reduce clashes and improve binding energy. No HotSpot residues were chosen. Calculations were carried out with the REF2015 score function[52].

Molecular dynamics simulations were performed using GROMACS 2018[32] with the CHARMM27[53] all atom force field. Designed proteins were solvated in an explicit cubic water box, and counter ions were added. Following system pressure/temperature equilibration and energy minimization steps (100 ps each), designs were simulated for 200 ns, and ranked by average backbone RMSD. Lowest ranking designs were simulated for an additional 200 ns. GPU acceleration was enabled.

Refolding of designed inhibitors was done with AbinitioRelax in Rosetta[34]. Fragments for each calculation were generated with the online server Robetta[54]. For each design, 50,000 trajectories were calculated. Energies of each trajectory were compared to RMSD to the initial design and designs whose 50 lowest energy trajectories were lowest in RMSD to original design were chosen for expression and purification.

Protein expression and purification

Miniprotein inhibitors were expressed in *E. coli* BL21(DE3) gold cells using a pET28b(+) vector with a thrombin cleavable N-terminal His-tag. Bacterial cultures were grown shaking at 225 rpm at 37ºC to OD_{600} of 0.4-0.8. Protein expression was induced with 0.5mM IPTG and the cultures were cooled to 18ºC and left to shake overnight. The following day, cultures were centrifuged at 5000 rpm for 7 min to pellet cells. Cells were lysed by sonication for 3 minutes (3 sec on/ 3 sec off at 70% power) in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 15 mM imidazole, 0.1 mg/mL lysozyme, HALT protease inhibitor cocktail, 20 mM BME). Lysate was then sonicated at 15,000 rpm for 15 min. The supernatant was then run on a 5 mL Ni-NTA column. Column was washed with 50 mL of wash buffer 1 (50 mM Tris pH 7, 1 M NaCl, 15 mM imidazole, 20 mM BME), then 100 mL of wash buffer 2 (50 mM Tris pH 6, 1 M NaCl, 15 mM imidazole, 20 mM BME). Protein was eluted from column with elution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 500 mM imidazole, 20 mM BME). His-tags were cleaved using bovine thrombin, then further purified using size-exclusion chromatography.

Low complexity domains for FUS (res. 1-214) and hnRNPA2 (res. 181-342) were expressed in E. *coli* BL21(DE3) Gold cells using a pHis vector gifted from the McKnight lab. The pHis vector contains an N-terminal mCherry fusion and a His-tag for purification. Bacterial cultures were grown shaking at 225 rpm at 37ºC to OD_{600} of 0.4-0.8. Protein expression was induced with 0.5mM IPTG and the cultures were cooled to 20ºC and left to shake overnight. The following day, cultures were centrifuged at 5000 rpm for 7 min to pellet cells. Cells were lysed by sonication for 3 minutes (3 sec on/ 3 sec off at 70% power) in lysis buffer (50 mM Tris-HCL (pH 7.5), 500 mM NaCl, 1% Triton X-100, 2 M guanidine hydrochloride (GdnHCl), 0.4 mg/mL lysozyme, 20 mM β-mercaptoethanol (BME), HALT protease inhibitor cocktail). Lysate was then sonicated at 15,000

rpm for 15 min. The supernatant was then run on a 10 mL Ni-NTA gravity column at 4ºC. Column was washed with 300 mL washing buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 20 mM BME, 0.1 mM PMSF, 2 M GdnHCl). Resin was washed with an additional 50 mL of wash buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 20 mM BME, 0.1 mM PMSF, 2 M urea). Protein was eluted with elution buffer (20 mM Tris- HCl pH 7.5, 500 mM NaCl, 250 mM imidazole, 20 mM BME, 0.1 mM PMSF, 2 M Urea). Protein was concentrated, concentration was measured using a NanoDrop (abs: 280 nm) and stored at -80 ºC until used.

TDP43 (res. 286-331) was expressed and purified with a maltose-binding protein (MBP) solubility tag as previously described by Cao and coworkers[36].

FUS_{1-214} used for the fibrillization assays (with no mCherry tag), was expressed and purified as previously described by Murray and coworkers[21].

In vitro phase separation experiments

For temperature-induced phase separation assays, 50 μM of mCh-FUS_{1-214}, mCH-$hnRNPA2_{181-342}$, or MBP-$TDP43_{286-331}$ in 1x PBS were mixed in solution with 50 μM Thioflavin T and 50 μM miniprotein inhibitor (unless otherwise stated) in a NUNC black 96 well optical bottom plate (final volume 100 uL). Samples were kept at 37ºC for 1 hour prior to experiment (or until ThT fluorescence readings had equilibrated). At time t=0, the plate was transferred to a Torey Pines orbital plate shaker cooled to 4ºC and shaken at low speed. To measure ThT fluorescence, plate was momentarily transferred to an FluoStar Omega fluorescence plate reader (ex. 440 nm, ex. 480

nm), then immediately placed back onto the cooled shaker. Readings were taken over the course of 1 hour, and experiment was performed in triplicate.

For centrifugation-based experiments, 50 μM mCh-FUS1-214 was mixed with 50 μM iFUS-G or iFUS-H in a 0.5 mL microcentrifuge tube at kept at 37°C for one hour. Sample was centrifuged at 4°C for 1 hour at 20,000 x g. Supernatant of samples before and after centrifugation was then measured using a SpectraMax M5 reader, reading mCherry fluorescence (ex: 587 nm, em: 610 nm).

Fibrillization assays

For FUS fibrillization, 50 μM FUS_{1-214} (no mCherry fusion) was quiescently incubated in a black 384-well optical bottom plate in 1x PBS at 37°C with various concentrations of miniprotein inhibitor. Turbidity of the solution was measured every 15 minutes by absorbance at 380 nm over the course of 1 week using a FluoStar Omega fluorescence plate reader, each condition was run in triplicate. Samples were then images by transmission electron microscopy.

Fibrillization of both tau k18 and aSyn were performed in 1x PBS at 37°C, with 50 μM protein and 50 μM ThT in a black NUNC 96 well optical bottom plate. Samples were agitated using double-orbital shaking at 700 rpm in a FluoStar Omega fluorescence plate reader, and ThT fluorescence was measured (ex. 440 nm, ex. 480 nm) to assess aggregation.

Fluorescent and transmission electron microscopy

All fluorescence microscopy was performed using a ZEISS Axio Observer D1 fluorescence microscope.

For electron microscopy, 5 uL of samples were spotted onto Formvar Carbon film 400 mesh copper grids (Electron Microscopy Sciences) and incubated for 4 minutes, then grids were stained with 5 uL uranyl acetate solution (2% w/v in water) for 2 minutes. Excess solution was blotted off, and grids were dried for 30 minutes. TEM images were taken using a JEOL 100CX TEM electron microscope at 100 kV.

Cell-based assays

HEK293T cells were grown and maintained at 37°C in 5% CO_2 in a humidified incubator using DMEM (Life Technologies, cat. 11965092) supplemented with FBS (10% vol/vol; Life Technologies, cat. A3160401), Glutamax (1%; Life Technologies, cat. 35050061), and penicillin/streptomycin (1%; Life Technologies, cat. 15140122). 100 uL of cells were plated onto black 96-well optical bottom tissue culture plates and allowed to grow to ~60% confluence. DNA for mCherry-fused LCD proteins was cloned into a PiggyBac expression vector (System Biosciences) and transfected into cells using a 1:20 dilution of Lipofectamine 3000 in OptiMEM media with P3000. 24 hours later, miniprotein inhibitors were transduced into the cells using a 1:20 dilution of Lipofectamine 2000 in OptiMEM media. 24 hours after this, cells were osmotically stressed with sorbitol (final concentration 0.4 M). Following a 1-hour incubation, cells were imaged using a fluorescent microscope to measure mCherry fluorescence. 10 images of random locations were taken for each well. Number of stress granules was quantified from the images using ImageJ's particle analyzer utility. To normalize for protein expression, whole well

images were taken with a Celigo Image Cytometer (Nexcelom), measuring confluence on the red channel to quantify total expression level of mCh-fused proteins. Total number of puncta were normalized by these confluence values. Experiments were performed in triplicate.

Alpha-synuclein HEK293T biosensor cells were gifted from the Marc Diamond lab at UTSW. Cells were grown and maintained using the same conditions as described above, and similarly plated onto black 96-well optical bottom tissue culture plates and allowed to grow to ~60% confluence. Alpha-synuclein fibrils were incubated overnight with iFUS-G, then transduced into the cells using Lipofectamine 2000 in OptiMEM media. The number of aggregated puncta was measured 48 hours later using a Celigo Image Cytometer (Nexcelom), measuring confluence on the green channel to quantify well confluence. Total number of puncta per well were counted using ImageJ particle analyzer function, and total counts were divided by the % confluence in each well for normalization. Experiments were performed in triplicate.

Circular dichroism

CD spectroscopy of designed miniproteins was carried out at a concentration of 0.20 mg/mL in 1x PBS buffer (pH 7.4) using in a crystal cuvette (path-length of 1mm) with a JASCO J-715 spectrophotometer. A PBS buffer blank was used for background subtraction.

ACKOWLEDGEMENTS

KAM is supported by the UCLA-Caltech Medical Scientist Training Program (GM08042) and UCLA Chemistry-Biology Interface training grant (USPHS National Research Service Award 5T32GM008496). We thank the Steve McKnight lab at Univ. of Texas Southwestern for protein

expression vectors and David Baker at Univ. of Washington for input regarding protein design. Funding for this work is provided by the Howard Hughes Medical Institute the National Institutes of Health.

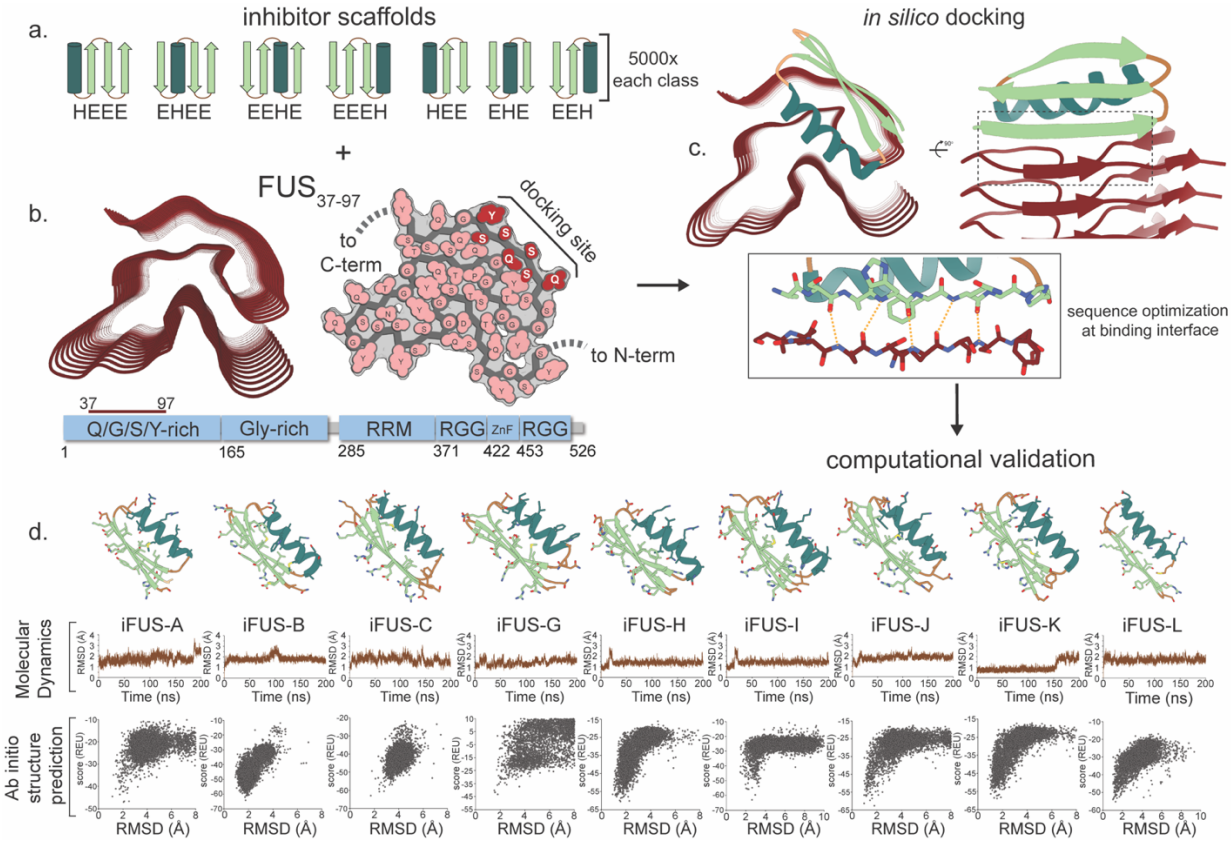


Figure 2.1: Computational design of FUS aggregation inhibitors. **a**. A library of *de novo* designed miniproteins was generated as inhibitor scaffolds. Seven unique classes were created, based on orientation of secondary structural elements (H=helix; E=sheet). **b**. The amyloid fibril structure of FUS served as the target for inhibitor design. The structure consists of residues 37-97, found in the N-terminal low-complexity, Q/G/S/Y-rich region of the protein. A single site near the N-terminus of the fibril structure was selected for inhibitor docking, residues 84-90. **c**. *In silico* docking of the designed miniprotein inhibitor library to the FUS structure was performed using Rosetta. Inhibitors dock to the exposed end of the fibril structure, capping its growth by preventing further addition of monomers. Upon docking, residues of the miniprotein inhibitor near the FUS binding interface are modified to optimize binding energy. **d**. To validate the folds and stability of top-ranking inhibitors, secondary computational methods were used. Long-range molecular dynamics simulations were used to assess the stability of each fold in a dynamic system over time.

Additionally, *ab initio* structure prediction, in which the inhibitor fold is predicted based on the primary sequence, was used to predict how likely the designed structures were to properly fold into their desired conformations. Nine designs were selected from the computational screen for experimental characterization.

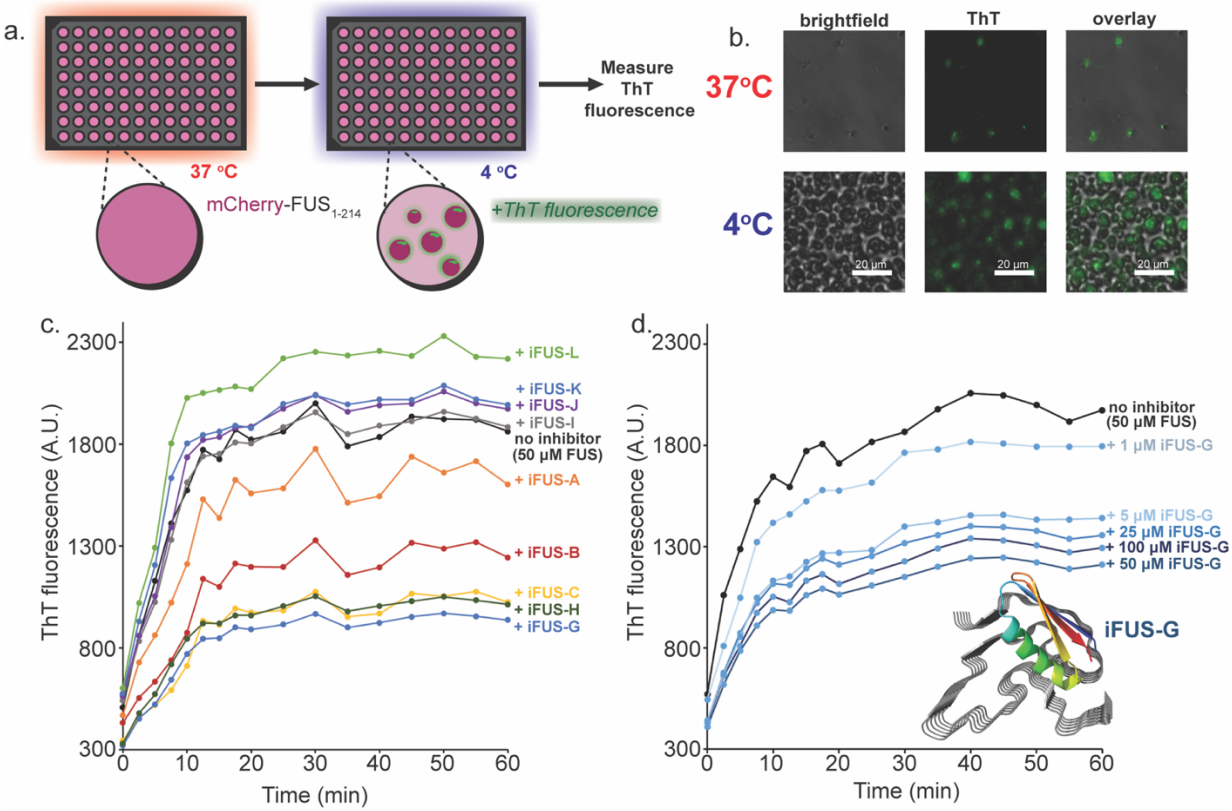


Figure 2.2: Designed inhibitors reduce FUS liquid-liquid phase separation in vitro. a. At 37°C the FUS low-complexity domain, in solution with Thioflavin T, shows little phase separation or fluorescence. When cooled to 4°C, phase separation droplets with bright ThT fluorescence rapidly appear. **b**. Brightfield and fluorescent microscopy images of 50 µM mCh-FUS$_{1-214}$ show little phase separation and ThT fluorescence at warmer temperatures, but abundant phase separation and fluorescence when cooled. ThT fluorescence can be seen localizing within the phase separation droplets. **c**. 50 µM mCh-FUS$_{1-214}$ was co-incubated with equimolar ratios of miniprotein inhibitors at 37°C prior to measurement. At time=0 min., the solution was cooled to 4°C and ThT fluorescence was measured to assess phase separation. Effects of inhibitors varied from inhibiting phase separation (iFUS-C/G/H) to promoting it (iFUS-J/K/L). **d**. Increasing concentration of inhibitor iFUS-G reduce FUS phase separation. At concentrations above equimolar ratios (inhibitor:FUS), the effects of iFUS-G are reduced.

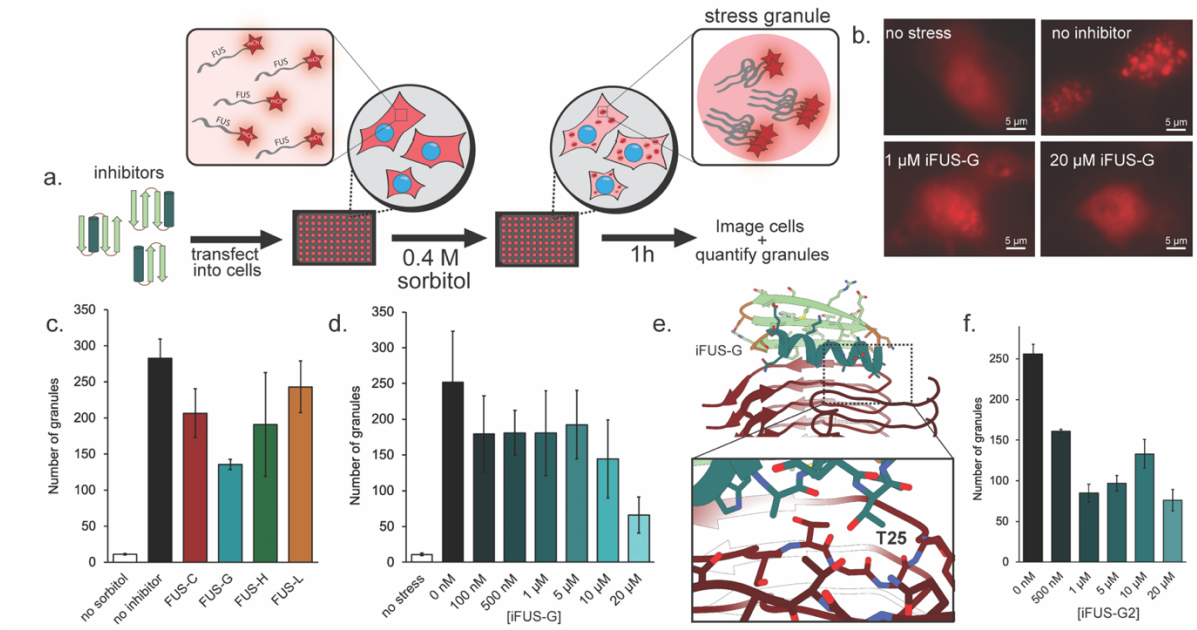


Figure 2.3: Miniprotein inhibitors prevent the incorporation of FUS into stress granules. a. A cell-based stress granule assay was used to assess the effects of the designed miniproteins on intracellular LLPS. Inhibitors were transfected into HEK293T cells expressing mCh-FUS1-214 and cells were later stressed with 0.4 M sorbitol to induce osmotic shock, which leads to aggregation of FUS in the form of stress granules. Following a 1-hour incubation with the sorbitol, cells were imaged, and the numbers of stress granules were quantified. **b**. Transfection of inhibitors into the cells resulted in a reduction in stress granule formation, as seen by fluorescent microscopy of cells treated with inhibitor iFUS-G. **c**. Quantification of stress granules in cells treated with different inhibitors. Cells not treated with sorbitol displayed little to no quantifiable granules. **d**. iFUS-G leads to a concentration-dependent reduction in stress granule formation. **e**. To improve the efficacy of iFUS-G, the inhibitor was re-run through the computational design pipeline (see Figure 2.1) to further improve FUS binding. The optimized inhibitor, iFUS-G2, featured a T25A mutation, in which the original threonine-25 residue of iFUS-G was converted to an alanine to avoid minor clashes with the FUS fibril backbone/sidechains. **f**. The optimized iFUS-G2 shows

improved reduction of stress granule formation, notably at 1 μM concentration. All error bars indicate +/- standard deviation.

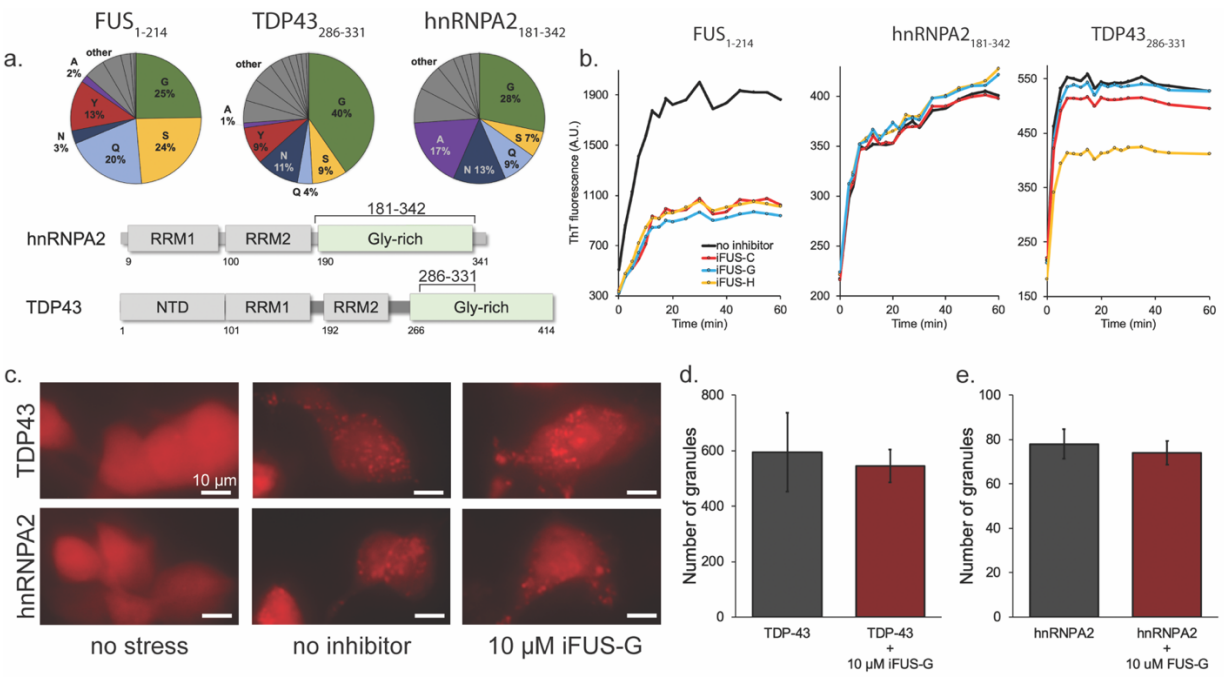
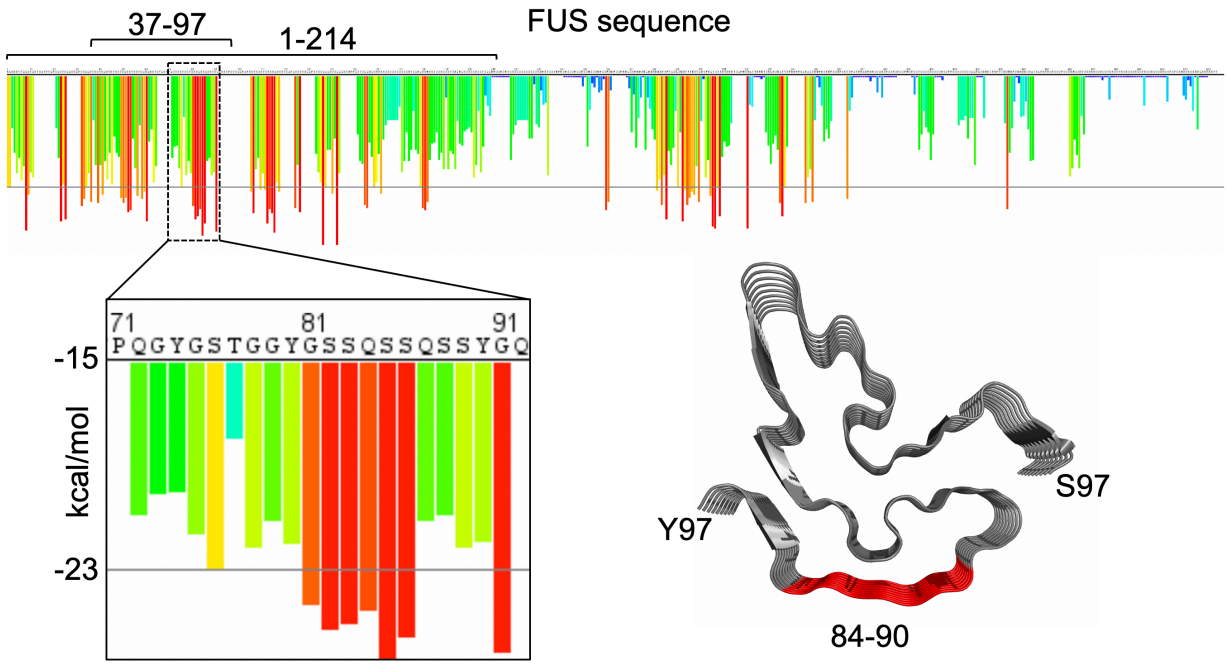


Figure 2.4: Designed miniproteins do not inhibit phase separation of TDP43 or hnRNPA2 low-complexity domains. a. Sequence compositions of FUS_{1-214}, $TDP43_{286-331}$, and $hnRNPA2_{181-342}$. All three low complexity domains have enrichment for glycine, serine, glutamine, and asparagine. FUS and $TDP43_{286-331}$ also show enrichment of tyrosine while $hnRNPA2_{181-342}$ has a high prevalence of alanine. The low complexity segments for both hnRNPA2 and TDP43 are derived from the glycine-rich domains of the full-length parent proteins. **b**. In vitro phase separation of FUS, hnRNPA2, and TDP43 LCD's with designed FUS inhibitors. Inhibitors iFUS-C, iFUS-G, and iFUS-H all significantly reduce FUS_{1-214} phase separation as measured by ThT fluorescence. The inhibitors show little effect on $hnRNPA2_{181-342}$ and $TDP43_{286-331}$, with the exception of iFUS-H and $TDP43_{286-331}$. **c**. Quantification of intracellular TDP43 stress granules with iFUS-G. Fluorescent TDP43 stress granules formed in sorbitol-stressed HEK293T cells overexpressing TDP43 are unaffected by the addition of FUS inhibitor iFUS-G. **d**. iFUS-G also has no effect on sorbitol-induced hnRNPA2 stress granule formation. e. Fluorescent microscopy of HEK293T cells overexpressing mCherry tagged TDP43 or hnRNPA2 LCD's. No granules are

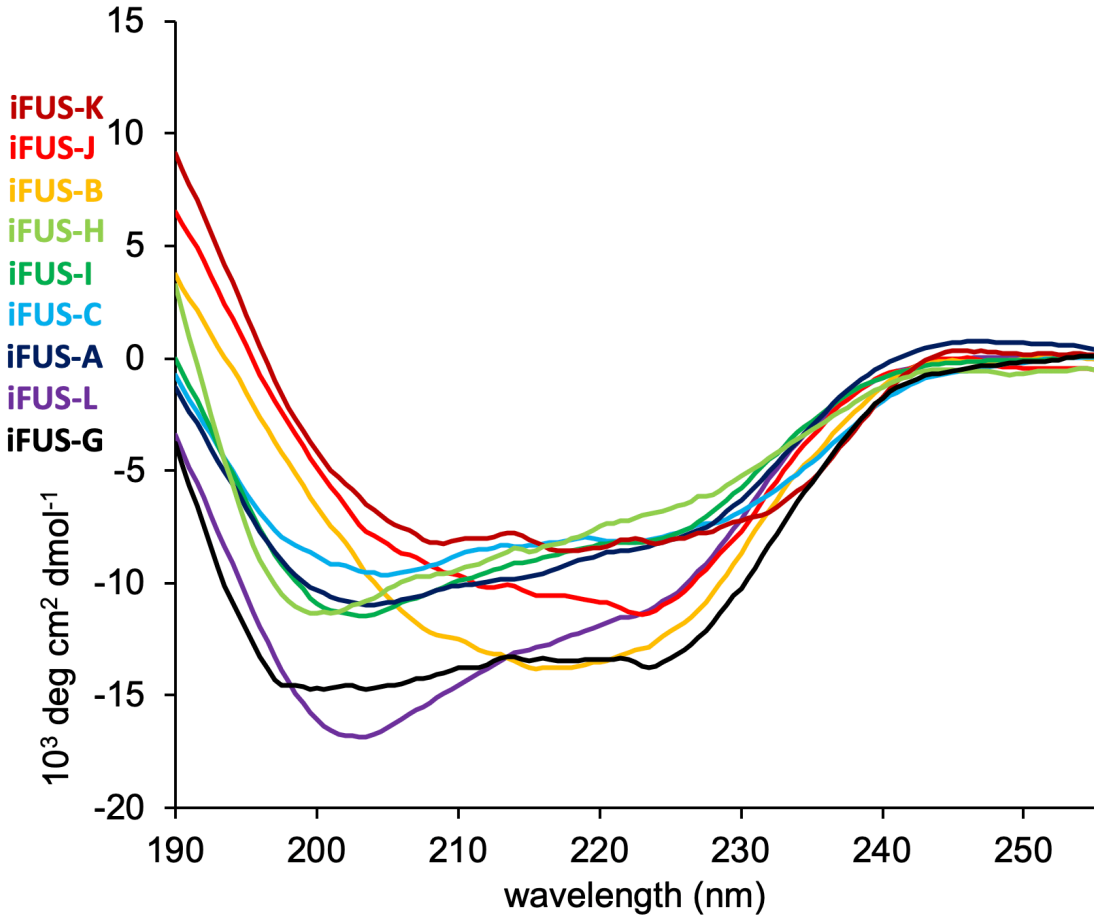
observed in “no stress” conditions (0.4 M sorbitol), while extensive granule formation is observed in stressed cells with and without added iFUS-G inhibitor. All error bars indicate +/- standard deviation.



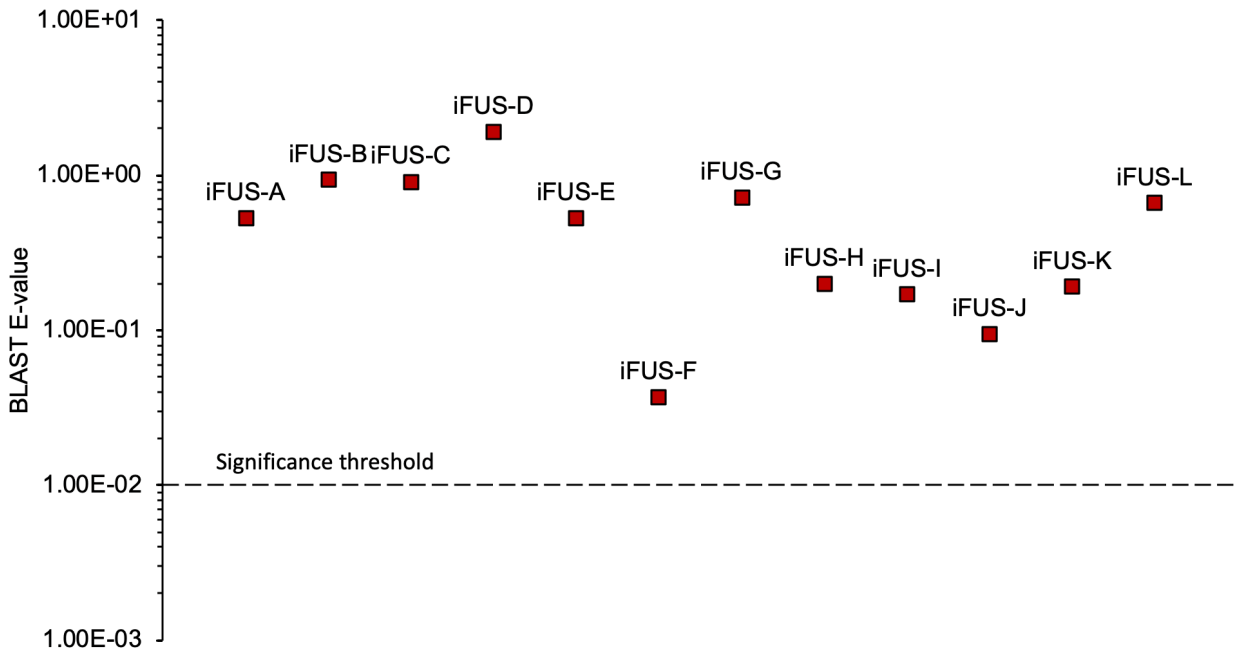
Supplementary Figure 2.1: Amyloidogenic regions of FUS. ZipperDB plot of FUS sequence to predict which regions of FUS are most likely to form amyloid structure. Bars indicate the score for the amino acid listed above, plus the following five residues. Lowest scoring energies (colored red) indicate those sequence elements most amyloid prone, with -23 kcal/mol set as the empirical threshold value. Residues 1-214, the low complexity region of the protein, have an enrichment of amyloidogenic segments, particularly residues 37-97, which compose the core of the determined fibril structure. Within that region, residues 84-90 are particularly amyloid-prone, and thus serve as the targeted binding site for inhibitor design.

inhibitor	sequence	inhibitor class
iFUS-A	SNAYSEAKGADPRNVNRAMKLAAEAAGADIHIDTSGDTIRITIHA	EHEE
iFUS-B	ARSSTVAQGSDPTIIRIIAWAAAWANNASMDDQTSGDTITIRIQQ	EHEE
iFUS-C	ASAHFQAHGQDAKKLATLLQRAAAAQGGRVDVKQDGDTVHLNLH	EHEE
iFUS-D	PRAHMQAHGIDEEQAKAVADAASWAAGGPVELRKDGDTFRVDVK	EHEE
iFUS-E	PRDHMKAVGADEWQIRAAAQAAADAAGGKNDVKRDGDTVEVNIQV	EHEE
iFUS-F	PRAVSSYNNGNAKQIQEAATKAAETAGTDVKVDNQGETIEVRMQY	EHEE
iFUS-G	ACAYAVFRGASEEQAKAWAKAMAD**T**LGAEVRMSTDGDVIHVTVEQ	EHEE
iFUS-G2	ACAYAVFRGASEEQAKAWAKAMAD**A**LGAEVRMSTDGDVIHVTVEQ	EHEE
iFUS-Gx	ACAYAVFRGASEEQAKAWAK**R**MAD**T**LGAEVRMSTDGDVIHVTVEQ	EHEE
iFUS-Gxx	ACAYAVFRGASEEQAKAWAK**R**MAD**Y**LGAEVRMSTDGDVIHVTVEQ	EHEE
iFUS-H	SNHSMSTRNDDSRAYKKAAKAAHDILGADVKVSEDGDEVEIRIHV	EHEE
iFUS-I	PRIEMVAVGIDWRQLHKAAKAAAEAANGDVRVTVDGDTVHVTIQ	EHEE
iFUS-J	DPTVETLAKQFKEAAEAAGAERVRVEVKDGKTVRVEVNGSFRASSHMTA	HEEE
iFUS-K	SSAMMRARGTDASTFETQVEAAATAAGAKVEHRRDGDTVTVHVHY	EHEE
iFUS-L	PLTVHSSTSGPRYVDSKASSDSEAAKAAWAAAKEAADRM	EEH

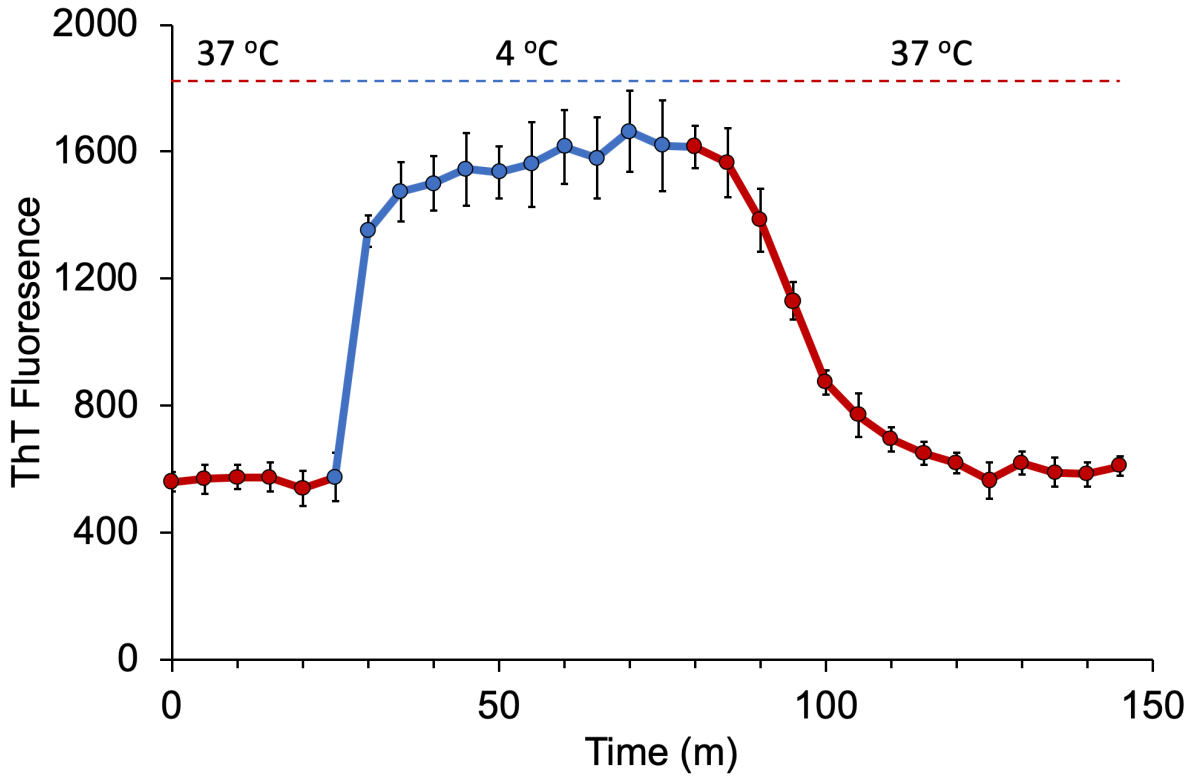
Supplementary Table 2.1: Sequences and class type of designed FUS inhibitors. Inhibitor iFUS-G was mutated (T25A) to yield the improved iFUS-G2 (indicated in green). Likewise, mutations in iFUS-G also were used as negative controls iFUS-Gx (A21R) and iFUS-Gxx (A21R/T25Y), indicated in red.



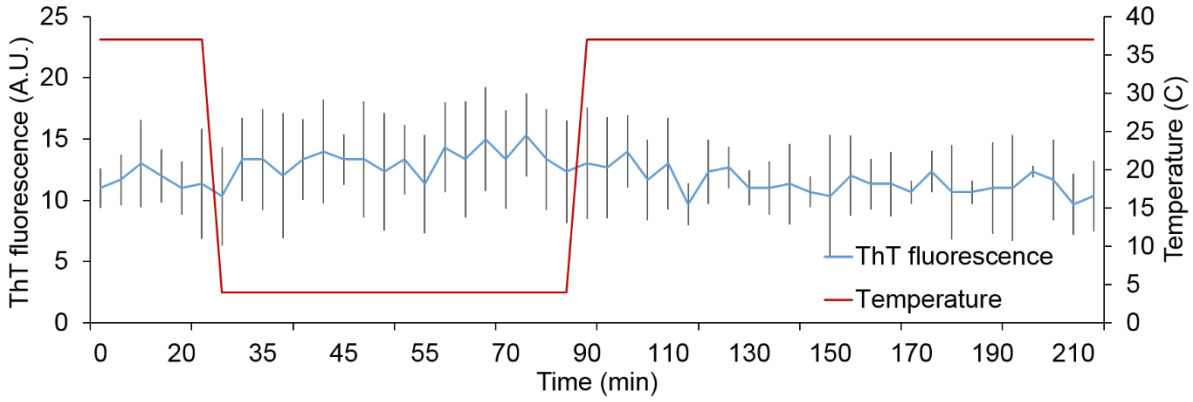
Supplementary Figure 2.2: Circular dichroism spectra for designed FUS inhibitors.



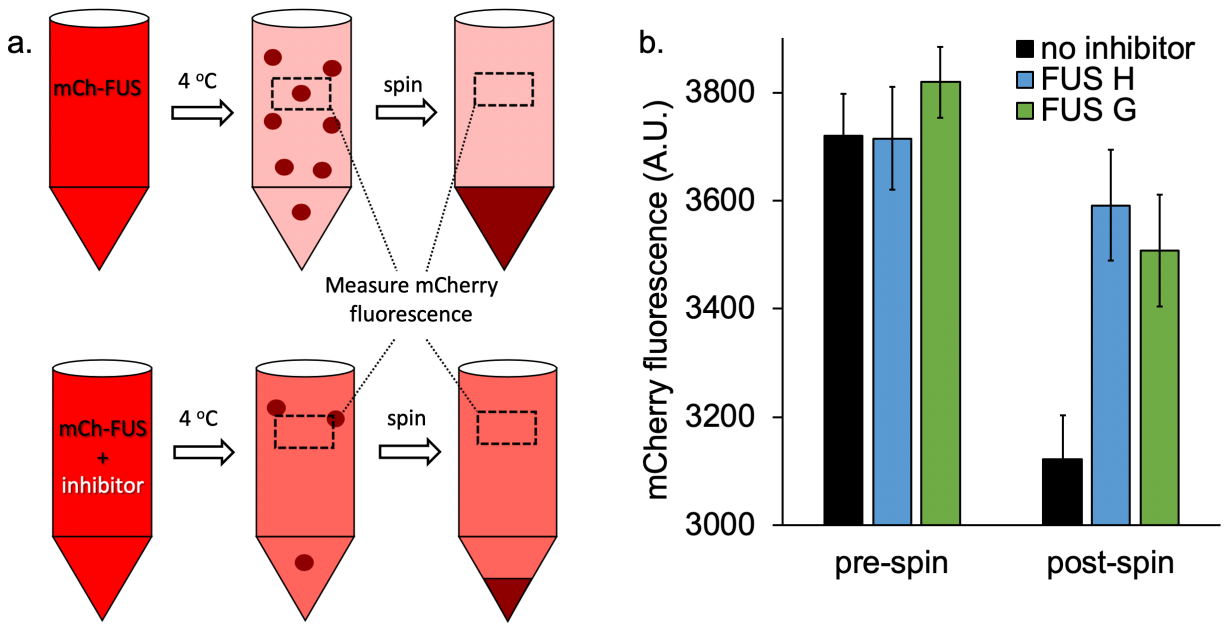
Supplementary Figure 2.3: BLAST E-values of designed inhibitors. The inhibitors are designed *de novo* without similarity to naturally occurring proteins. All designs are above the E-value threshold of 1.00E-02, the statistical cutoff for significant sequence similarity to another known protein sequence.



Supplementary Figure 2.4: FUS engages in reversible phase separation as a function of temperature. ThT fluorescence of mCh-FUS_{1-214} remains constant while maintained at 37°C. When cooled to 4°C, ThT fluorescence rapidly increases and plateaus. When the sample is again returned to 37°C, ThT signal gradually returns to the original baseline, indicating dissolution of the ThT-positive phase separation droplets.

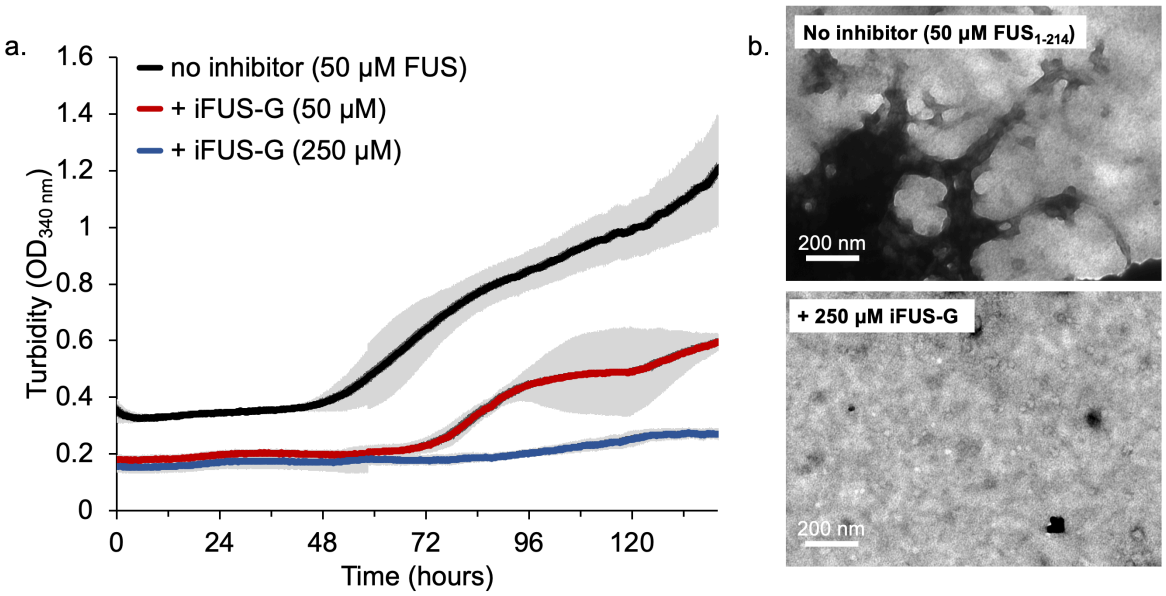


Supplementary Figure 2.5: ThT fluorescence at different temperatures. The fluorescence of 50 μM Thioflavin T in 1x PBS was measured at 37°C, 4°C, and again at 37°C. No significant change in fluorescence was measured as a function of temperature.

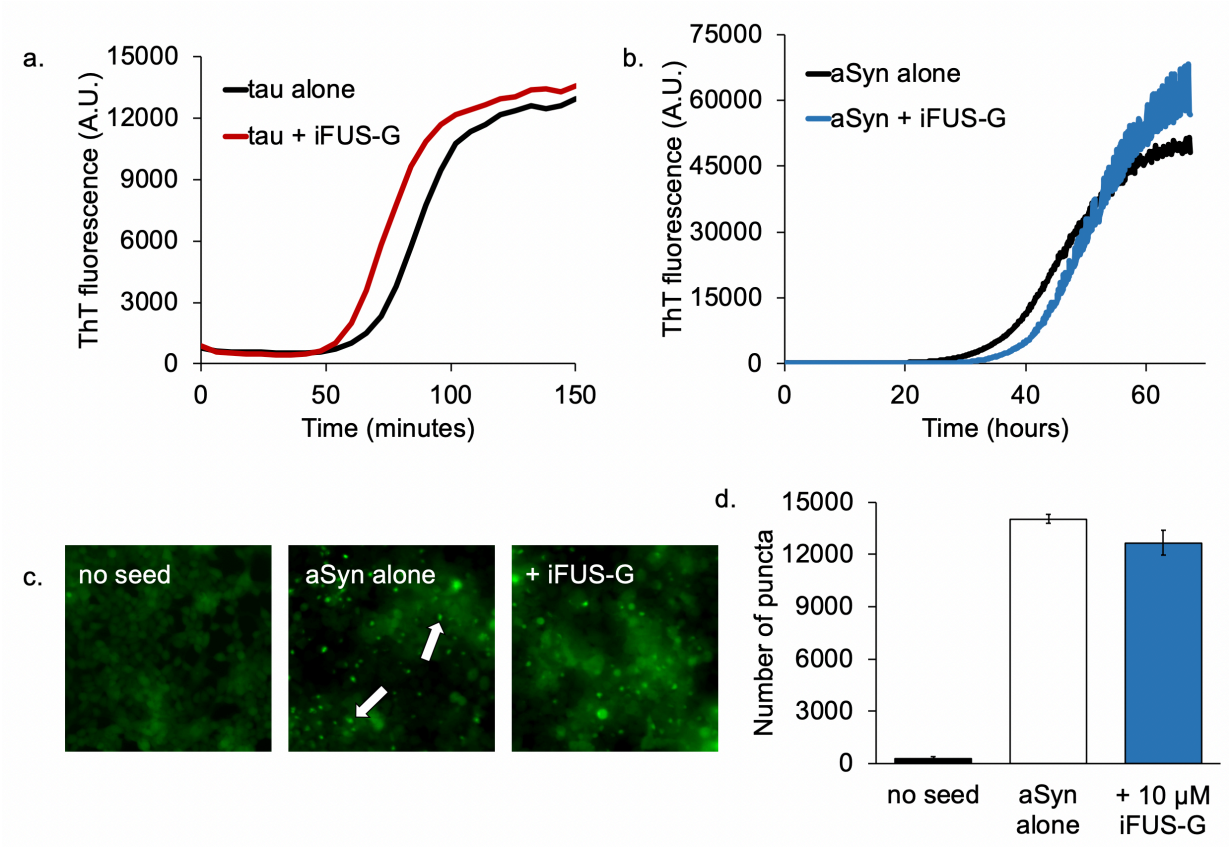


Supplementary Figure 2.6: *In vitro* analysis of effects of inhibitors on FUS phase separation *in vitro*. a. 50 µM of mCherry-fused FUS_{1-214} was mixed with 50 µM inhibitors iFUS-G or iFUS-H and cooled to 4°C in induce phase separation. Samples were then spun at high speed to pellet out the dense phase from the surrounding light phase. The mCherry fluorescence of the supernatant was then quantified before and after centrifugation. **b**. With no inhibitor added, a lower amount of mCherry is detected in the post-spin supernatant than compared to samples with inhibitor added. During protein LLPS in general, the protein concentrations in both the light and dense phase remain constant. More or less protein may enter the dense phase, but that will only influence to total volume of dense phase compared to light, not its concentration. If simply less FUS LLPS occurred as a result of the inhibitor, we would expect the mCherry concentration in the post-spin supernatant (the light phase) to be the same throughout all samples, and only the volume of dense phase in the pellet would change. However, as we observe a higher mCherry concentration in the light phase of inhibitor treated samples. This indicates that the inhibitors are enabling more soluble

protein to persist in the light phase before entering the dense phase. This could be due to the inhibitors altering the saturation concentration required for LLPS to occur.

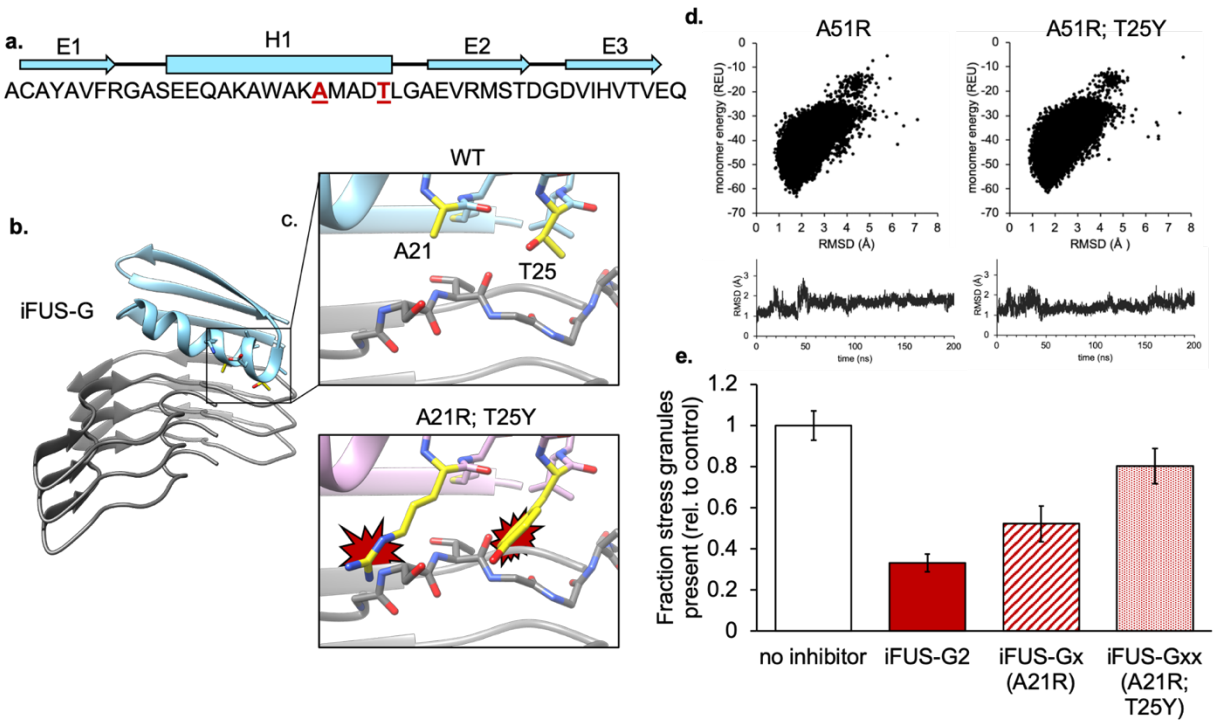


Supplementary Figure 2.7: **Effects of inhibitor iFUS-G on FUS fibrillization. a.** Increasing concentrations of iFUS-G reduce the aggregation of FUS_{1-214}, as measured by turbidity (OD_{340}). **b.** Electron microscope images reveal fibrillar aggregates of FUS form in the absence of inhibitor, but do not form in the presence of 250 µM iFUS-G.

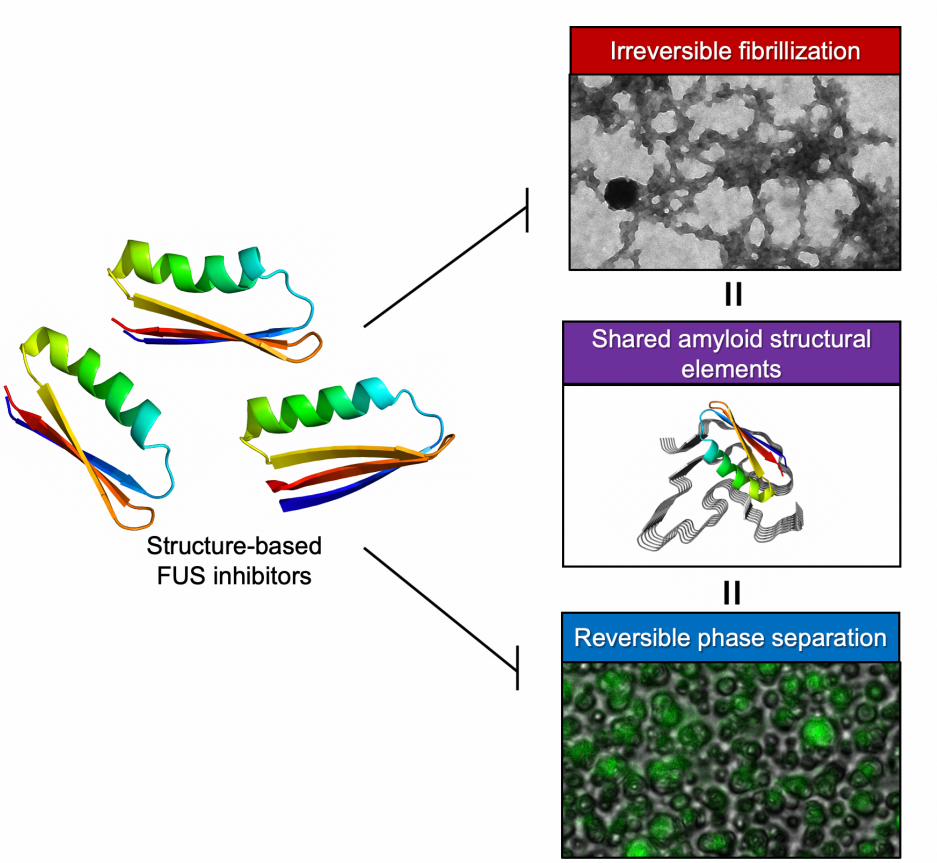


Supplementary Figure 2.8: Effects of iFUS-G on aggregation of other amyloid proteins. **a-b**. To assess if the designed FUS inhibitor iFUS-G non-specifically affects the fibrilization of amyloid proteins other than FUS, iFUS-G was tested in aggregation assays of tau and alpha-synuclein (aSyn). **a**. 50 µM of tau microtubule binding domain (k18) was mixed with 50 µM iFUS-G and allowed to aggregate by shaking at 37°C in PBS. Tau aggregation was monitored by Thioflavin T (ThT) fluorescence, revealing that iFUS-G only has a small effect on tau aggregation. **b**. Similarly, iFUS-G (50 µM) added to aSyn (50 µM) shows little effect on aSyn aggregation. **c**. Fluorescent microscope images of aSyn biosensor cells with no aSyn fibrils added show no visible puncta ("no seed"), while addition of fibrils results in many visible puncta ("aSyn alone"; indicated by white arrows). Incubation of the aSyn fibrils with iFUS-G (10 µM final concentration) prior to

transduction also results intracellular aSyn aggregation. **d**. Quantification of images reveals that iFUS-G has no significant effect on aSyn seeding.



Supplementary Figure 2.9: Designed steric clashes in the binding interface between iFUS-G and the FUS fibril reduce the efficacy of phase separation inhibition. a. Sequence of iFUS-G, highlighting secondary structure, and two residues (A21 and T25) in the binding interface with FUS. **b.** iFUS-G bound to FUS. **c.** Mutations introduced to disrupt the interaction are A21R and T25Y, both residues resulting in bulky steric clashes with the fibril. **d.** The mutations did not affect the overall computed fold or stability of the designed protein, as shown by *ab initio* refolding calculations (top) and molecular dynamics simulations (bottom). e. The effects of the mutations on FUS stress granule inhibition. Both the A21R single mutant and the A21R/T25Y double mutant reduce the inhibition of stress granule formation (inhibitor concentration = 1 μM).



Supplementary Figure 2.10: Proposed model of FUS inhibition. Inhibitors targeting the fibrillar structure of FUS are capable of inhibiting both FUS fibrillization and LLPS. This implies that both fibrillization and LLPS contain shared structural elements that drive their formation.

REFERENCES

1. Ferrari, R., Kapogiannis, D., Huey, E. D. & Momeni, P. FTD and ALS: a tale of two diseases. *Curr. Alzheimer Res.* **8**, 273–294 (2011).
2. Nowicka, N., Juranek, J., Juranek, J. K. & Wojtkiewicz, J. Risk Factors and Emerging Therapies in Amyotrophic Lateral Sclerosis. *Int. J. Mol. Sci.* **20**, (2019).
3. Tsai, R. M. & Boxer, A. L. Treatment of Frontotemporal Dementia. *Curr. Treat. Options Neurol.* **16**, 319 (2014).
4. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
5. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *J. Biol. Chem.* **294**, 7128–7136 (2019).
6. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
7. Nomura, T. *et al.* Intranuclear Aggregation of Mutant FUS/TLS as a Molecular Pathomechanism of Amyotrophic Lateral Sclerosis. *J. Biol. Chem.* **289**, 1192–1202 (2014).
8. Li, Y. R., King, O. D., Shorter, J. & Gitler, A. D. Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.* **201**, 361–372 (2013).
9. Zhang, P. *et al.* Chronic optogenetic induction of stress granules is cytotoxic and reveals the evolution of ALS-FTD pathology. *eLife* **8**,.
10. Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678–690 (2015).

11. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).

12. Griner, S. L. *et al.* Structure-based inhibitors of amyloid beta core suggest a common interface with tau. *eLife* **8**, e46924 (2019).

13. Sievers, S. A. *et al.* Structure-Based Design of Non-Natural Amino Acid Inhibitors of Amyloid Fibrillation. *Nature* **475**, 96–100 (2011).

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

15. 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).

16. Bloom, G. S. Amyloid-β and Tau: The Trigger and Bullet in Alzheimer Disease Pathogenesis. *JAMA Neurol.* **71**, 505–508 (2014).

17. Sangwan, S. *et al.* Inhibition of synucleinopathic seeding by rationally designed inhibitors. *eLife* **9**,.

18. Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6469–6473 (1998).

19. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β-sheets that assemble networks. *Science* **359**, 698–701 (2018).

20. Gui, X. *et al.* Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* **10**, 1–12 (2019).

21. 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).

22. Murray, D. T. & Tycko, R. Side Chain Hydrogen-Bonding Interactions within Amyloid-like Fibrils Formed by the Low-Complexity Domain of FUS: Evidence from Solid State Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **59**, 364–378 (2020).

23. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

24. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016).

25. Chevalier, A. *et al.* Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).

26. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9**, 1–10 (2018).

27. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).

28. Fitzpatrick, A. W. & Saibil, H. R. Cryo-EM of amyloid fibrils and cellular aggregates. *Curr. Opin. Struct. Biol.* **58**, 34–42 (2019).

29. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **6**, (2011).

30. Lin, Y.-R. *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5478–E5485 (2015).

31. Silva, D.-A., Correia, B. E. & Procko, E. Motif-Driven Design of Protein-Protein Interfaces. *Methods Mol. Biol. Clifton NJ* **1414**, 285–304 (2016).

32. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).

33. Buchko, G. W. *et al.* Cytosolic expression, solution structures, and molecular dynamics simulation of genetically encodable disulfide-rich de novo designed peptides. *Protein Sci. Publ. Protein Soc.* **27**, 1611–1623 (2018).

34. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871 (2005).

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

36. 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).

37. 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).

38. McDonald, K. K. *et al.* TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum. Mol. Genet.* **20**, 1400–1410 (2011).

39. Wilson, E. B. The Structure of Protoplasm. *Science* **10**, 33–45 (1899).

40. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

41. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).

42. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

43. Wegmann, S. *et al.* Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, (2018).

44. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**, 163-178.e19 (2017).

45. Peran, I. & Mittag, T. Molecular structure in biomolecular condensates. *Curr. Opin. Struct. Biol.* **60**, 17–26 (2020).

46. Van Melckebeke, H. *et al.* Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 13765–13775 (2010).

47. Ng, A. H. *et al.* Modular and tunable biological feedback control using a de novo protein switch. *Nature* **572**, 265–269 (2019).

48. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).

49. Bick, M. J. *et al.* Computational design of environmental sensors for the potent opioid fentanyl. *eLife* **6**, e28909 (2017).

50. Yoshizawa, T., Nozawa, R.-S., Jia, T. Z., Saio, T. & Mori, E. Biological phase separation: cell biology meets biophysics. *Biophys. Rev.* (2020) doi:10.1007/s12551-020-00680-x.

51. Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).

52. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

53. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).

54. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526-531 (2004).

CHAPTER 3

Discovering amyloid-related diseases by mapping mutations in low-complexity protein domains to known pathologies: the case of Keratin-8 in alcoholic liver disease

INTRODUCTION

Protein aggregation is a pathological characteristic of many diseases, including Alzheimer's (AD) and Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), prion diseases, systemic amyloidosis, and even certain metabolic diseases like Type II diabetes. In these disease contexts, normally soluble, globular proteins assemble into long insoluble fibrils, known as amyloid[1]. Amyloid fibrils deposit within tissues, either intra- or extracellularly, where they interfere with native cellular functions, illicit toxicity, and contribute to disease pathogenesis. Atomic structures determined of amyloid fibrils reveal they are composed of repeating layers of protein monomers which have adopted a primarily beta-sheet conformation. The beta-sheets of the fibril core tightly interdigitate, forming a structural motif known as a steric zipper, in which the amino acid side chains interlock like the teeth of a zipper down the fibril axis[2]. These highly stable interactions likely contribute to the resistance of amyloid fibrils to denaturation and proteolytic degradation. Analysis of sequences capable of adopting steric zipper conformations has successfully identified protein regions which drive amyloid formation[3].

While protein aggregation was once believed to be strictly pathogenic, there is growing evidence that proteins also undergo a functional form of aggregation to drive native cellular processes. These states of protein self-assembly are often reversible and contribute to the formation of dynamic membraneless cellular bodies such as stress granules, P-bodies, Cajal bodies, and

nuclear paraspeckles[4,5]. Unlike disease-associated amyloid aggregates, these functional assemblies associate and dissociate reversibly. Hughes and colleagues demonstrated that this reversibility may be driven by specific structural motifs in protein low-complexity domains (LCDs). Termed LARKS (low-complexity amyloid-like kinked segments), these motifs form fibrillar structures with a kinked beta-strand conformation as opposed to the predominantly pleated strands of steric zippers in pathogenic amyloid fibrils[6]. In both reversible and pathogenic amyloid, identical beta-strands of different protein monomers, kinked or pleated, are stacked to form beta-sheets than run the length of the fibril. The proteins FUS, TDP43, and hnRNPA2 have an enrichment of LARKS and are all known to reversibly phase separate[7–9], and full-length fibril structures of both proteins exhibit a high degree of kinking within the beta-sheets[10] (hnRNPA2 structure unpublished). However, all three proteins are known to aggregate irreversibly in several diseases, and it has also been shown that fibrous FUS aggregates can arise from soluble phase separated droplets[11,12].

The molecular mechanisms that drive the transition from reversible to irreversible amyloid aggregation are still incompletely understood. A number of pathogenic mutations within the LCDs of proteins known to undergo both reversible and irreversible self-assembly have been demonstrated to increase aggregation propensity, including in FUS[13], hnRNPA1/hnRNPA2[11], and TDP-43[14]. Here we seek to better understand this transition from reversibility to irreversibility through structural and biochemical characterization of disease-related mutations occurring within LARKS. We use a computational screen to identify mutations in LARKS that may drive the transition from a kinked beta-sheet conformation to a pleated one. Our computational approach identified many known aggregation-promoting mutants while also finding new variants not previously associated with protein aggregation. Among these are mutations in the low-complexity region of Keratin-8 (KRT8). We demonstrate that LARKS

mutations in KRT8 promote both its reversible and irreversible aggregation, highlighting the amyloid-like nature of KRT8 aggregation. Furthermore, atomic structures of KRT8 LARKS reveal that pathogenic mutations convert the non-pleated wildtype protein into a pleated conformation, shedding light on the molecular mechanisms underlying aggregation-promoting mutations in low-complexity proteins.

RESULTS

Computational screen of disease-related mutations in LARKS

As irreversible amyloid fibrils are characterized by pleated beta-sheets and reversible LARKS have kinked beta-sheets, we hypothesize that mutations in reversibly aggregating proteins may promote pathogenic irreversible aggregation by driving a transition from a kinked to pleated beta-sheet conformation which favors steric zipper formation. To study this, we first used computational threading to identify LARKS, the kinked segments thought to drive reversible self-assembly, throughout the proteome. This was achieved by threading protein sequences onto the kinked backbones of three different LARKS structures (Figure 3.1a) and using the protein modeling software Rosetta to assess the energy of the structure. This threading approach has been successfully used to identify steric zipper segments in amyloid proteins[3]. Once identified, the LARKS sequences were cross-referenced with three mutational databases (OMIM[15], Uniprot[16], and ClinVar[17]) to identify disease-related mutations occurring within the LARKS. As a control, benign single-nucleotide polymorphisms (SNPs) occurring within the LARKS sequences were also analyzed. Analysis of the mutations reveals it is largely mutations from glycines to either polar or hydrophobic residues (Figure 3.1c), while benign SNPs led to less disruptive changes in amino acid properties (e.g. Gly to Ser) (Supplementary Figure 3.1).

The wild-type and mutant LARKS sequences were then threaded onto the backbone of a pleated steric zipper to assess if the mutation increased the propensity of the sequence to form a pleated sheet as opposed to a kinked one. The calculated energies of the threaded zipper structures versus the difference in energy between the wild-type and mutant zippers were plotted (Figure 3.1b). From this we observe a population of LARKS mutations that shows a decrease in zipper energy (i.e. more likely to form a zipper) in both their absolute score as well as compared to wild type. This cluster is also relatively free of benign SNPs. Within this population we find many mutations previously identified to increase aggregation propensity (Figure 3.1d). Examples include TDP-43 G294V[14], G294A[9], and G295S[9,14] and hnRNPA2 D290V[11,18] (Supplementary Table 3.1). We also observe a large enrichment of mutations occurring in the head domain of the intermediate filament protein Keratin-8 (KRT8).

KRT8, like other alpha-keratin proteins, contains a central coiled-coil domain flanked by an N-terminal head domain and a C-terminal tail domain which are both low-complexity in sequence composition[19]. The KRT8 mutants are composed of glycines converting to alanine, cysteine, or valine and tyrosines converting to either histidine or cysteine. Both glycines and tyrosines have been identified as critical residues for maintaining liquid-liquid phase separations of low-complexity proteins[20]. For experimental validation we selected three of these KRT8 mutants (G62C, Y54H, and G55A), all of which are found in patients with cryptogenic liver disease[21], as well as three FUS mutants (G191S, G225V, G230C).

Structure determination of wild-type and mutant LARKS of KRT8

To assess the structural impacts of the pathogenic mutants, we determined the atomic structures of several LARKS from KRT8 for both the wild-type and mutant sequences. Segment crystal structures of $KRT8_{58\text{-}64}$ (WT: SGMGGIT; G62C: SGMGCIT) for the wild-type and G62C

mutation were determined. Additionally, the structure of $KRT8_{52-58}$ (GGYAGAS) for the G55A mutation was also solved. High resolution diffraction data were collected for the wild-type sequence (GGYGGAS) of $KRT8_{52-58}$ but we were unable to determine the phases of the structure with molecular replacement using beta-strands as search models, presumably due to its adoption of a highly kinked conformation. However, a recently determined structure from hnRNPA1 (GGGYGGS) very closely resembles the wild-type $KRT8_{52-58}$ sequence, particularly the residues at and surrounding the mutation site, and was used for comparison[22].

WT $KRT8_{58-64}$ (SGMGGIT) crystallized as an anti-parallel class-7 steric zipper with a long PEG 1000 molecule binding along the fibril axis in the space between beta-sheets (Figure 3.2a)[2]. Two ethanol molecules are observed coordinated near the C-terminus. The interface appears to be formed by the hydrophobic interactions of the methionine and isoleucine with the tyrosine and backbone of the adjacent sheet. At glycine 62 (and glycine 61), we observe a highly extended backbone conformation, in which the dihedral angles at both residues approach nearly $|180^{o}|$. In contrast, the structure of $KRT8_{58-64}$ G62C (SGMGCIT) adopts are nearly completely pleated beta-sheet conformation (Figure 3.2b). A now class-2 parallel steric zipper, the sheets associate again through hydrophobic interactions with the methionine and isoleucine but are further stabilized by interaction with the mutant cysteine. Isopropanol is also observed coordinated in the zipper interface. Importantly, instead of the extended conformation adopted by the glycine in the wild-type structure, the mutant cysteine is in a pleated conformation typical of most irreversible amyloid proteins.

The G55A $KRT8_{52-58}$ (GGYAGAS) structure forms an anti-parallel class-6 steric zipper, with sheets interacting via the tyrosine and two alanines, including the G55A mutant alanine (Figure 3.2c). The glycine at the N-terminus assumes a highly extended conformation like the

SGMGGIT structure. But similar to the other mutant structure, the backbone at the mutant alanine is pleated. Comparison with GGGYGGS from hnRNPA1, which shares the same GGYGG sequence motif as the wild-type $KRT8_{52-58}$ (GGYGGAS), shows that a glycine at that position also adopts a highly extended backbone conformation. An overlay of both WT/mutant structure pairs highlights that mutations from wild-type glycines facilitate a transition in backbone from a non-ideal, highly extended conformation to a pleated one (Figs 2d-e). From these fibril structures we observe that both the G62C and G55A mutations in KRT8 convert a non-pleated beta-sheet into a pleated one, transitioning a LARKS-like conformation into a steric zipper.

Characterization of the aggregation-promoting effects of pathogenic KRT8 mutations

Our structural analysis highlighted a key difference in the backbone conformations between the wild-type and mutant KRT8 fibrils. We next sought to characterize the effects these mutations have on KRT8 aggregation. The head domain of KRT8 (residues 1-90) was expressed and purified with an N-terminal mCherry solubility tag for the wild-type and G62C, Y54H, and G55A mutants. Aggregation kinetics assays using the fluorescent amyloid dye Thioflavin T (ThT) reveal that extended shaking at 37ºC results in the aggregation of all four $KRT8_{1-90}$ constructs, however the three mutants aggregated much more quickly and extensively compared to wild-type (Figure 3.3a). Electron micrographs of each sample revealed clumped aggregates with similar morphologies for both the wild-type and mutants (Figure 3.3b). X-ray powder diffraction of each sample displayed a cross-beta diffraction pattern typical of amyloid fibrils. For all four samples we observe diffraction rings near 10 Å and ranging between 4.3-4.6 Å (Supplementary Figure 3.2). While most amyloid proteins exhibit diffraction rings closer to 4.8 Å, diffraction of previously determined LARKS structures also displayed rings near 4.5 Å,

indicating that such a diffraction pattern may be a feature of low-complexity protein aggregates. Other intermediate filament proteins known to phase separate also have diffraction rings near 4.6-4.7 Å[23]. Stability analysis was performed on the aggregates, showing that the mutants were slightly less resistant to SDS denaturation compared to wild type (Supplementary Figure 3.3).

Having observed the effects of each mutation on irreversible aggregation, we next asked if these mutations affect reversible keratin aggregation. Keratins and other intermediate filament proteins have been shown to phase separate reversibly both *in vitro* and in cells[23]. To quantitatively measure the effects each mutant has on phase separation, we developed a fluorescent assay in which the sample is maintained at 37ºC, then cooled to 4ºC to induce phase separation, and then warmed back to 37ºC to melt the phase separations. Thioflavin T fluorescence occurs during phase separation at cool temperatures but diminishes when the sample is warmed (Figure 3.3e). This effect has been previously demonstrated to occur for phase separation droplets of hnRNPA1[22]. Wild-type and mutant samples at 37ºC showed a low baseline of ThT fluorescence, but when cooled to 4ºC, all mutants, particularly G55A and Y54H, showed much more robust phase separation compared to wild-type (Figure 3.3f). It then took several hours of melting at 37ºC to return back to baseline. This same series of experiments was performed using FUS and three disease-related mutations (G191S, G225V, and G230C) identified in our screen (Supplementary Figure 3.4). The results are similar to KRT8, in that each of the FUS mutants reversibly aggregated much more extensively than the wild type. Interestingly, the FUS phase separations took less than an hour to melt back to baseline, whereas KRT8 took nearly five hours.

The effects of ethanol and the seeded aggregation of KRT8 from ASH liver tissue

KRT8 is known to aggregate pathologically, primarily in the form of Mallory Denk bodies (MDBs) in liver disease. Given our structural and biochemical findings of KRT8 aggregation, as well as the fact that MDBs are most frequently observed in ASH and alcoholic cirrhosis, we next sought to further explore KRT8 aggregation in the context of alcoholic liver disease[24]. Two out of three of the crystal structures of KRT8 segments that were solved showed the peptides in complex with either ethanol or isopropanol, each with distinct binding sites (Figure 3.4a-b). To further assess the effects of ethanol on KRT8 aggregation, wild-type $KRT8_{1-90}$ was aggregated in the presence of 0%, 1%, and 2.5% ethanol. Increasing concentrations of ethanol had a pronounced effect on promoting KRT8 aggregation (Figure 3.4c), while it had no effect on the aggregation of tau (Figure 3.4d), and mild effects on alpha-synuclein (Supplementary Figure 3.5), both of which are known amyloid proteins.

The *in vitro* aggregation of canonical amyloid proteins can be seeded by the addition of tissue that contains amyloid aggregates. For example, aggregation of tau can be seeded by the addition of Alzheimer's disease brain extracts containing neurofibrillary tangles, amyloid aggregates composed of tau[25]. To test the seeded aggregation of KRT8, we aggregated monomeric $KRT8_{1-90}$ in the presence of liver tissue extracts from patients with either ASH or hepato-cellular carcinoma (HCC) (Figure 3.4e). The HCC samples had no effect on KRT8 aggregation, while two out of three of the ASH samples significantly enhanced aggregation. Similarly, we observed KRT8 aggregation was also seeded by the addition of preformed recombinant aggregates (Supplementary Figure 3.6). We next aimed to identify a compound which could mitigate KRT8 aggregation. A small screen of ~20 on-hand compounds either known or predicted to reduce amyloid aggregation, including EGCG, curcumin, methylene blue, and others, identified the antibiotic demeclocycline HCl as capable of reducing KRT8

aggregation at even low nanomolar concentrations (Figure 3.4f). Both demeclocycline and EGCG also reduced KRT8 phase separation (Supplementary Figure 3.7).

DISCUSSION

Our principal finding is that alcoholic liver disease appears to be an amyloid-related condition, in which KRT8 aggregation is associated with pathology. The same general procedure that led us to this conclusion can possibly be followed to discover other diseases that have amyloid involvement. First, assessing the propensity of sequences to adopt LARKS conformations can identify LCD-containing proteins able to engage in amyloid-like interactions. Of the ~1700 proteins predicted to contain LARKS, many of them contain disease-related mutations within the LARKS sequence. In this study, we only analyze the 400 top-ranked LARKS (as scored by the structural threading algorithm), leaving many more to be screened and characterized[6]. We also focus exclusively on single missense mutations, but investigation into other types of variants, such as deletions of translocations, may prove useful for future investigation[26]. Second, by predicting the impact the mutations in LARKS have on potential steric zipper formation, one can identify mutations that may impact protein aggregation. Expansion of this approach to include disease-associated mutations within entire LCDs, not just LARKS, may cast an even broader net to find more aggregation-promoting mutations. By identifying pathogenic mutations predicted to enhance steric zipper formation, this workflow provides an approach to find proteins with a baseline propensity to aggregate in a disease-associated context.

Our analysis of KRT8 shows that each of the mutations predicted to promote steric zipper formation accelerate aggregation and confirm that KRT8 can undergo amyloid-like aggregation. While not widely regarded as an amyloid protein, KRT8 is known to aggregate in liver disease within Mallory-Denk bodies[27], and all three aggregation-promoting mutations analyzed in this

study were initially associated with liver disease[28]. Composed primarily of aggregated KRT8 and lesser amounts of KRT18, MDBs are cytoplasmic inclusions that occur within hepatocytes most frequently in alcoholic steatohepatitis (ASH) or alcoholic cirrhosis[29]. Some have previously speculated about the amyloid nature of MDBs. They have been demonstrated to bind luminescent conjugated oligothiophenes (LCOs) h-HTAA and p-FTAA, both of which selectively bind proteins in cross-beta conformations[30]. Additionally, infrared spectroscopy shows cytokeratin transitioning from a predominantly helical conformation to predominantly beta-sheet when incorporated into MDBs[31]. Increased expression of keratins, as well as hyperphosphorylation, has been shown to occur during cellular stress, and the KRT8 mutations lead to visible KRT8 aggregation in cell culture models[28,32,33]. It is during these periods of increased cellular concentration that factors like ethanol may further tip KRT8 to the point of aggregation.

Just as the aggregation of other amyloid proteins can be seeded by preformed fibrils, we observe that KRT8 aggregation can be seeded by ASH liver tissue extracts. MDBs present in the livers of patients with ASH will reduce in size and quantity with cessation of alcohol consumption. However, if the patient returns to alcohol consumption after this, the MDBs return more rapidly and extensively than before[34]. Latent KRT8 aggregates may remain present during the alcohol cessation period and quickly seed the formation of subsequent aggregates, contributing to what was originally described as the "toxic memory" response of MDBs[29]. Given the limited therapeutic options available for the treatment of ASH and alcoholic cirrhosis, targeting KRT8 aggregation could present a new route for drug development, as has been done with other amyloid diseases. However, similar to other amyloid diseases, the causative versus correlative nature of KRT8 aggregates in relationship to liver disease progression requires further elucidation--whether these aggregates act as bystanders or directly contribute to pathogenesis merits additional study[35].

By analyzing the conversion of LARKS into steric zippers, this work provides insight into the molecular mechanisms driving irreversible amyloid aggregation. LARKS resemble amyloid fibrils in structure, existing as mated beta-sheets stacked upon one another[6]. However, kinks present in the protein backbone of the LARKS disrupt the interaction between the mated sheets, reducing the formation of steric zippers, the tight interdigitation of side chains believed to stabilize amyloid structures. This disruption may lead to the observed reversibility of LARKS. These motifs are found in many proteins known to form LLPS, including TDP-43[36] and FUS[10]. To better understand the cluster of mutations we identified in our computational screen, we compared the relative frequencies at which the wild-type and mutant residues are found beta-sheets throughout known protein structures (Supplementary Figure 3.8). We do not see a strong trend of mutations shifting residues to those more likely to be found in beta-sheets. In contrast, we also analyzed the frequencies at which the wild-type and mutant residues are found in known LARKS structures (Supplementary Figure 3.9). The mutant residues are much less likely to be found in LARKS compared to the wild type. It appears that disruption of the LARKS conformation, notably at the kinked glycine residues, could be an important factor in driving aggregation as opposed to simply adopting a more beta-sheet prone sequence.

Interestingly, instead of highly kinked beta-sheets for the wild type LARKS observed in FUS/TDP-43/hnRNPA1/hnRNPA2, the glycines of the wild type KRT8 segments assume a highly extended conformation, with both dihedral angles approaching $|180^{o}|$. This extended conformation has also been observed in the structure of a LARKS from the nucleoporin protein Nup54 (data unpublished). Additionally, evidence of these extended conformations of can be found in the full-length fibril structures of FUS and TDP43 (Supplementary Figure 3.10). Both kinked and extended beta-sheet conformations may serve the same function, to disrupt the pleated beta-sheet

interdigitation that permits steric zipper formation. Threading of sequences onto extended-beta sheet structures, as was done to screen for LARKS, could be a useful future step to identify similar motifs important for regulating amyloid structure.

MATERIALS AND METHODS

Computational identification of LARKS mutants

LARKS present in the human proteome were identified as previously described[6]. Information about disease related mutations was based on records from ClinVar[17], OMIM[15] and UniprotKB[16]. The records were retrieved in April 2018 with Python scripts utilizing NCBI E-utilities or custom APIs provided by OMIM and UniprotKB. Single Nucleotide Polymorphism records were retrieved with NCBI E-utilities from Entrez SNP database[37]. UniprotKB mapping tool was used to convert UniprotKB accessions to Entrez Gene identifiers. Wild-type and mutant LARKS sequences were threaded onto steric zipper backbones using ZipperDB[3], energies of each were then compared and plotted.

Crystallization

All peptides used for crystallization were purchased from GenScript and crystals were grown in the following conditions using hanging drop vapor diffusion: SGMGGIT: 0.1 M phosphate/citrate (pH 4.2), 40% v/v ethanol, 5 % w/v PEG 1000; SGMGCIT: 0.2M sodium citrate tribasic dehydrate, 0.1M HEPES (pH7.5), 20% v/v 2-propanol; GGYAGAS: 2.5M sodium chloride, 100mM sodium acetate/acetic acid (pH 4.5), 0.2M lithium sulfate.

X-ray crystallography data collection and processing

All x-ray diffraction data was collected using beamline 24-ID-E of the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL, USA) using a beam wavelength of 0.971 Å and temperature of 100 K. Data was collected using 5º oscillations with a detector distance of 140 mm, using an ADSC Q315 CCD detector. Indexing and integration was performed using CCP4, and molecular replacement was done using Phaser, utilizing a library of poly-alanine beta-strands, as well as a selection of kinked strands, as search models. Manual adjustments were performed with COOT during iterative rounds of processing with Refmac to refine the final atomic models. Data and refine statistics are listed in Supplementary Table 3.2.

Protein expression and purification

Recombinant KRT8(1-90) and FUS(1-214) were purified using a pHis-parallel-mCherry vector, using a previously described method[7]. Briefly, proteins were overexpressing in BL21(DE3) Gold *E. coli* cells. Cultures were grown to an OD_{600}=0.4-0.8, then induced with 0.5M IPTG overnight. Cells were pelleted by centrifugation, and the clarified lysate was purified using gravity Ni-NTA columns.

Recombinant tau microtubule binding domain (k18) and α-synuclein were expressed and purified as previously described[38,39].

In vitro aggregation assay

Frozen aliquots of KRT8(1-90) wild-type and mutants were thawed on ice and diluted to 100 µM in buffer containing 20 mM Tris HCl (pH 7.5), 200 mM NaCl, 0.5 mM EDTA, 50 µM Thioflavin T (ThT), and 5 µM TCEP to a final volume of 200 µL in black Nunc 96-well optical bottom plates (Thermo Scientific). A single PTFE bead (0.125 inch diameter) was added to each well to facilitate

agitation. Plates were incubated in a microplate reader (FLUOstar OMEGA, BMG Labtech) for ~100 hours at 37°C with 700 rpm double orbital shaking. Fluorescent measurements were recorded every 15 minutes using λ_{ex} = 440 nm and λ_{em} = 480 nm.

For reversible aggregation assays, the same sample concentrations were used. Sample was maintained at 37°C in the microplate reader with 700 rpm double orbital shaking, with readings taken every 5 minutes. Once the ThT levels were at a baseline, the plate was transferred to a Torrey Pines orbital plate shaker maintained at 4°C with orbital shaking set at maximum. The plate was manually and briefly transferred back to the microplate reader to take fluorescence measurements every five minutes. After 60 minutes, the plate was then transferred back to the microplate reader and kept there to measure melting, with readings again taken every 5 minutes.

Liver tissue seeded assays were carried out in 1x PBS with 100 μM $KRT8_{1-90}$ and 50 μM ThT. Liver extract was in a final dilution of $1/100^{th}$ from the stock extract. Sample was agitated with a PTFE bead with double orbital shaking at 700 rpm.

Both tau k18 and alpha-synuclein aggregation assays were performed in 1x PBS with 50 μM protein monomer, 50 μM ThT, and agitated with a PTFE bead with double orbital shaking at 700 rpm.

For small molecule inhibition assays, 100 mM DMSO stocks for each compound were diluted with 1x PBS into a 1 mM working stock. Compounds were combined with 50 μM KRT8, 50 μM ThT in 1x PBS and agitated with a PTFE bead with double orbital shaking at 700 rpm.

Transmission electron microscopy

6 μL of aggregated wild-type and mutant $KRT8_{1-90}$ samples (taken from in vitro aggregation experiments) was spotted onto Formvar Carbon film 400 mesh copper grids (Electron Microscopy

Sciences) and incubated for 4 minutes. Grids were stained with 6 µL uranyl acetate solution (2% w/v in water) for 2 minutes. Excess solution was removed by blotting and air dried for 30 minutes. TEM images were acquired with a JOEL 100CX TEM electron microscope at 100 kV.

X-ray fiber powder diffraction

Aggregated samples of KRT8 or FUS were centrifuged at 15,000 rpm for 30 minutes and buffer was exchanged with water twice. Samples were suspended between two siliconated glass capillaries ~1 mm apart, forming a bridge between the two capillaries. Sample was allowed to dry and the capillary with the dried aggregate was mounted on an in-house x-ray diffraction machine and diffracted with x-rays for 5 minutes, with the diffraction pattern collected on a CCD detector.

SDS denaturation experiments

KRT8 and FUS samples were aggregated in 1x PBS, 50 µM ThT for ~5 days in a FLUOstar OMEGA microplate reader (BMG Labtech) at 700 rpm double orbital shaking in a Nunc 96-well optical bottom plate. Samples were then incubated with SDS solution (in water) to final concentrations ranging between 0-2.5 % SDS for 20 minutes. ThT fluorescence of each sample was measured using λ_{ex} = 440 nm and λ_{em} = 480 nm using the microplate reader.

Extraction of patient liver tissue samples

Tissue from histopathologically confirmed liver disease cases were fresh-frozen and extracted without freeze-thaw. Extraction of protein aggregates from patient liver tissues was performed using a modified protocol from McGee and colleagues[40]. 250 mg of frozen tissue was cut and thawed, then homogenized using a manual tissue homogenizer. Sample was diluted 5-fold in

extraction solution (0.25 M sucrose (pH 8.5), 5 mM Tris HCl, and 5 mM EDTA) on ice. Sample was then centrifuged at 4°C at 4000 x g for 20 minutes. Supernatant was discarded and pellet was resuspended in the extraction solution and respun for an additional 20 minutes. Washes were repeated until the layer of floating lipid present after each centrifugation was no longer present. Solution was then passed through a 40 μm cell strainer to remove any intact debris. Filtrate was spun for 15 minutes at 400 x g, and the pellet was washed twice. Pellet was then resuspended in 1 x PBS for use in seeding experiments.

ACKNOWLEDGEMENTS

We thank Duilio Cascio and Michael Collazo of the UCLA Dept. of Energy Institute Macromolecular Crystallization Core Technology Center for assistance in structure determination and the UCLA Translational Pathology Core Laboratory for their assistance in acquiring tissue samples. We acknowledge USPHS National Research Service Award 5T32GM008496 (KAM), the UCLA-Caltech Medical Scientist Training Program (KAM) and HHMI for support. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are 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 beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

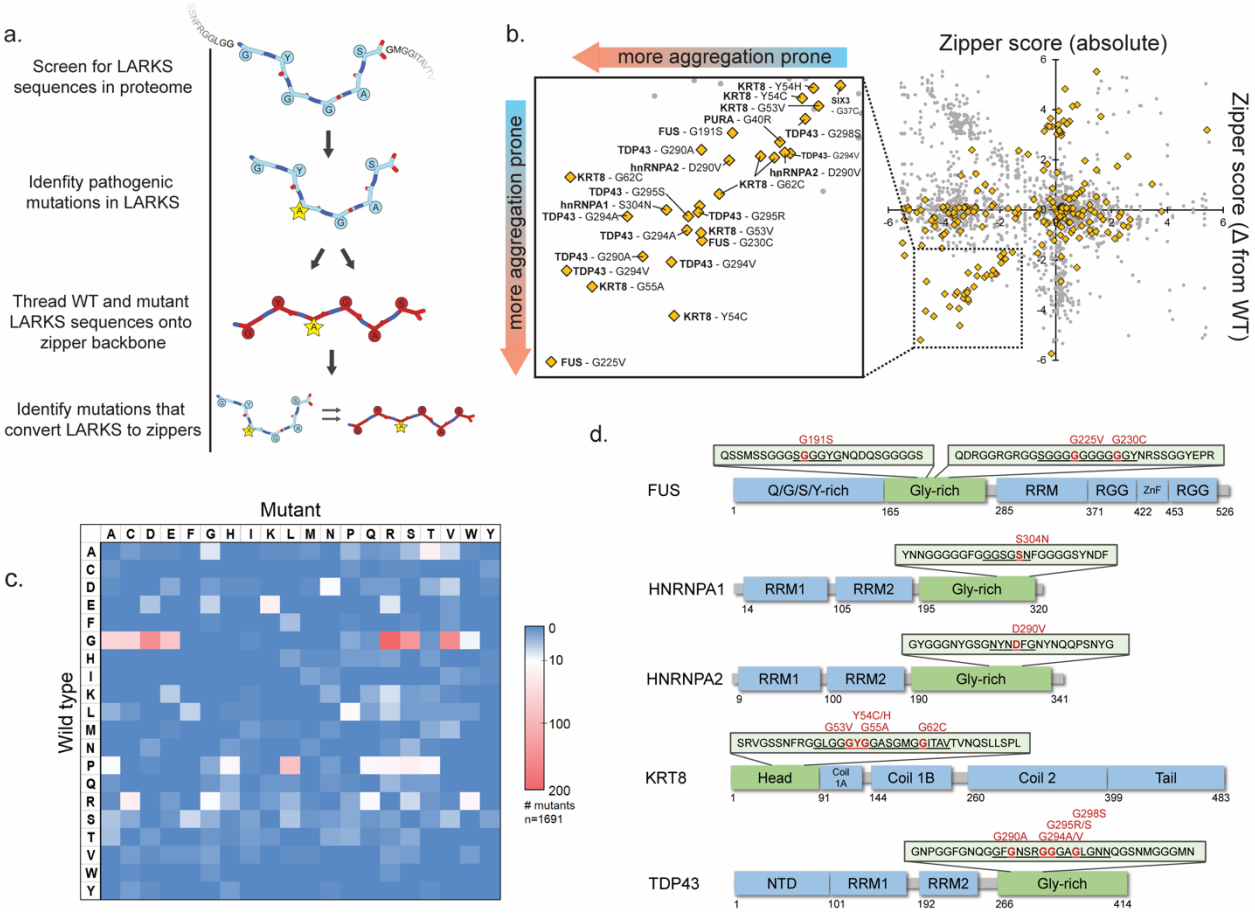


Figure 3.1: Computational search for pathogenic mutations that potentiate amyloid aggregation. **a.** Methodology used to screen for mutations in LARKS that enable the conversion to a steric zipper. (i) The human proteome was screened for sequences capable of adopting LARKS-like structures, a structural motif underlying reversible aggregation. (ii) Mutational databases were then queried to identify disease-linked variants that occur within the identified LARKS regions. (iii) The wild-type and mutant LARKS sequences were then scored for their capacity to form steric zippers, the core structural motif underlying irreversible amyloid aggregation. (iv) Mutations in LARKS sequences that increase their propensity to form steric zippers were identified and further characterized. **b.** Plots of the steric zipper scores for pathogenic mutations in LARKS (yellow) compared to benign SNPs (grey). Insert highlights a population of mutations with increased likelihood of zipper formation. Many mutations known to amyloid

aggregation were identified, including those is FUS, TDP43, hnRNPA1, and hnRNPA2. **c.** Heat map of pathogenic mutations occurring in LARKS (n = 1691). Transition of wild type glycine to polar or hydrophobic residues, and proline to leucine mutations, are the mostly commonly observed variants. **d.** Select proteins with LARKS containing pathogenic mutations predicted to increase zipper formation. Among those identified proteins already characterized to reversibly and irreversibly aggregate (FUS/TDP43/hnRNPA) is the intermediate filament protein Keratin-8 (KRT8).

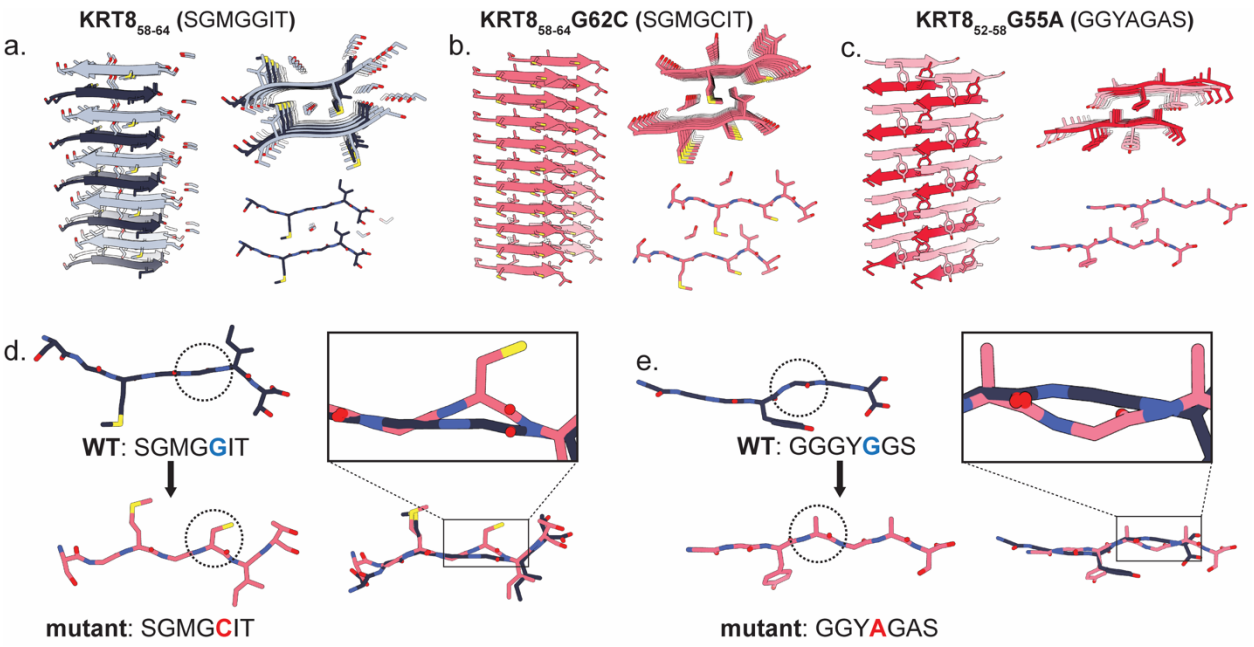


Figure 3.2: Atomic structures of Keratin-8 amyloid segments. Structures of LARKS regions from KRT8 that contain pathogenic mutations were crystalized to assess the affect the mutations have on fibril structure. **a-b.** Fibril structures of the wild-type $KRT8_{58-64}$ segment SGMGGIT (**a**) and the corresponding G62C mutant structure SGMGCIT (**b**). Three views of the fibril structures are shown, from the side (left), looking down the fibril axis (top right) and showing a single fibril layer in cross-section (bottom right). The wildtype fibril contains an anti-parallel beta-sheet with PEG and ethanol molecules bound along the fibril axis. The central glycine residues (Gly61/Gly62) adopt a nearly linear beta-strand conformation. The G62C mutant fibril is a parallel beta-sheet, whose steric zipper interface is driven by hydrophobic interactions between Met60,Cys62, and Ile63. **c.** The $KRT8_{52-58}$ G55A mutant structure GGYAGAS was also determined, revealing an anti-parallel beta-sheet with a steric zipper interface formed by Tyr54, Ala55, and Ala57. **d**. Alignment of the wild-type and mutant $KRT8_{58-64}$ segments shows a transition in backbone conformation from a highly extended to a pleated beta-sheet at the site of the mutation (circled). Alignment of the beta-strands from both structures highlights the G62C

mutation results in the non-pleated Gly62 converting to a pleated conformation as Cys62 (shown in inset). **e**. Comparison of the G55A mutant $KRT8_{52-58}$ GGYAGAS segment with a structure from hnRNPA1 whose sequences closely matches the corresponding KRT8 wild-type (hnRNPA1: GGGYGGS vs. KRT8: GGYGGAS) shows a similar trend in transitioning from an extended beta-sheet in the wild-type to a pleated conformation in the mutant. Similar to G62C, backbone overlay the G55A mutation converts a non-pleated Gly55 into a pleated Ala55 (right).

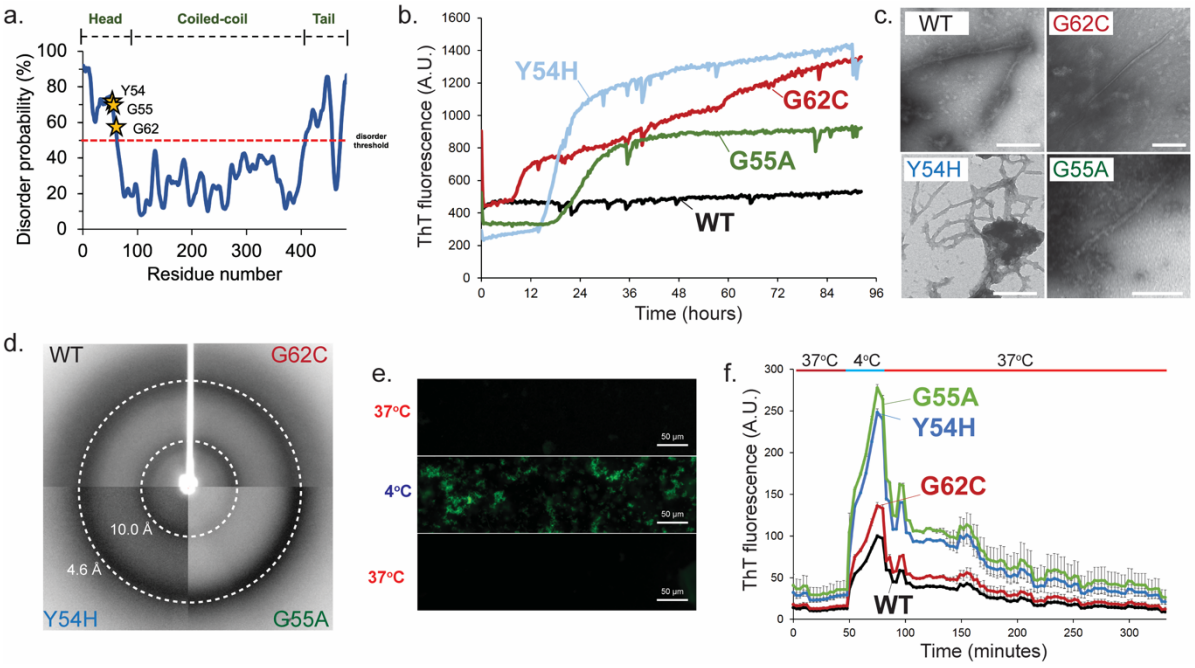


Figure 3.3: The effects of mutations on the *in vitro* aggregation of KRT8. a. Like other aggregation-prone low-complexity domains, the Head domain of KRT8 is predicted to be disordered, and contains three disease-associated mutations, Y54H, G55A, and G62C. PrDOS prediction shows both the Head and Tail KRT8 domains, both of which have low-complexity sequence compositions, are above the threshold for predicted disorder (50%). **b**. Thioflavin T-based aggregation kinetics of KRT8 show that disease-related mutations present in LARKS sequences of the protein (Y54H, G55A, G62C) greatly enhance aggregation compared to wild-type (WT). **c**. Electron micrographs of both wild-type and mutant KRT8 samples show fibrillar aggregates of varying morphology (scale bar: 400 nm). Very few fibrils were observed for wild type while many were seen for the three mutants. **d**. X-ray powder diffraction of KRT8 aggregates produces a cross-beta diffraction pattern characteristic of amyloid fibrils. Diffraction rings are observed at 4.6 Å and 10 Å. **e**. Cooling of KRT8 from 37ºC to 4ºC produces ThT-positive aggregates, which dissociate upon warming back to 37ºC. **f**. Aggregation by cooling occurs more extensively in the mutant forms of KRT8 compared to wild type, as assessed by ThT fluorescence.

Samples were initially maintained at 37ºC, cooled to 4ºC, then melted back to 37ºC. All aggregates melt over the course of several hours after warming back to 37ºC.

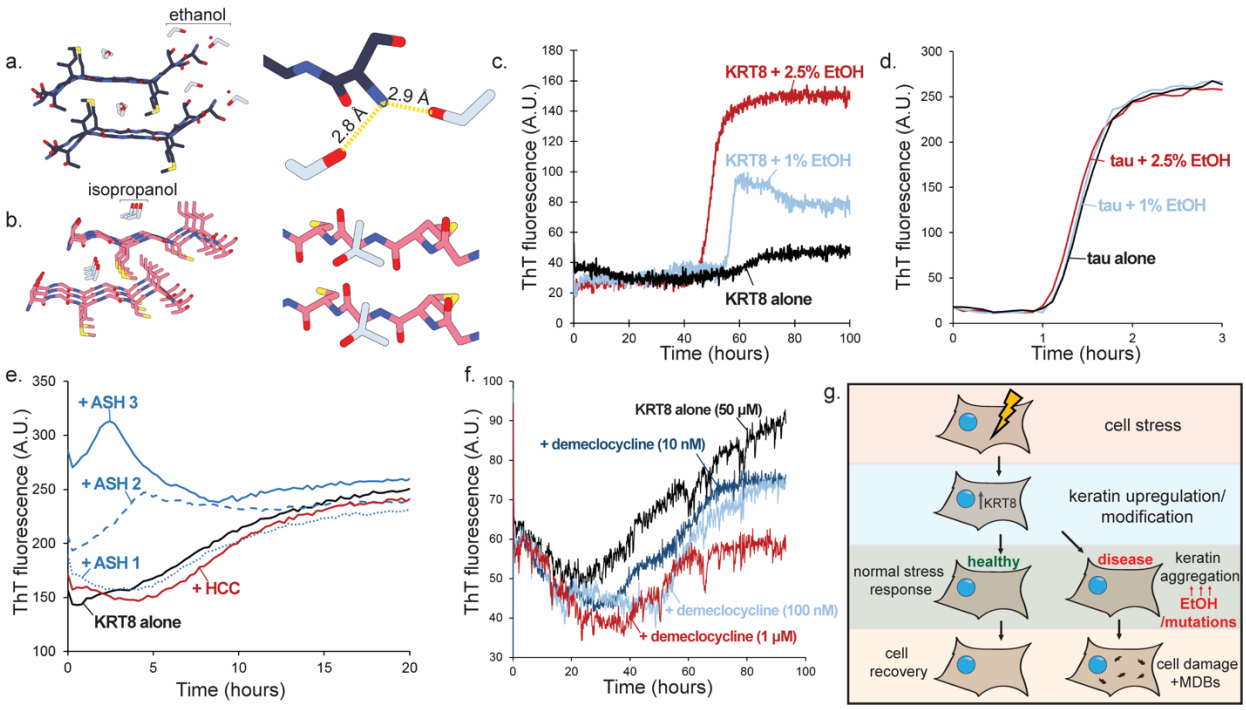
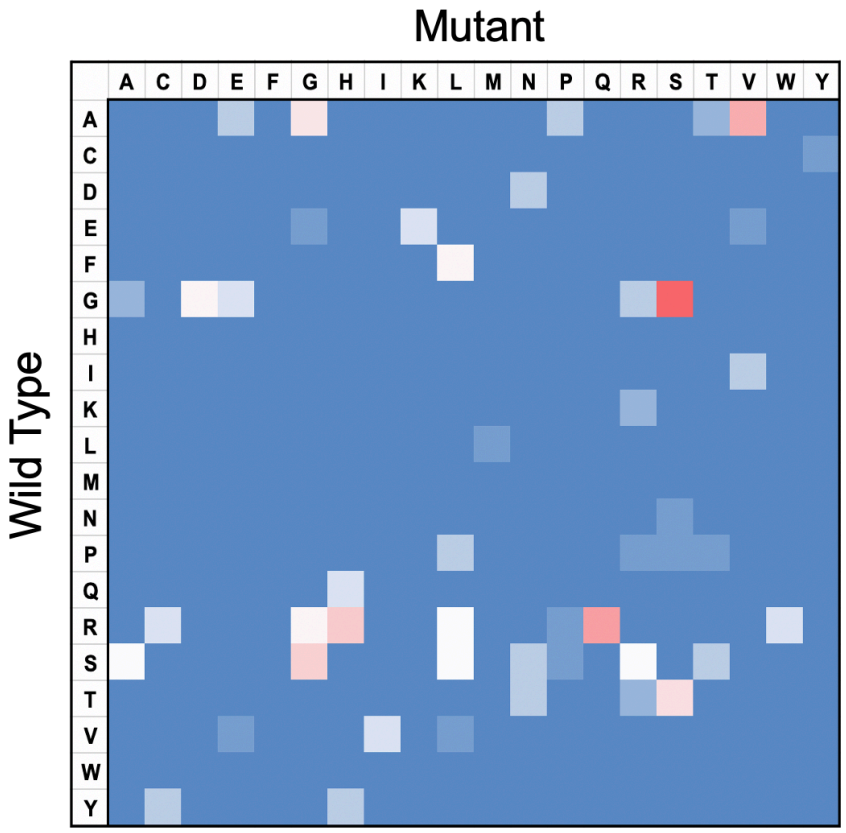


Figure 3.4: **Both ethanol and alcoholic steatohepatitis liver tissue extracts promote the aggregation of KRT8. a-b**. Structures of wild-type and mutant $KRT8_{58-64}$ segments co-crystallized with alcohol. The wild-type SGMGGIT structure (**a**) shows two ethanol molecules (gray) per asymmetric unit coordinated to the N-terminus of the peptide, while mutant SGMGCIT (**b**) displays an isopropanol binding alongside the middle of the peptide along the fibril axis, occupying the space between the beta-sheets (in gray) in the steric zipper. Both structures illustrate that alcohols can directly associate with KRT8 amyloid fibrils. Aggregated KRT8 is the primary component of Mallory Bodies, a common pathological feature observed in hepatocytes during alcoholic liver disease. **c**. ThT aggregation kinetics assay of wild-type KRT8 with increasing concentrations of ethanol (0, 1, and 2.5%). Higher concentrations of ethanol were tested but led to rapid protein precipitation from solution (data not shown). **d**. Aggregation of the tau microtubule binding domain (k18) in the presence of ethanol. **e**. Extracts of liver tissue with alcoholic steatohepatitis (ASH) or hepatocellular carcinoma (HCC) were used to seed the aggregation of

KRT8. **f**. The effects of demeclocycline HCl on KRT8 aggregation. **g**. In a state of cellular stress, increased expression of KRT8 may contribute to a normal stress response. However, in the context of disease, particularly in the presence of ethanol and/or mutations which may occur in liver disease, KRT8 may become aggregated. This may result in the formation of MDBs or an insufficient stress response.

protein	mutation residue #	WT	mutant	Predicted LARKS	Predicted zipper
FUS	191	G	S	SGGGYG	SSGGYG
	225	G	V	SGGGGG	SGGGVG
	230	G	C	GGGGGY	GGGCGY
GATAD1	59	G	A	GFGAAT	AFGAAT
hnRNPA1	304	S	N	GGSGSN	GGSGNN
hnRNPA2	290	D	V	SGNYND	SGNYNV
	290	D	V	NYNDFG	NYNVFG
KRT8	53	G	V	GLGGGY	GLGGVY
	53	G	V	LGGGYG	LGGVYG
	54	Y	C	YGGASG	CGGASG
	54	Y	C	GYGGAS	GCGGAS
	54	Y	H	GYGGAS	GHGGAS
	55	G	A	GYGGAS	GYAGAS
	62	G	C	MGGITA	MGCITA
	62	G	C	SGMGGI	SGMGCI
	62	G	C	GGITAV	GCITAV
PURA	40	G	R	GGGGSG	GRGGSG
TDP43	290	G	A	GFGNSR	GFANSR
	290	G	A	GNSRGG	ANSRGG
	294	G	A	GNSRGG	GNSRAG
	294	G	A	SRGGGA	SRAGGA
	294	G	V	NSRGGG	NSRVGG
	294	G	V	RGGGAG	RVGGAG
	294	G	V	GNSRGG	GNSRVG
	295	G	R	GGGAGL	GRGAGL
	295	G	S	GGGAGL	GSGAGL
	298	G	S	AGLGNN	ASLGNN

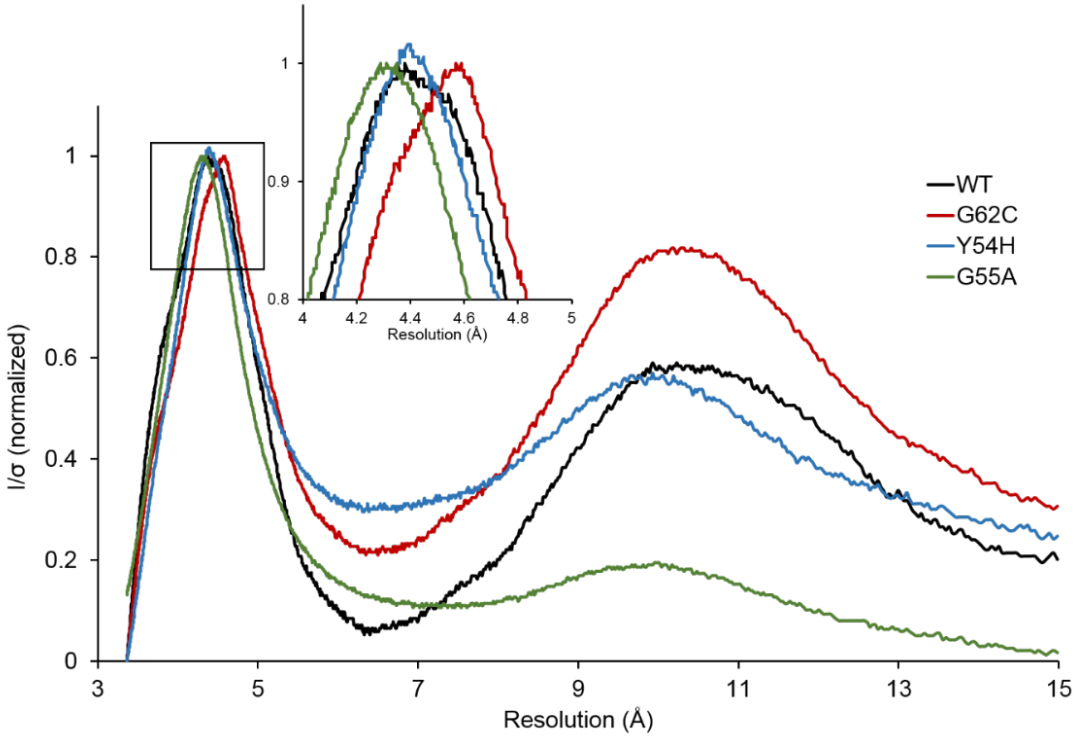
Supplementary Table 3.1: List of pathogenic mutations in LARKS identified to increase the likelihood of steric zipper formation. Proteins like FUS, hnRNPA1, hnRNPA2, and TDP43 are known to form reversible phase separations as well as irreversible amyloid aggregates. Numerous mutations occurring in the head domain of KRT8 were also identified, primarily associated with liver disease. Mutations in GATAD1 (GATA zinc finger domain containing 1) and PURA (transcriptional activator of Pur-alpha) were also identified by the *in silico* screen.



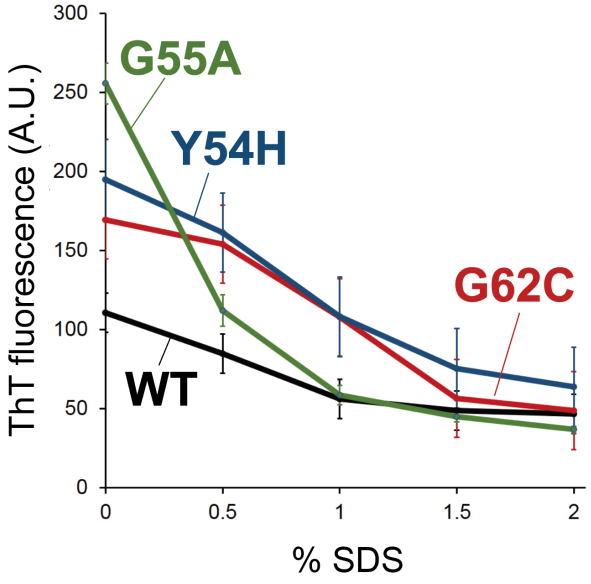
Supplementary Figure 3.1: Heatmap for benign SNPs found in LARKS. Wild type residues on shown on the y-axis, and mutant amino acids on the x-axis. Blue indicates low frequency; red indicates high frequency. The most commonly observed polymorphisms are G→S, A→V, R→Q, R→H, and S→G. The benign nature of these variants may be due to the non-disruptive nature of the amino acid substitution. For example, R→Q, R→H, and A→V are all substitutions for amino acids of similar properties (i.e. polar to polar, hydrophobic to hydrophobic). Glycine and serine residues are highly enriched in LARKS and low-complexity domains. The G→S and S→G variants may indicate a degree of interchangeability between glycine and serine residues is tolerated for function to be maintained.

Crystal Peptide	^{58}SGMGGIT^{64}	^{58}SGMGCIT^{64}	^{52}GGYAGAS^{58}
Data Collection			
Beamline	APS 24-ID-E	APS 24-ID-E	APS 24-ID-E
Spacegroup	$P2_1$	$P2_1$	P1
Resolution (Å)	1.1	1.7	1.1
Unit cell dimensions: a,b,c (Å)	8.36, 51.61, 9.53	4.75, 46.17, 10.33	9.43, 10.49, 16.63
Unit cell angles: α,β,γ (°)	90.0,109.1,90.0	90.0, 103.3, 90.0	88.9, 76.3, 74.1
Measured reflections	7713	2087	3928
Unique reflections	2926	472	1830
Overall completeness (%)	93.3	93.5	72.0
Last shell completeness (%)	62.7	71.4	50.7
Overall redundancy	2.6	4.4	1.5
Last shell redundancy	2.0	2.6	1.5
Overall Rsym	0.182	0.192	0.059
Last shell Rsym	0.351	0.947	0.109
Overall I/Sigma	3.3	13.0	7.43
Last shell I/Sigma	1.4	2.1	3.93
Last Shell (Å)	1.13-1.10	1.76-1.70	1.28-1.10
Refinement			
R_{work}	0.157	0.291	0.280
R_{free}	0.169	0.269	0.306
RMSD bond length (Å)	0.010	0.022	0.001
RMSD angle (°)	1.5	2.0	0.6
Number of peptide atoms	98	43	41
Number of solvent atoms	19	4	0
Average B factor of peptide (Å^2)	9.1	30.1	7.1
Average B factor of solvent (Å^2)	18.0	34.1	--

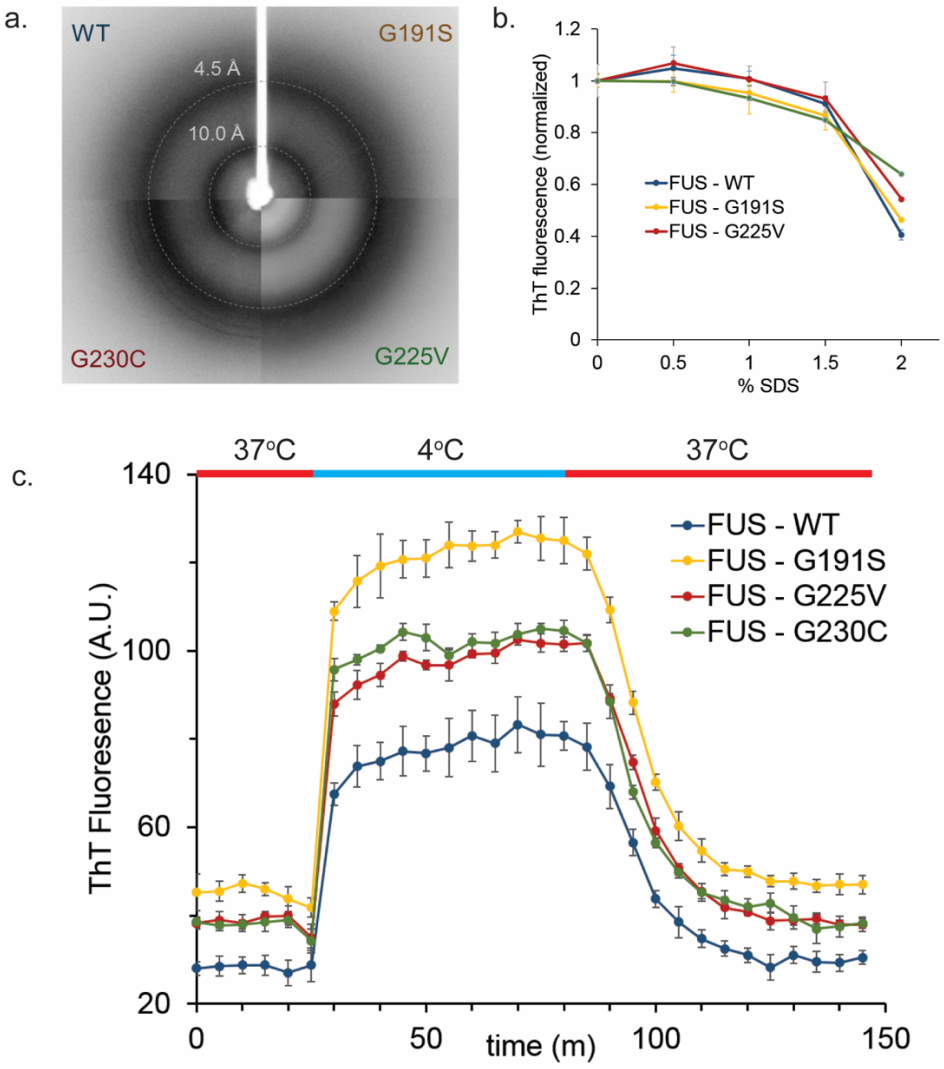
Supplementary Table 3.2: Table of KRT8 segment microcrystal data collection and refinement



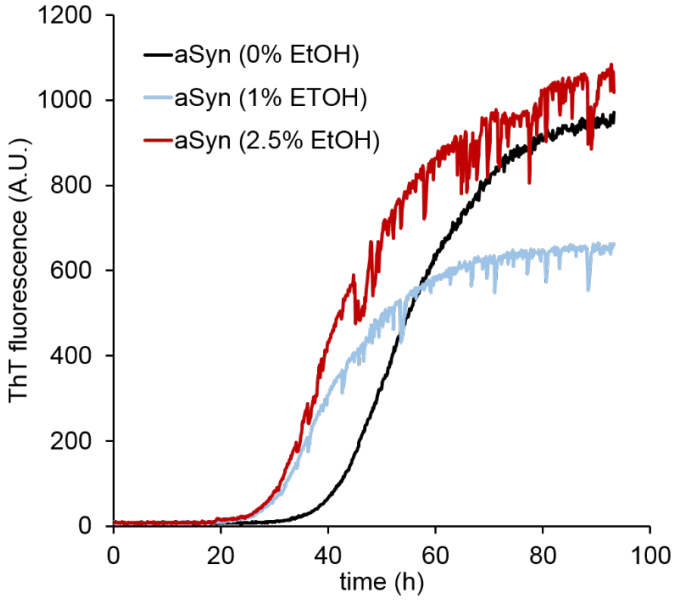
Supplementary Figure 3.2: Radial profiles of x-ray powder diffraction of KRT8 aggregates. A slight difference between the wild-type and mutants forms of KRT8 aggregates can be observed between the rings ranging 4.3-4.6 Å. The G55A mutant also produces a much weaker ring near 10 Å compared to the others.



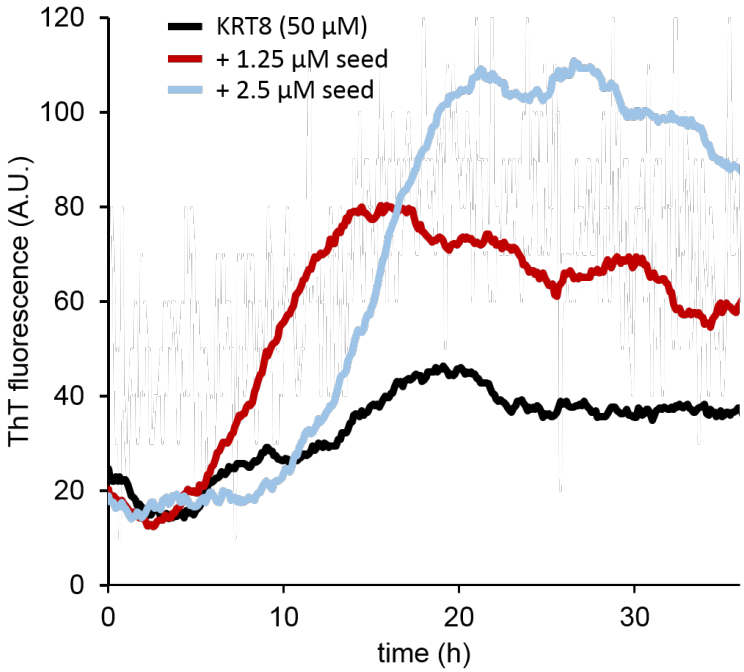
Supplementary Figure 3.3: Stability of the wild-type and mutant KRT8 aggregates to SDS denaturation. At 0% SDS, wild type KRT8 forms less fibrils overall compared to all three mutants, as measured by ThT fluorescence.



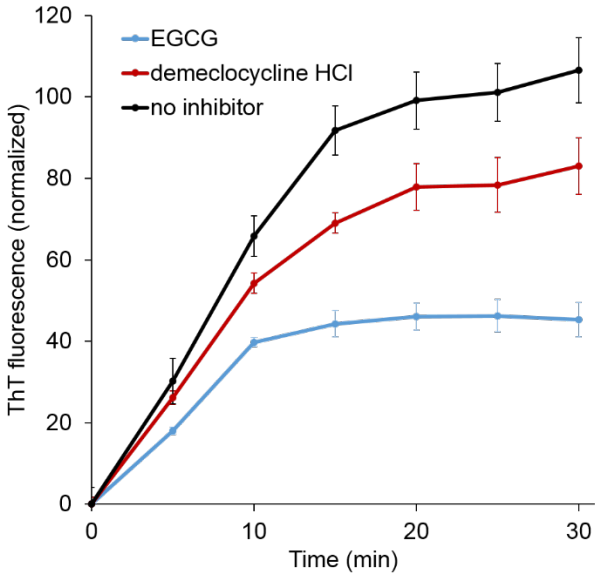
Supplementary Figure 3.4: The effects of pathogenic LARKS mutations on FUS aggregation. **a**. Aggregates of both wild-type and mutant (G191S, G230C, and G225V) FUS all produce cross-beta diffraction indicative of amyloid formation. **b**. FUS aggregates display similar stabilities to SDS denaturation, with the wild-type being slightly less stable at 2% SDS compared to the mutants. **c**. Like KRT8, FUS forms ThT positive phase separations when cooled to 4°.



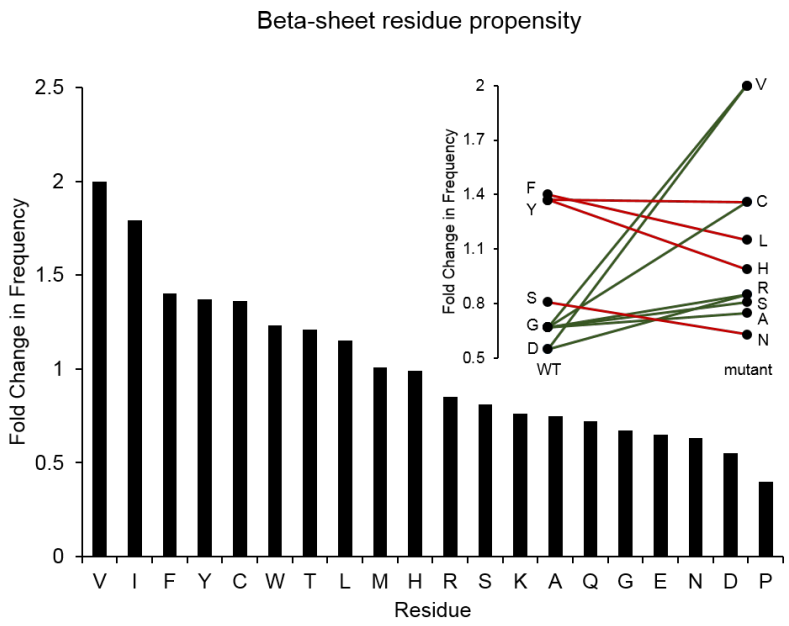
Supplementary Figure 3.5 The effects of ethanol on alpha-synuclein aggregation. Both 1% and 2.5% ethanol appear to reduce the lag-time required for alpha-synuclein (50 μM) to aggregate.



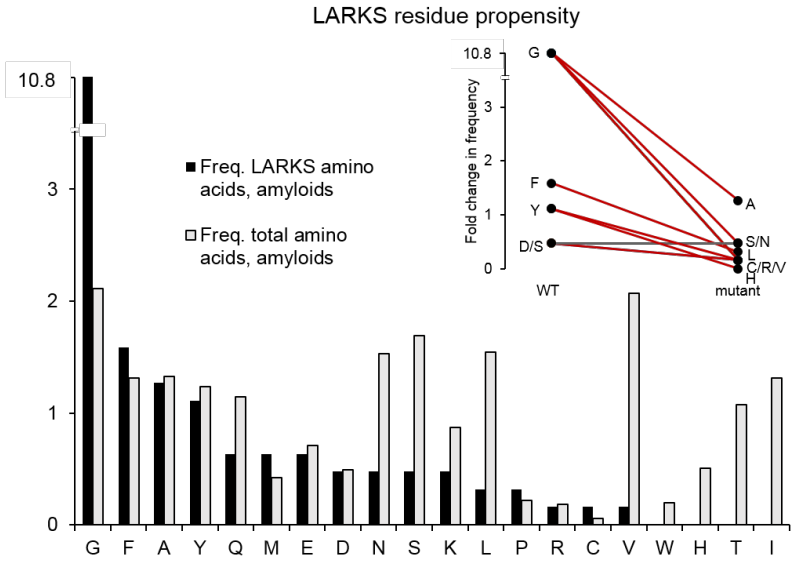
Supplementary Figure 3.6: Seeded aggregation of KRT8 by preformed recombinant KRT8 aggregates. The addition of pre-aggregated KRT8 enhances the aggregation of monomeric protein (50 µM), presumably through template seeding.



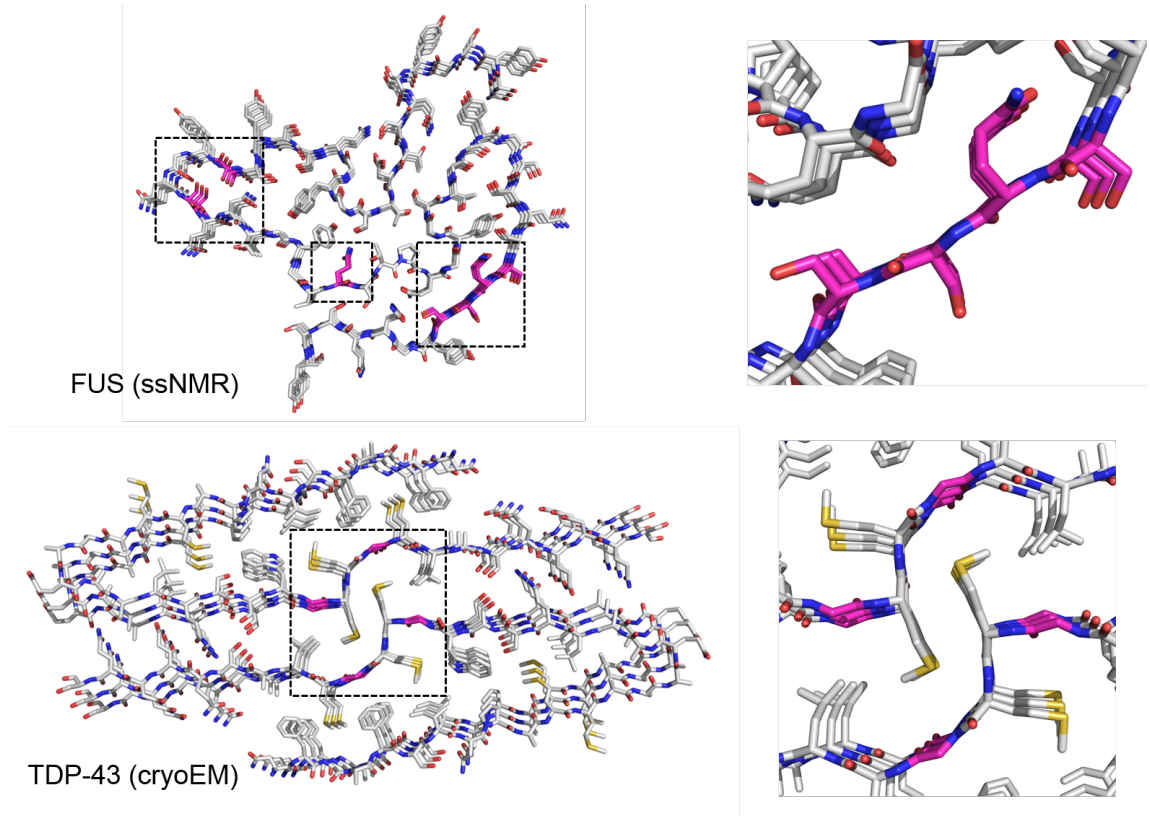
Supplementary Figure 3.7: The effects of small molecules on KRT8 phase separation. KRT8 samples were maintained at 37ºC at time point 0, then cooled to 4ºC, leading to ThT phase separations. At 50 µM KRT8 concentration, addition of demeclocycline HCl (10 µM) or the known anti-amyloid compound EGCG (10 µM) reduces ThT signal.



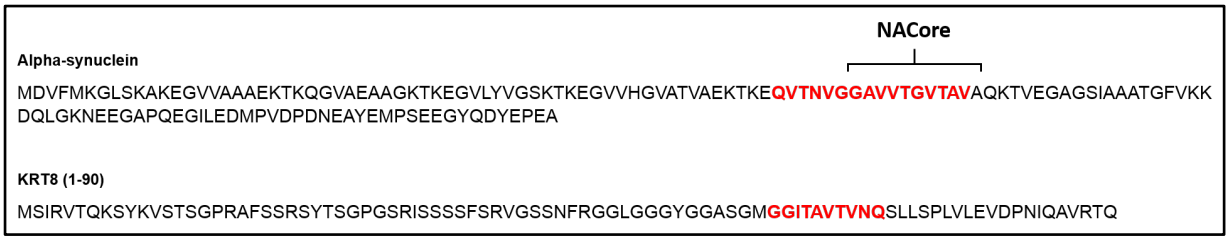
Supplementary Figure 3.8: Residue frequencies in beta-sheets. The relative frequencies in which each amino acid is observed in beta-sheets (bar graph). Analysis of the disease-related mutations predicted to convert LARKS to zippers shows little correlation of residues mutating to more beta-sheet prone amino acids compared to wild-type (insert).



Supplementary Figure 3.9: Residue frequencies in LARKS. Comparison of relative amino acid frequencies between all experimentally determined structures of LARKS versus amyloid fibrils (bar graph). There is a drastically higher enrichment of glycines in LARKS compared to canonical amyloid structures, likely contributing to their kinked structures. Pathogenic mutations predicted to convert LARKS to steric zippers lead to a transition from residues enriched in LARKS to those that are not.



Supplementary Figure 3.10: Atomic structures of FUS and TDP-43 contain extended beta-sheets. Both structures have regions of highly extended beta-sheets (magenta).



Supplementary Figure 3.11: Sequence comparison of KRT8$_{1\text{-}90}$ with alpha-synuclein. The head domain of KRT8 has an enrichment of glycine and serine residues. Additionally, it has a hydrophobic region (highlighted in red) that resembles the amyloidogenic core of alpha-synuclein, known as the NACore.

REFERENCES

1. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
2. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
3. 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).
4. Aguzzi, A. & Altmeyer, M. Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends Cell Biol.* **26**, 547–558 (2016).
5. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles | Journal of Cell Biology | Rockefeller University Press. https://rupress.org/jcb/article/210/4/529/38290/Prionlike-domains-in-RNA-binding-proteins-are.
6. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698 (2018).
7. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).
8. Ryan, V. H. *et al.* Mechanistic view of hnRNPA2 low complexity domain structure, interactions, and phase separation altered by disease mutation and arginine methylation. *Mol. Cell* **69**, 465-479.e7 (2018).

9. Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A. & Patel, B. K. Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis. *Front. Mol. Neurosci.* **12**, (2019).

10. 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).

11. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

12. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).

13. Nomura, T. *et al.* Intranuclear Aggregation of Mutant FUS/TLS as a Molecular Pathomechanism of Amyotrophic Lateral Sclerosis. *J. Biol. Chem.* **289**, 1192–1202 (2014).

14. Sun, C.-S. *et al.* The Influence of Pathological Mutations and Proline Substitutions in TDP-43 Glycine-Rich Peptides on Its Amyloid Properties and Cellular Toxicity. *PLOS ONE* **9**, e103644 (2014).

15. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), 2018.

16. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).

17. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

18. Murray, D. T. *et al.* Structural characterization of the D290V mutation site in hnRNPA2 low-complexity-domain polymers. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9782–E9791 (2018).

19. Bragulla, H. H. & Homberger, D. G. Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *J. Anat.* **214**, 516–559 (2009).

20. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699.e16 (2018).

21. Ku, N.-O., Gish, R., Wright, T. L. & Omary, M. B. Keratin 8 Mutations in Patients with Cryptogenic Liver Disease. *N. Engl. J. Med.* **344**, 1580–1587 (2001).

22. Gui, X. *et al.* Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* **10**, 1–12 (2019).

23. Lin, Y. *et al.* Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **167**, 789-802.e12 (2016).

24. Jensen, K. & Gluud, C. The Mallory body: morphological, clinical and experimental studies (Part 1 of a literature survey). *Hepatol. Baltim. Md* **20**, 1061–1077 (1994).

25. Morozova, O. A., March, Z. M., Robinson, A. S. & Colby, D. W. Conformational features of tau fibrils from Alzheimer's disease brain are faithfully propagated by unmodified recombinant protein. *Biochemistry* **52**, 6960–6967 (2013).

26. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**, 163-178.e19 (2017).

27. Strnad, P., Zatloukal, K., Stumptner, C., Kulaksiz, H. & Denk, H. Mallory–Denk-bodies: Lessons from keratin-containing hepatic inclusion bodies. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1782**, 764–774 (2008).

28. Ku, N.-O., Gish, R., Wright, T. L. & Omary, M. B. Keratin 8 Mutations in Patients with Cryptogenic Liver Disease. *N. Engl. J. Med.* **344**, 1580–1587 (2001).

29. Denk, H., Stumptner, C. & Zatloukal, K. Mallory bodies revisited. *J. Hepatol.* **32**, 689–702 (2000).

30. Mahajan, V. *et al.* Cross β-Sheet Conformation of Keratin 8 Is a Specific Feature of Mallory–Denk Bodies Compared With Other Hepatocyte Inclusions. *Gastroenterology* **141**, 1080-1090.e7 (2011).

31. Kachi, K., Wong, P. T. T. & French, S. W. Molecular Structural Changes in Mallory Body Proteins in Human and Mouse Livers: An Infrared Spectroscopy Study. *Exp. Mol. Pathol.* **59**, 197–210 (1993).

32. Asghar, M. N. *et al.* The Amount of Keratins Matters for Stress Protection of the Colonic Epithelium. *PLOS ONE* **10**, e0127436 (2015).

33. Owens, D. W. *et al.* Human keratin 8 mutations that disturb filament assembly observed in inflammatory bowel disease patients. *J. Cell Sci.* **117**, 1989–1999 (2004).

34. Klassen, L. W., Tuma, D. & Sorrell, M. F. Immune mechanisms of alcohol-induced liver disease. *Hepatology* **22**, 355–357 (1995).

35. Zatloukal, K. *et al.* From Mallory to Mallory–Denk bodies: What, how and why? *Exp. Cell Res.* **313**, 2033–2049 (2007).

36. 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).

37. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **47**, D23–D28 (2019).

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

39. Sangwan, S. *et al.* Inhibition of synucleinopathic seeding by rationally designed inhibitors. *eLife* **9**, e46775 (2020).

40. Morton, J. A., Fleming, K. A., Trowell, J. M. & McGee, J. O. Mallory bodies--immunohistochemical detection by antisera to unique non-prekeratin components. *Gut* **21**, 727–733 (1980).

CHAPTER 4

Extended beta-sheets contribute to reversible amyloid formation

INTRODUCTION

The conversion of proteins into the amyloid state is a hallmark of numerous diseases, wherein proteins that are normally soluble, many of which are intrinsically disordered, aggregate into long unbranched, insoluble fibrils[1]. Such fibrils are observed in neurodegenerative diseases like Alzheimer's and Parkinson's disease, prion diseases, systemic amyloid diseases like AL and AA amyloidosis, and even metabolic disorders like Type II diabetes[2]. Pathogenic amyloid fibrils are exceptionally stable and form irreversibly within tissues. This high degree of stability is explained by the unique atomic structure of the fibrils[3]. Amyloid structure is characterized by repeating layers of protein monomers stacked upon each other as predominantly beta-sheets. The vertical stacking is stabilized by backbone hydrogen bonding between strand layers and certain side chain interactions, including pi-pi interactions of stacked aromatic residues or the formation of polar zippers caused by hydrogen-bonding (H-bonding) of polar side chains[3,4]. Opposing beta-sheets of the fibrils will often interact through the interdigitation of hydrophobic side chains, forming a steric zipper[5]. Combined, the molecular forces stabilizing the amyloid architecture produce aggregates that are resistant to cellular degradation, enabling their contribution to disease pathology.

Despite their numerous associations with disease states, functional amyloid assemblies have also been observed[6–10]. Several RNA binding proteins form amyloid-like fibrils via their low-complexity domains (LCDs), such as FUS (fused in sarcoma), TDP43 (TAR DNA-binding protein

43), and hnRNPA1/hnRNPA2 (heterogenous nuclear ribonucleoproteins), which contribute to hydrogel formation and are speculated to play a role liquid-liquid phase separations[11–14]. While pathogenic amyloids are typically irreversible once formed, these LCD-containing proteins (i.e. FUS, TDP43, hnRNPA) form temperature-labile amyloid fibrils, capable of associating and dissociating dynamically. It has been demonstrated that the temperature-labile assembly of LCD segments into hydrogels is due to the formation of amyloid structures with kinked beta-sheet backbones, as opposed to the normally pleated beta-sheets found in irreversible amyloid structures[15]. Termed LARKS (low-complexity, aromatic-rich, kinked segments), these kinked amyloid motifs interact weakly through polar atoms and aromatic side chains, as opposed to the tight interdigitation and hydrophobic packing observed in typical steric zippers. Similar kinked amyloid motifs have also been observed to regulate the reversible fibrillization of hnRNPA1[16]. Despite these early findings, many questions remain regarding the molecular interactions regulating reversible amyloid formation.

In this work we aim to further explore the structural motifs underlying functional amyloid proteins. Recent advances in structural biology, namely those in micro-focus X-ray diffraction, cryo-electron microscopy (cryoEM) and solid-state nuclear magnetic resonance (ssNMR) have enabled the determination of numerous amyloid fibril structures from both recombinant and tissue-derived sources[17–19]. Analysis of these structures highlights important distinctions between proteins which undergo reversible and irreversible assembly. Chiefly among these findings is the prevalence of highly extended beta-sheets in the reversible structures, in which the peptide backbone adopts a nearly linear conformation, with both dihedral angles approaching $|180^{o}|$. Like the kinks observed with LARKS, these extended motifs result in beta-sheet architecture that deviates from the ideal pleated conformation. Computational analysis of these extended motifs

highlights their potential role in fibril disassembly and identifies specific intra-residue interactions which might stabilize their conformation.

RESULTS

Proteins backbone conformations in amyloid fibril structures

Early inquiries into the structure of amyloid protein fibrils using X-rays revealed cross-beta diffraction patterns which are now one of their defining biophysical characteristics[20]. Subsequent structural studies of amyloid proteins confirmed that this pattern is a result of beta-strands stacking into vertical sheets along the fibril axis through backbone hydrogen bonding[2]. These sheets then interact with other sheets primarily through sidechain packing to form steric zippers, a highly stable motif in which side chains interdigitate like the teeth of a zipper down the length of the fibril axis[19]. Since those studies, numerous structures of amyloid proteins have been solved, from both synthetic and tissue-derived sources, all of which follow these same structural principles. However, beyond this generalized amyloid architecture, heterogeneity between amyloid folds is becoming increasingly apparent as more structures are determined.

While all amyloid fibrils are primarily composed of beta-sheets, the extent and precise conformation of those beta-sheets within each structure varies. The beta-sheet conformations of amyloid fibrils can be roughly grouped into three categories based on the dihedral angles of their peptide backbones: kinked, pleated, and extended (Figure 4.1). Typical beta-sheets are pleated, in which the peptide backbone zigzags to accommodate the tetrahedral bond orientation of the Cα atoms. The majority of residues in amyloid fibrils are pleated beta-sheets. However, amyloid fibril structures, particularly those derived from full-length proteins, are not exclusively pleated. Kinked backbone beta-sheets permit reverse turns in the protein chain to allow compact assembly of

overall fibril fold. Extensive kinking is observed in LARKS motifs interspersed throughout amyloid-prone low-complexity regions[15]. Backbone kinks are most frequently observed at glycine residues due to conformational flexibility allowed by their lack of sidechain. Lastly, the beta-sheet can adopt extended conformations in which the pleating becomes stretched to form linear or near-linear peptide angles. Here both phi and psi dihedral angles approach |180°|. Many examples of all three of these beta-sheet categories (kinked/pleated/extended) are observed in known amyloid fibril structures (Figure 4.1b).

Reversible amyloid structures are enriched with extended beta-strands

To better understand the roles each of these three beta-sheet categories play in amyloid proteins, we analyzed the structures of ten different full-length amyloid fibrils determined by either ssNMR or cryoEM (Figure 4.2). Structures include the following: amyloid-beta(1-42) (Aβ) found in Alzheimer's disease[21], islet amyloid polypeptide (IAPP) observed in pancreatic islet cells during Type II diabetes[22], three structures of tau from patient brain tissue with Alzheimer's disease[23,24], chronic traumatic encephalopathy (CTE)[25] or Pick's disease[26], α-synuclein from multiple system atrophy (MSA) patient brain tissue[27], β2-microglobulin from dialysis-associated amyloidosis tissue[28], transthyretin (TTR) from ATTR cardiac amyloidosis tissue[29], and λ6 light chain from AL amyloidosis tissue[30]. Additionally, structures from proteins known to undergo reversible fibrillization were analyzed, FUS[14] and two unique TDP43 structures (dagger and R-fold)[31]. Examples of pleated, kinked, and extended beta-sheets are observed, as defined and color-coded by their dihedral angles (Figure 4.2a). Quantitative analysis reveals that the majority of residues in all structures are in pleated-beta sheet conformations, composing 79% of ordered residues on average (Figure 4.2b). However, delineation between the reversible and irreversible amyloid proteins are observed when analyzing the kinked and extended strand composition. The FUS and

TDP43 R-fold have the highest percentage of kinked residues, 31% and 32%, respectively, compared to the average 19%. Comparison of extended beta-sheet prevalence reveals FUS and both TDP43 structures have a significantly higher enrichment: 7%, 12%, and 5% compared to the average 2%. Extended strand motifs are absent in the majority of the irreversible amyloid structures analyzed.

We next sought to measure how these differences in backbone dihedral angles translate to inter-residue distances within each fold. To accomplish this, we measured the atomic distances between all possible Cα *i* and *i*+4 pairs in a host of amyloid structures (Figure 4.3). Structures analyzed include several pleated steric zipper segments ("zippers"), several LARKS structures ("LARKS"), and multiple full-length amyloid structures. Additionally, fibril structures from Keratin-8 ("KRT8") and Nucleoporin 54 ("Nup54") were measured, as both intermediate filament and FG-repeat nucleoporin proteins have demonstrated reversible self-assembly through liquid-liquid phase separation of their LCDs and contain predicted LARKS[15,32,33] (Supplementary Table 4.3). For comparison, the globular structure from proteinase K as well as a one of its helical segments ("globular" and "helix") (Figure 4.3a)[34]. The atomic distances were classified into four distinct categories: helical, kinked, pleated, and extended. Each classification was created based on the internal control structures analyzed—i.e. the "helical" category encompassed all distances found in the proteinase K α-helix, "pleated" was based on the steric zipper segment distances, "kinked" were those measurements between "helix" and "pleated", and "extended" were all distances greater than the steric zippers. Like the Ramachandran angle analysis, virtually all structures contain distances in the "pleated" and "kinked" ranges. Again, a distinction between proteins that reversibly and irreversibly aggregate is seen in the "extended" distances. TDP43, FUS, KRT8, Nup54, all of which reversibly aggregate, have multiple Cα *i*/*i*+4 distances >14 Å,

while only a single measurement in that range is observed in the irreversible structures (tau AD filament). Overall comparison of measurements confirms that distances over 14 Å are almost exclusively observed in the reversible category (Figure 4.3c). Examples of these extended distances can be observed in the amyloid structures of both FUS and TDP43 (Figure 4.3d).

Energetic analysis of pleated and extended beta-sheet fibrils

Having observed the prevalence of extended beta-sheet motifs in the amyloid structures analyzed, we next investigated their role in fibril stability. Two amyloid peptide structures from $KRT8_{58\text{-}64}$ were previously determined for the wildtype ($_{58}SGMGGIT_{64}$) and G62C mutant ($_{58}SGMGCIT_{64}$) sequences. The G62C mutation leads to increased aggregation of KRT8 and is associated with liver disease[35]. The wildtype $_{58}SGMGGIT_{64}$ peptide adopts an extended beta-sheet conformation at the two central glycine residues. The conversion of glycine-62 to cysteine in the mutant shifts the extended backbone from an extended beta-sheet to a pleated beta-sheet (Figure 4.4a). To analyze the effects each backbone has on fibril stability, DFT-based calculations were carried out for each structure. For each calculation, a pair of parallel strands from both structures were stripped of sidechain atoms and then geometry optimized. Total energies of the strands as pairs were calculated then subtracted from the energies of the strands after being separated, yielding an energy of separation for both structures (Figure 4.4b). Comparison of separation energies shows the extended strand separation energy (-15.6 kcal/mol) is much less than the pleated strand (-23.0 kcal/mol). The same calculations were performed for the structures as anti-parallel strand pairs and showed the same trend: -32.2 kcal/mol (extended) vs. -54.9 kcal/mol (pleated) (Supplementary Figure 4.1).

Previous work by Newberry et al. has demonstrated that extended beta-strands are stabilized by intra-residue H-bonding, known as C5 H-bonding[36]. Linearization of the peptide

dihedral angles allows sufficient overlap between the amide proton σ* orbital and the *p*-type orbital of the oxygen lone pair from the adjacent carbonyl (Figure 4.4c), which allows H-bonding to occur. We hypothesized that this type of interaction may also be present in the extended beta-strands found in reversible amyloid structures (Figure 4.4d). To test this, non-covalent interaction (NCI) analysis was performed on the same KRT8 peptide structures used in the separation energy calculations, $_{58}$SGMGGIT$_{64}$ and $_{58}$SGMGCIT$_{64}$. As shown in Figure 4.4e, areas of favorable non-covalent interaction can be visualized as blue/green density. For both structures, extensive interactions are present along the interface between the two paired sheets, due to the extensive H-bonding found between layers in beta-sheet/amyloid structures. Inspection of the two central gly61 and gly62 residues, which are linearized in the extended $_{58}$SGMGGIT$_{64}$ structure reveals interaction density in the areas where C5 H-bonding would occur (O-H distance of 2.22 Å). In contrast, no such density is observed in the $_{58}$SGMGCIT$_{64}$ pleated structure, as the pleat increases the distance between the H-bond donor/acceptor pair and decreases the chance for orbital overlap (O-H distance of 2.87 Å). Interestingly, some density is observed in the pleated structure between the carbonyl oxygen and Cα protons at gly61/gly62, another conjectured intra-residue H-bond. For a control, we also performed the NCI analysis on the C5 H-bonded peptide structures studied by Newberry and colleagues. As expected, C5 H-bond-like interaction density is also observed in these structures (Supplementary Figure 4.2).

DISCUSSION

The correlation of a protein's structure to its function is a foundational concept in biochemistry. Despite the well-ordered nature of amyloid fibrils, the link between their structures and possible functions remains poorly understood. Amyloid fibrils formed by protein LCDs are

capable of undergoing reversible association and dissociation, implying a possible functional role. However, the same fibrils can convert to an irreversible form observed in pathologies of multiple diseases, further confounding our understanding of the structure/function link. In this work we attempt to better understand this link by comparatively analyzing the structures, specifically the backbone conformations, of LCD-derived and other amyloid fibrils. Our previous work describing LARKS motifs in LCDs postulates that kinks in the protein backbone facilitate the labile interactions important for reversible amyloid assembly. LARKS are observed in fibril segments structures, such as those from Nup98 and hnRNPA1 (Figure 4.1), but also the full-length structures from FUS and TDP43 (Figure 4.2).

In addition to kinked backbones, here we highlight extended beta-sheets as a recurrent motif in amyloid protein structure, particularly those similar to LARKS in sequence composition. For example, the fibril structure from Nup54 has a sequence composition similar to known LARKS structures, rich in glycine and aromatic residues, but forms a near-linear beta-sheet, arguably the opposite of a kinked beta-sheet. Our DFT calculations show that these extended beta-sheet backbones form less stable dimers than their pleated counterparts. Previous theoretical studies by Schiener and colleagues demonstrates such extended geometry results in significantly weaker NH-O cross β-strand hydrogen bonds[37]. Thus, by adopting an extended conformation, the interactions between strands in beta-sheets in the reversible amyloid proteins may more readily dissociate. Since these linearized beta-strands are likely inherently less stable than pleated ones, additional forces may be responsible for stabilizing them. We hypothesize that chiefly among these forces is the C5 H-bond, as demonstrated by the NCI analysis. Energetic contribution from C5 H-bonding may help maintain the linearity of the beta-strand during reversible amyloid assembly and disassembly.

As both kinked and extended beta-strands appear to contribute to fibril instability, it appears that amyloid reversibility may be a result of deviations from ideal (i.e. pleated) beta-sheet conformation. Such backbone aberrations likely stem from the sequence of each protein. LARKS are mostly composed of glycines, phenylalanine/tyrosine, and serine. These same residues are also the most frequently observed in extended beta-sheet conformations, according to a global analysis of high-resolution protein structures[36]. Likewise, analysis of all known amyloid structures confirms that glycine/serine/phenylalanine/tyrosine are the among residues most frequently in the extended beta conformation (Supplementary Figure 4.3), with glycine by far the most common. As such, we term these extended beta-sheet motifs EAGLS (extended amyloid-like glycine-rich low-complexity segments). The sequence bias in the low complexity domains of FUS, hnRNPA1/2, TDP43, and other proteins known to reversibly aggregate aligns with the sequence bias of LARKS and EAGLS. This may imply that these two structural motifs represent the evolved function of such LCDs, to partially destabilize fibril assembly. More information regarding the structure of protein LCDs is needed.

The interchangeability between LARKS and EAGLS is also unknown. As shown in Figure 4.2, the three full-length amyloid structures from proteins known to reversibly aggregate, FUS, TDP43 (R-fold) and TDP43 (dagger-fold) all have lower than average pleated residue frequency, and higher than average extended beta-sheet frequency. The FUS and TDP43 R-fold have a very high degree of kinked residues, while the TDP43 dagger-fold does not. However, the TDP43 dagger-fold has by far the highest frequency of extended residues, seemingly exchanging kinked residues for extended ones. Many factors like sequence composition may influence structural differences like this; the TDP43 dagger fold is much less rich in glycines, for example. Additional

structures of low-complexity amyloid fibrils and their correlation to reversibility will hopefully aid in illuminating questions like this.

In this study we have exclusively focused on the differences in backbone conformation between amyloid structures. However, many other forces likely participate in determining fibril structure and stability. Charge interactions, pi-pi stacking, H-bonding, etc. between residue sidechains is an important consideration, as are solvent effects, the roles of potential cofactors and post-translational modifications, and many others[38]. Here we have highlighted a potential role peptide backbone conformation, particularly extended beta-sheets, may play in regulating reversibility. We anticipate knowledge of the structure/function relationships in fibrils to grow as the structural space of amyloid proteins continues to expand.

MATERIALS AND METHODS

Crystallization of Nup54 segment

The FGTGFG peptide segment from Nucleoporin54 was purchased from GenScript and crystallized by hanging drop vapor diffusion in the following conditions: (waiting on MPH for this)

X-ray diffraction data was collected using the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL, USA) at beamline 24-ID-E, using a temperature of 100 K and a beam wavelength of 0.971 Å. Data was collected using 5^{o} oscillations with a detector distance of 140 mm, using an ADSC Q315 CCD detector. Indexing and integration was performed using CCP4, and molecular replacement was done using Phaser, utilizing a library of poly-alanine beta-strands as search models. Manual adjustments were performed with COOT during iterative rounds of processing with Refmac to refine the final atomic models.

Structure analysis

See Supplementary Tables 4.1 and 4.2 for PDB codes of all amyloid structures analyzed.

Computational methods

Starting with structures of KRT8 segments $_{58}$SGMGGIT$_{64}$ and $_{58}$SGMGCIT$_{64}$ obtained via X-ray diffraction, sidechains were removed, and hydrogens were added using GaussView 6. Hydrogen positions were optimized at the B3LYP-D3/6-311++G(d,p) level of theory, constraining heavy atoms to their crystallographic positions. Energy calculations were performed at the same level of theory with the addition of the PCM solvation model for water. NCI plots were generated with nciplot[39] and visualized in opensource PyMOL[40].

ACKNOWLEDGEMENTS

KAM is supported by the UCLA-Caltech Medical Scientist Training Program (GM08042). KAM and DE are supported by the UCLA Chemistry-Biology Interface training grant (USPHS National Research Service Award 5T32GM008496). We thank Duilio Cascio and the UCLA DOE Macromolecular Crystallization Core facility. Funding for this work is provided by the Howard Hughes Medical Institute and NIH. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are 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 beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

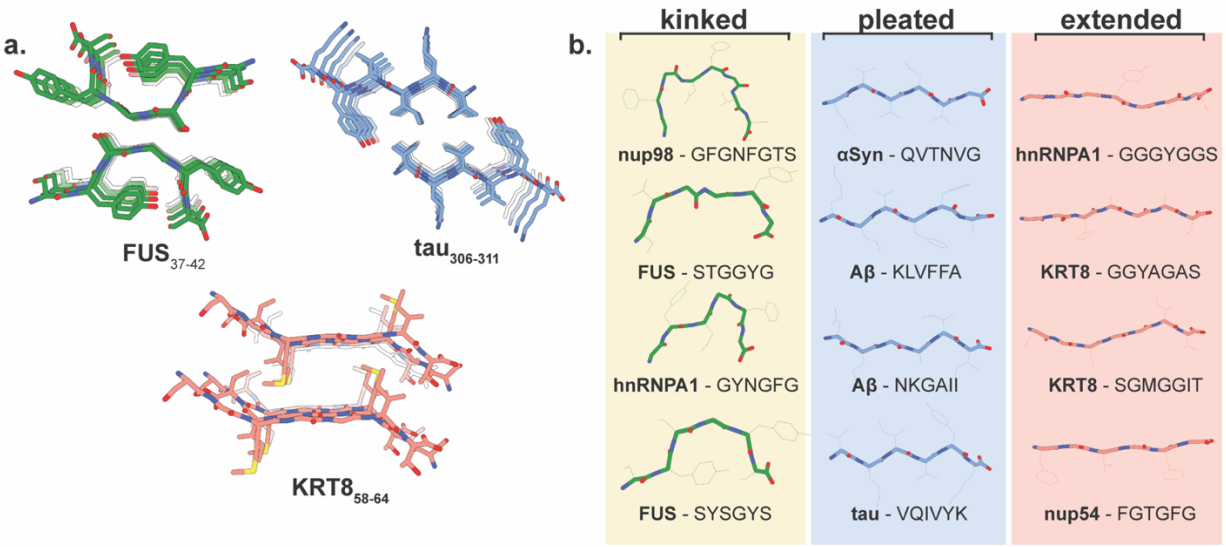


Figure 4.1: Protein backbone conformations found in amyloid structures. a. Cross-sectional view of several amyloid fibrils. Structural determination of amyloid fibrils has revealed their primarily beta-sheet protein backbones can adopt many different conformations. Peptide segments determined from the RNA-binding protein FUS (SYSGYS, green) adopts a kinked beta conformation, while a segment from the microtubule binding domain of tau (VQIVYK, blue) is exclusively pleated. A structure from the intermediate filament protein Keratin-8 (SGMGGIT, red) features two central glycine residues in an extended, nearly linear conformation. **b**. Variations in backbone conformation are found throughput many different amyloid proteins which can be classified into three general categories: kinked, pleated, and extended.

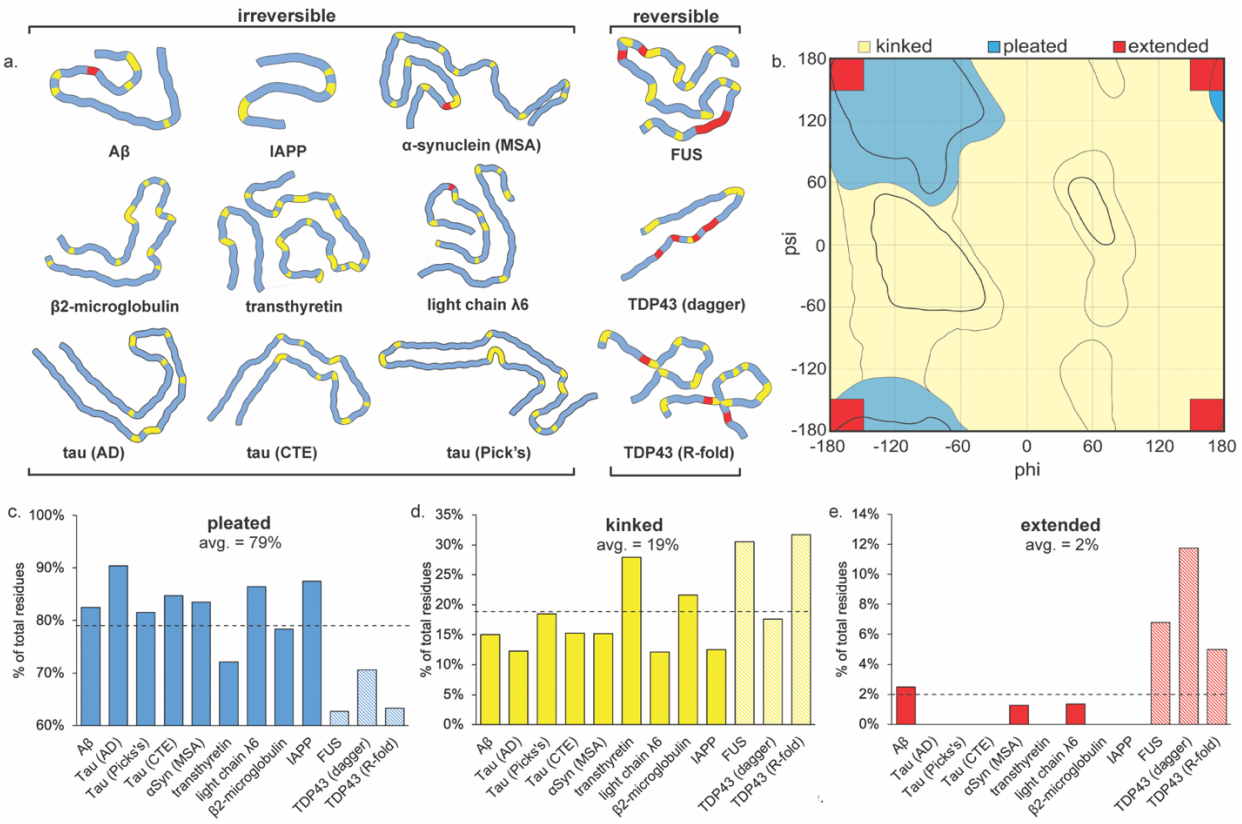


Figure 4.2: Comparison of backbone conformations for full-length amyloid structures. a. Atomic-level structures of full-length amyloid proteins, highlighting backbone conformation. Coloring indicates whether the backbone is in a pleated (blue), kinked (yellow), or extended (red) beta-sheet. Structures from amyloid-β (Aβ), islet amyloid polypeptide (IAPP), tau filaments derived from patients with chronic traumatic encephalopathy (CTE) or Pick's disease, FUS, α-synuclein filaments from patients with multiple system atrophy (MSA), β2-microglobulin, transthyretin filaments from patients with TTR amyloidosis, light chain λ6 filaments from patients with AL amyloidosis, and TDP43 are shown. **b**. Ramachandran plot illustrating the dihedral angle criteria used to define the kinked/pleated/extended conformations of the amyloid structure backbones. Extended beta-strands are classified by both phi and psi angles > |150°|. **c-e**. Quantification of backbone dihedral angles for each amyloid structure, calculated as percentage of total residue number, showing the prevalence of pleated (c), kinked (d), and extended (e) beta-

sheet residues for each structure. FUS and TDP43, known to undergo reversible aggregation, are indicated by striped boxes. A significant enrichment of extended beta-sheets is observed for both FUS and TDP43 compared to the other structures.

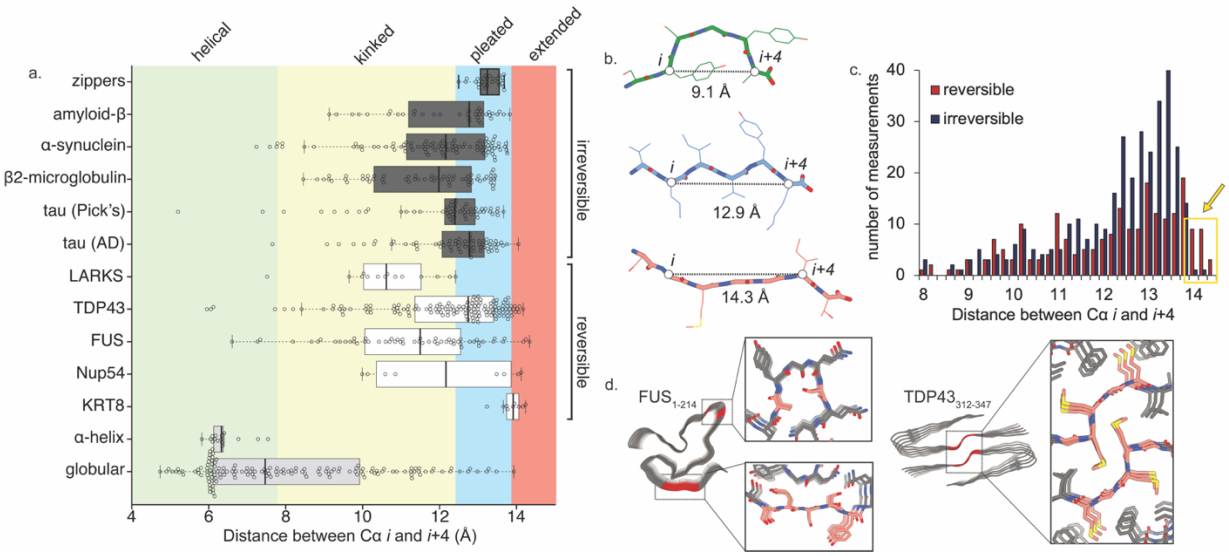


Figure 4.3: Regions of increased inter-residue distance are enriched in reversible amyloid proteins. a. Distances between Cα *i* and *i*+4 for select amyloid protein structures. Distances can be qualitatively grouped into four categories: helical, kinked, pleated, and extended. "Zippers" and "LARKS" were calculated from several model steric zipper and LARKS segments, respectively. "Globular" measurements were calculated from the globular structure of proteinase K, and "helix" was calculated from a helical segment of proteinase K. Comparison of those proteins known to undergo reversible versus irreversible aggregation reveals an enrichment of "extended" type atomic distances in the reversible group. Box plots show means and interquartile range, whiskers show min/max values. **b**. Examples of Cα *i* and *i*+4 measurements taken from three amyloid peptide segments, highlighting the increased distances between residues of the extended beta-strand (red). **c**. Quantification of Cα *i* and *i*+4 inter-residue distance measurements, comparing proteins known to undergo reversible and irreversible amyloid aggregation. Very high inter-residue distances (>14 Å) are almost exclusively seen in the reversible amyloid proteins (highlighted in yellow box). **d**. Examples of high inter-residue distances can be found in the

structures of FUS and TDP43 (indicated in red). Both proteins contain stretches of amino acids in beta-sheet conformations with nearly no pleating.

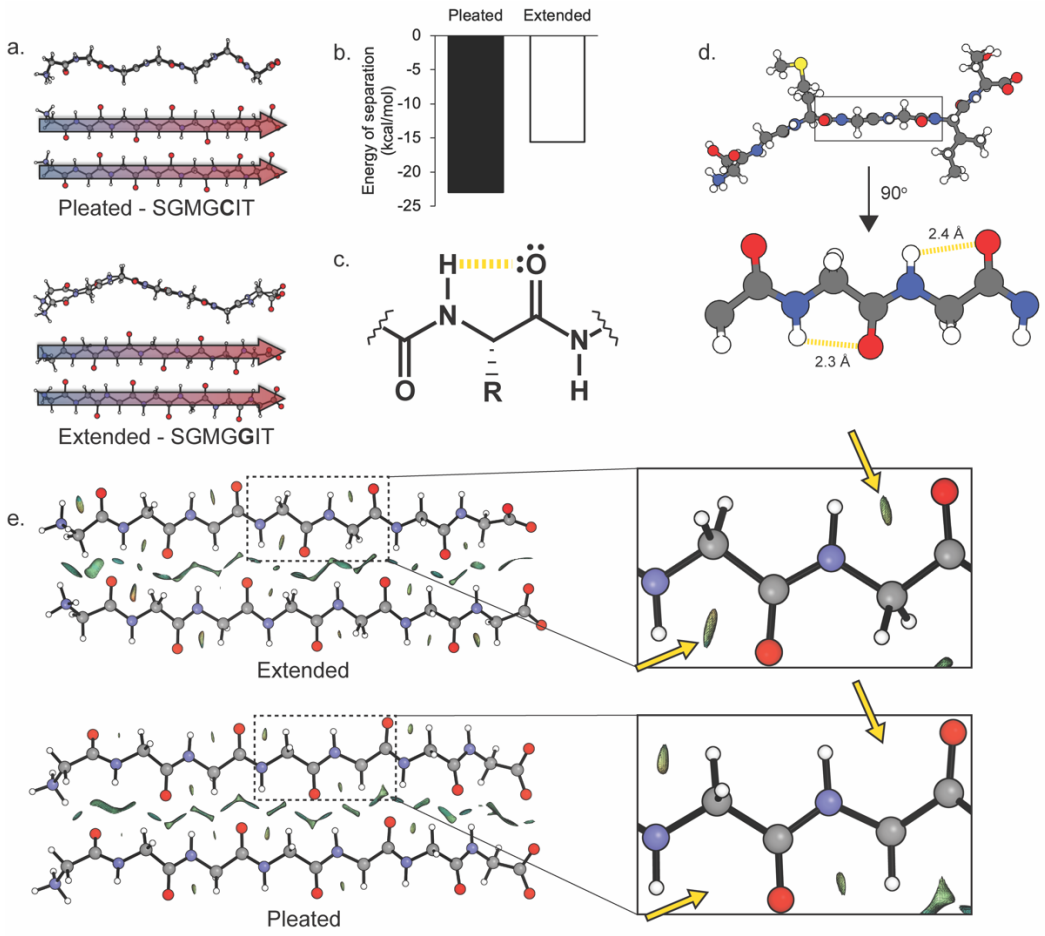


Figure 4.4: Extended beta-sheet conformations decrease fibril dissociation energy and are stabilized by inter-residue hydrogen-bonding. a. Two fibril structures were determined for the wild-type ($_{58}$SGMGGIT$_{64}$) and aggregation-prone mutant sequences ($_{58}$SGMGCIT$_{64}$) of Keratin-8. The G62C mutation converts of an extended beta-sheet in the wild type, to a pleated one in the mutant. **b**. The energy of separation for a pair of strands was determined for both the extended (SGMGGIT) and pleated (SGMGCIT) structures using quantum calculations. More energy is required to separate the pleated strands compared to the extended. **c**. Peptide backbones with near linear dihedral angles can engage in intra-residue hydrogen-bonding between the amide proton and the oxygen lone pair of the adjacent carbonyl. This is known as C5 hydrogen-bonding. **d**. The central glycine residues of the $_{58}$SGMGGIT$_{64}$ structure are in a conformation allowing potential

C5 H-bonding. **e**. Non-covalent interaction (NCI) analysis of both the extended SGMGGIT and pleated SGMGCIT peptides. Density indicates areas of energetically favorable interaction. C5 H-bonding like interactions are seen for both central glycine residues of the extended strand (yellow arrows), which are absent in the pleated strand.

Protein	PDB code
Amyloid-beta	5oqv
IAPP	6y1a
Alpha-synuclein (MSA)	6xyo
Beta2-microglobulin	6gk3
transthyretin	6sdz
Light chain lambda 6	6hud
Tau (AD)	6hre
Tau (CTE)	6nwp
Tau (Pick's)	6gx5
FUS	5w3n
TDP43 (dagger-fold)	6n37
TDP43 (R-fold)	6n3c

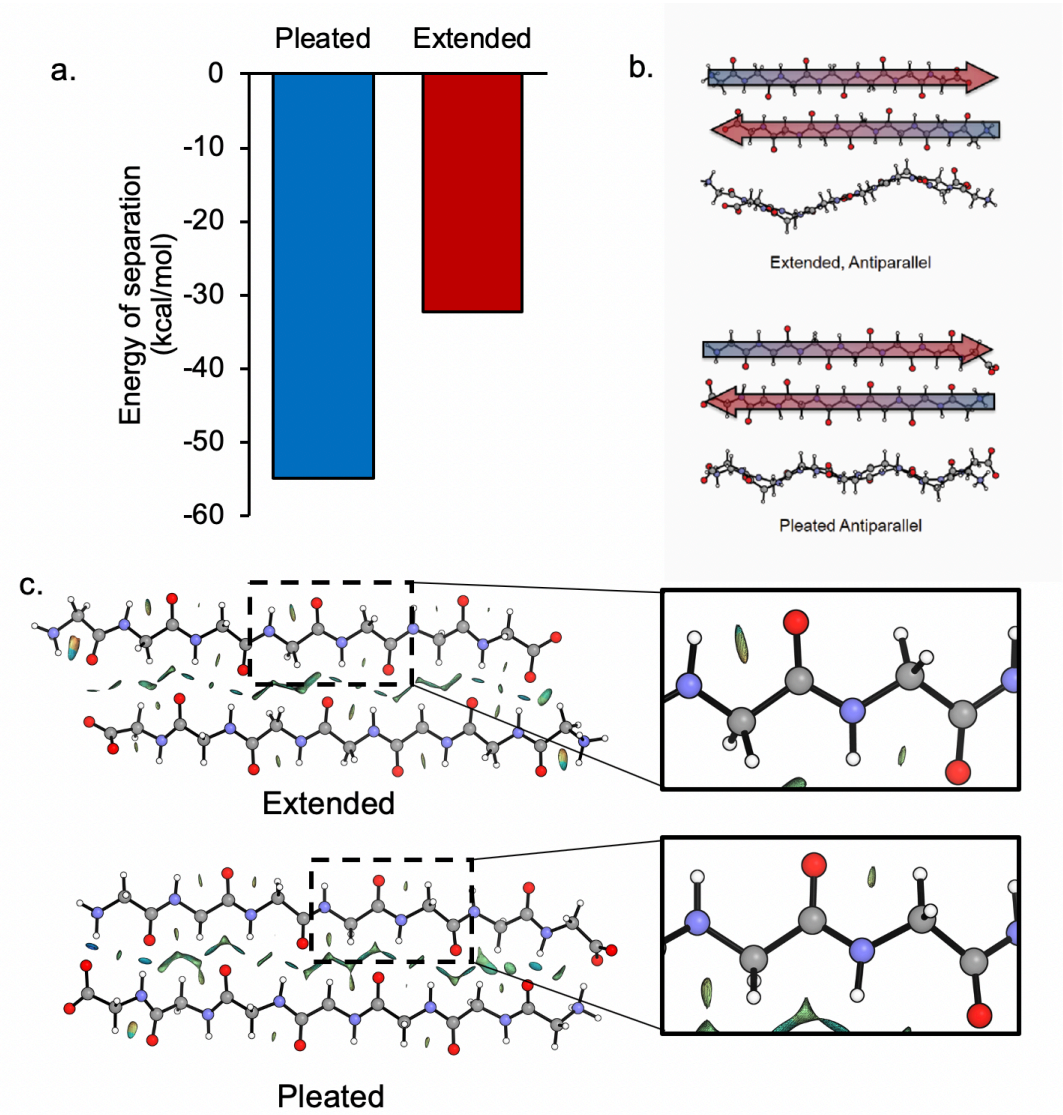
Supplementary Table 4.1: **List of protein structure used for dihedral angle analysis**

Protein	PDB code
zippers	6on9, 2omm, 3ow9
Amyloid-beta	5oqv
Alpha-synuclein	6cu7
Beta2-microglobulin	6gk3
Tau (Pick's)	6gx5
Tau (AD)	6hre
LARKS	6bxx, 6bzp, 6bwz
TDP43	6n37, 6n3c
FUS	5w3n
Nup54	--
KRT8	--
Alpha-helix	5kxv
Globular	5kxv

Supplementary Table 4.2: List of protein structures used for atomic distance analysis

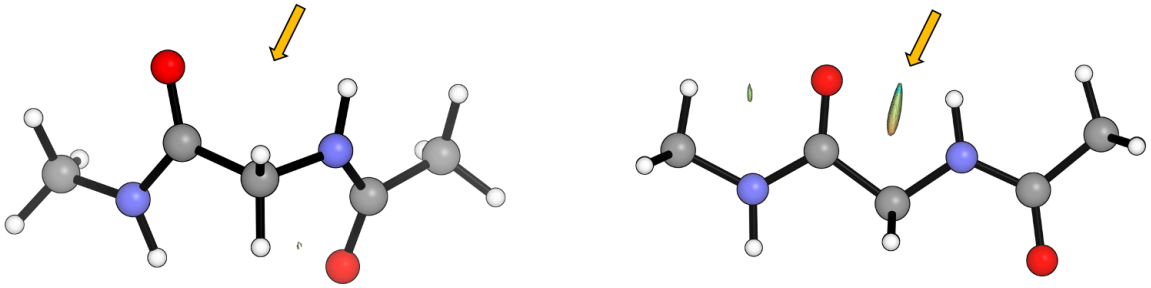
Crystal Peptide	^{63}FGTGFG^{68}
Data Collection	
Beamline	APS 24-ID-E
Space group	$P2_1$
Resolution (Å)	1.2
Unit cell dimensions: a,b,c (Å)	15.71, 9.40, 23.26
Unit cell angles: α,β,γ (°)	90.0,108.5,90.0
Measured reflections	8482 (448)
Unique reflections	2137 (143)
Overall completeness (%)	98.8 (92.3)
Overall redundancy	4.0
Overall Rsym	0.186 (0.386)
Overall I/Sigma	5.4 (2.6)
Last Shell (Å)	1.23-1.20
Refinement	
R_{work}	0.172 (0.276)
R_{free}	0.193 (0.360)
RMSD bond length (Å)	0.021
RMSD angle (°)	2.93
Number of peptide atoms	85
Number of solvent atoms	9
Average B factor of peptide (Å^2)	5.4
Average B factor of solvent (Å^2)	14.5
PDB ID code	--

Supplementary Table 4.3: Nup54 peptide segment crystallographic data and refinement

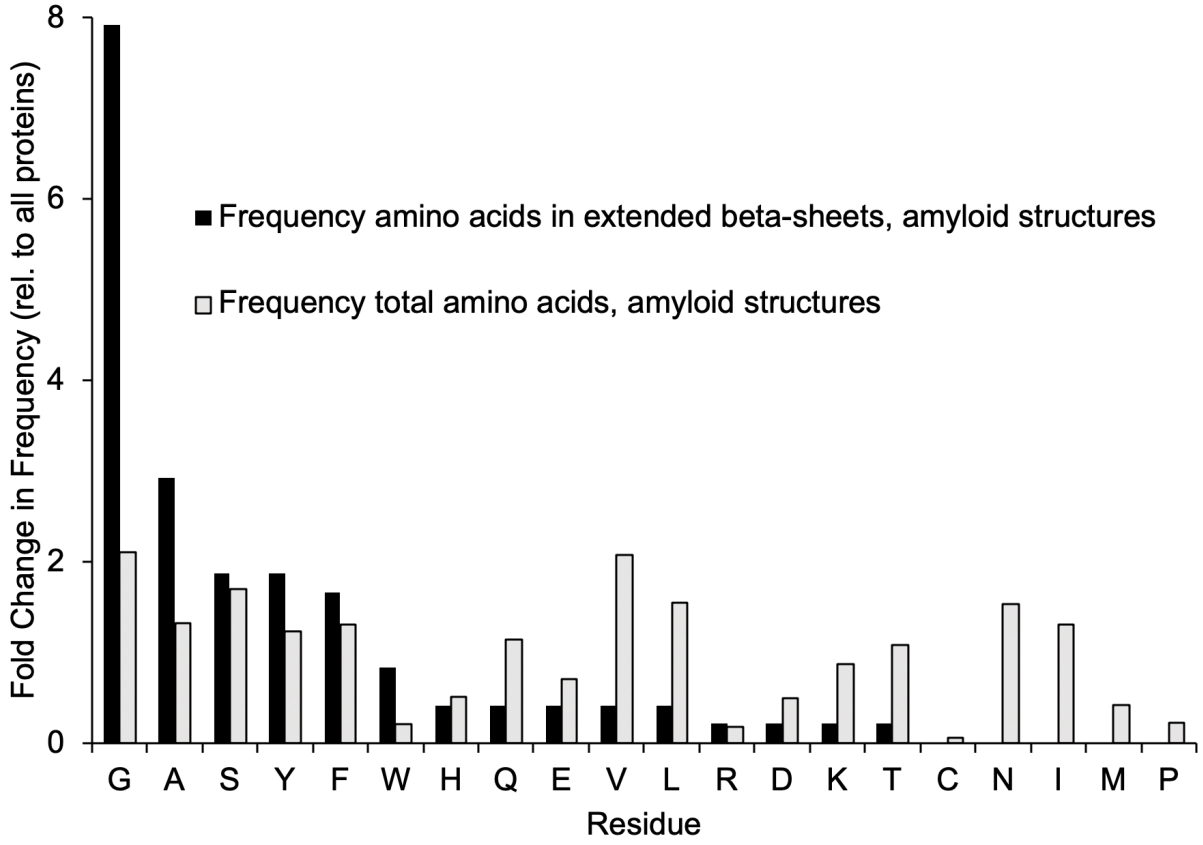


Supplementary Figure 4.1: Quantum calculations of anti-parallel beta-sheets from KRT8. To assess if the difference in energy of separation between the extended and pleated beta-sheets of KRT8 amyloid structures is affected by parallel vs. anti-parallel beta-sheet orientation (see Figure 4.4), calculations were performed on structures in anti-parallel conformation as well. **a-b**. Computed energies of separation for KRT8 amyloid segments $_{58}$SGMGGIT$_{64}$ (extended) and $_{58}$SGMGCIT$_{64}$ (pleated) in anti-parallel conformation show that the pleated form is more stable compared to the extended. c. Non-covalent interaction (NCI) analysis of the two structures shows that the extended form is stabilized by C5 H-bond like interactions in the central glycine residues

(G61 and G62) between the amide hydrogen and carbonyl oxygen. In contrast, conversion to a pleated conformation removes this interaction.



Supplementary Figure 4.2: Non-covalent interaction analysis of AcDegNHMe conformers. To verify that non-covalent interaction (NCI) analysis would detect C5 H-bonding interactions in amyloid structures, NCI analysis was performed on two conformers of AcDegNHMe described by Newberry et al., which samples donor-acceptor distances between the amide hydrogen and carbonyl oxygen. The conformations were then subjected to natural bond orbital (NBO) analysis to assess orbital overlap, and subsequent energetic contribution, between the donor and acceptor pair (carbonyl *p*-type lone pair and σ* orbital of the hydrogen). The conformation on the right was deemed to have significant orbital overlap and energetic stabilization indicating a C5 H-bond, while the left conformation did not. NCI analysis of both reveals the C5 H-bond containing conformation (right) does in fact show interaction density between the C5 H-bond donor and acceptor, while the C5 H-bond absent conformation (left) lacks any NCI density.



Supplementary Figure 4.3: Residue frequency analysis of extended beta-sheets in amyloid structures. Survey of >200 amyloid structures determined through crystallography, cryoEM or ssNMR shows that glycine, alanine, serine, tyrosine and phenylalanine are most commonly found in extended beta-sheet conformations (phi and psi > |150^{o}|), shown in black. This is compared to the overall residue frequency of found in these amyloid structures (gray).

REFERENCES

1. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
2. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
3. Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).
4. Stanković, I. M., Niu, S., Hall, M. B. & Zarić, S. D. Role of aromatic amino acids in amyloid self-assembly. *Int. J. Biol. Macromol.* (2020) doi:10.1016/j.ijbiomac.2020.03.064.
5. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
6. Chuang, E., Hori, A. M., Hesketh, C. D. & Shorter, J. Amyloid assembly and disassembly. *J. Cell Sci.* **131**, (2018).
7. Barnhart, M. M. & Chapman, M. R. Curli Biogenesis and Function. *Annu. Rev. Microbiol.* **60**, 131–147 (2006).
8. Bradley, M. E., Edskes, H. K., Hong, J. Y., Wickner, R. B. & Liebman, S. W. Interactions among prions and prion 'strains' in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **99 Suppl 4**, 16392–16399 (2002).
9. Balbirnie, M., Grothe, R. & Eisenberg, D. S. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2375–2380 (2001).

10. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328–332 (2009).

11. Kato, M. *et al.* Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753–767 (2012).

12. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).

13. Kato, M. & McKnight, S. L. Cross-β Polymerization of Low Complexity Sequence Domains. *Cold Spring Harb. Perspect. Biol.* **9**, a023598 (2017).

14. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627.e16 (2017).

15. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β-sheets that assemble networks. *Science* **359**, 698–701 (2018).

16. Gui, X. *et al.* Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* **10**, 1–12 (2019).

17. Fitzpatrick, A. W. & Saibil, H. R. Cryo-EM of amyloid fibrils and cellular aggregates. *Curr. Opin. Struct. Biol.* **58**, 34–42 (2019).

18. Tycko, R. Solid-state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem.* **62**, 279–299 (2011).

19. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).

20. Astbury, W. T., Dickinson, S. & Bailey, K. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **29**, 2351-2360.1 (1935).

21. Gremer, L. *et al.* Fibril structure of amyloid-β(1-42) by cryo-electron microscopy. *Science* **358**, 116–119 (2017).

22. Röder, C. *et al.* Amyloid fibril structure of islet amyloid polypeptide by cryo-electron microscopy reveals similarities with amyloid beta. *bioRxiv* 2020.02.11.944546 (2020) doi:10.1101/2020.02.11.944546.

23. Falcon, B. *et al.* Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol. (Berl.)* **136**, 699–708 (2018).

24. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).

25. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).

26. Falcon, B. *et al.* Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).

27. Schweighauser, M. *et al.* Structures of α-synuclein filaments from multiple system atrophy. *bioRxiv* 2020.02.05.935619 (2020) doi:10.1101/2020.02.05.935619.

28. Iadanza, M. G. *et al.* The structure of a β2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat. Commun.* **9**, 4517 (2018).

29. 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).

30. Swuec, P. *et al.* Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat. Commun.* **10**, 1269 (2019).

31. 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).

32. Lin, Y. *et al.* Toxic PR poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **167**, 789-802.e12 (2016).

33. Dormann, D. FG-nucleoporins caught in the act of liquid–liquid phase separation. *J. Cell Biol.* **219**, (2020).

34. Masuda, T. *et al.* Atomic resolution structure of serine protease proteinase K at ambient temperature. *Sci. Rep.* **7**, 45604 (2017).

35. Ku, N.-O., Gish, R., Wright, T. L. & Omary, M. B. Keratin 8 Mutations in Patients with Cryptogenic Liver Disease. *N. Engl. J. Med.* **344**, 1580–1587 (2001).

36. Newberry, R. W. & Raines, R. T. A prevalent intraresidue hydrogen bond stabilizes proteins. *Nat. Chem. Biol.* **12**, 1084–1088 (2016).

37. Scheiner, S. Relative strengths of NH..O and CH..O hydrogen bonds between polypeptide chain segments. *J. Phys. Chem. B* **109**, 16132–16141 (2005).

38. Murray, D. T. & Tycko, R. Side Chain Hydrogen-Bonding Interactions within Amyloid-like Fibrils Formed by the Low-Complexity Domain of FUS: Evidence from Solid State Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **59**, 364–378 (2020).

39. Johnson, E. R. *et al.* Revealing noncovalent interactions. *J. Am. Chem. Soc.* **132**, 6498–6506 (2010).

40. Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 2.3.