Title
Structural and Biochemical Studies of the Amyloid-Forming Proteins Tar DNA Binding Protein 43, Superoxide Dismutase 1, and Alpha-Synuclein

Author
Guenther, Elizabeth Lynn

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structural and Biochemical Studies of the Amyloid-Forming Proteins TAR DNA Binding Protein 43, Superoxide Dismutase 1, and Alpha-Synuclein

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

by

Elizabeth Lynn Guenther

2017

ABSTRACT OF THE DISSERTATION

Structural and Biochemical Studies of the Amyloid-Forming Proteins
Tar DNA Binding Protein 43, Superoxide Dismutase 1, and Alpha-Synuclein

by

Elizabeth Lynn Guenther
Doctor of Philosophy in Biochemistry and Molecular Biology
University of California, Los Angeles, 2017
Professor David S. Eisenberg, Chair

The formation of elongated, unbranched fibrillar protein aggregates, termed amyloid, has been linked to neurodegenerative diseases including Alzheimer's Disease, Parkinson's Disease and Amyotrophic Lateral Sclerosis (ALS). For decades, scientists have sought to determine the structures and polymorphic features of these aggregates to better understand their complex assembly and role in disease. Here, we show structures from three amyloid proteins, TAR DNA Binding Protein 43 (TDP-43), Superoxide Dismutase 1 (SOD1), and α-synuclein, offering insight into amyloid and protein aggregation in general.

First, we investigated TDP-43 as a model for the polymorphic capabilities of pathological amyloid aggregation. From x-ray diffraction, electron diffraction and cryoEM, we show that the segment, ^{247}DLIIKGISVHI^{257}, from the second RNA recognition motif forms an array of amyloid polymorphs. These associations include seven distinct interfaces displaying five different symmetry classes of steric zippers. The polymorphic nature of this segment illustrates at the molecular level how amyloid proteins can form diverse fibril structures.

In the second section, we investigated the role of the low complexity domain (LCD) of TDP-43 in reversible stress granule aggregation and irreversible pathogenic aggregation. We determined six atomic resolution segment structures from the LCD of TDP-43 that offer insight into regions of the protein responsible for aggregation. We illustrate how three of these structures, ^{300}GNNQGSN^{306}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402} form fibrils and exhibit characteristic steric zippers, indicative of pathogenic aggregation. We present an atomic structure for segment ^{312}NFGAFS^{317}, which reveals a fibril with a kinked interface that is reversible and may play a role in stress granule assembly. We present a model that combines previous literature with our new structural findings, to illustrate how stress granule assembly can lead to pathogenic aggregation.

Third, we structurally characterized peptide segments from SOD1 because pathological deposition of mutated SOD1 accounts for ~20% of the familial ALS (fALS) cases. We used a computational approach to discover four segments from the protein that form fibril-like aggregates and subsequently solved the structure of three of those segments: ^{101}DSVISLS^{107}, ^{147}GVIGIAQ^{153} and ^{147}GVTGIAQ^{153}. We found, through the use of proline mutations, that two of these ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153} are likely to trigger the aggregation of full-length SOD1, suggesting common molecular determinants of fALS and sALS.

In the fourth section, we investigated whether amyloid oligomers or fibrils were responsible for SOD1 toxicity in ALS. We determined the corkscrew-like structure of a cytotoxic segment of SOD1 in its oligomeric state. Through time course cytotoxicity assays, we demonstrated that the toxicity is likely a property of soluble oligomers and not large insoluble aggregates. Our work adds to evidence that the toxic oligomeric entities in protein aggregation diseases contain antiparallel, out-of-register β-sheet structures and identifies a target for structure-based therapeutics in ALS.

In the final section, we characterized the amyloidogenic core of α-synuclein, the main component of Lewy bodies, the neuron-associated aggregates seen in Parkinson's disease. We

utilized a new technique known as Micro-electron diffraction (MicroED) to solve the structure of crystals of an 11-residue segment we term NACore that are invisible by optical microscopy. The structure reveals protofibrils built of pairs of face-to-face β-sheets that exhibit a fiber diffraction pattern similar to full-length α-synuclein. We present a model of the toxic, full-length α-synuclein fibril, incorporating the NACore and an additional segment.

En masse, the structures in this dissertation offer broad insights into pathological and reversible protein aggregation. The structures provide a starting point for additional studies aimed at developing therapeutics for amyloid formation in neurodegenerative disease.

The dissertation of Elizabeth Lynn Guenther is approved.

James U. Bowie

Feng Guo

Todd Yeates

David S. Eisenberg, Committee Chair

University of California, Los Angeles

2017

To my mom – my best friend, role model and the most generous person I knew. A year gone, but always in my heart. You ARE my inspiration.

To my dad – you taught me that gender doesn't dictate dreams. You were the glue that held together my world when it crumbled last year. You ARE my rock.

To my friends and family - who love me at my worst, encourage me to pursue my dreams, and teach me how to grow with life experiences. You make my life better.

Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

They say it takes an army to raise a child, and I am now convinced it takes an army to earn a PhD. This work would not have been possible without the support of my loved ones. I first need to thank my parents who have supported me in every dream I have ever had. Thank you for never allowing me to accept less than my best and for teaching me that every day we are alive is a gift to be cherished. Thank you for treating my friends like family and for being the best role models of what it means to have a career, be an involved parent, and still be generous with your time. Most importantly, thanks to both of you for refusing to let me leave graduate school when mom was diagnosed last year and passed away. Your selfless love and dad's support throughout this year is the only reason this dissertation is complete. Lastly, thank you to my family, specifically my Grump, and my friends from New Jersey, Boston College and Los Angeles. You are all the pieces that make up my universe. Thank you all for challenging me to grow and become a better person each day - for that I am indebted to you.

Ultimately, the personal connections are what helped me navigate the stress of graduate school, but I would have never been here without the support of some amazing scientists. Thank you to my undergraduate research mentor, Heather Cooke, for taking me under her wing as a freshman at Boston College and showing me early on that a woman could be a successful researcher. Thank you to my mentor here at UCLA, David Eisenberg, for re-invigorating my passion for science during my recruitment trip, before I ever even enrolled. Your passion to leave a mark on the world and effect change through amyloid research is contagious. The freedom with which you allow us to operate and your trust in us as students gave me the confidence to develop research projects and pursue difficult problems. It was an honor to work for you. Finally, I also need to thank my committee members, Jim Bowie, Feng Guo, and Todd Yeates for challenging me to develop independent projects and offering invaluable suggestions at our structure group meeting. It was a privilege working in Boyer with all of you.

The greatest aspect of my graduate career was the opportunity to work with so many amazing collaborators. None of this work would have been possible without their incredible input and contributions to my dissertation research. Below, I have detailed the contributions of my co-authors for each draft and published manuscript.

Chapter 1 is a draft manuscript for Nature Structure and Molecular Biology by Elizabeth Guenther, Peng Ge, Hamilton Trinh, Michael, Sawaya, Duilio Cascio, David Boyer Tamir Gonen, Hong Zhou and David Eisenberg. Biochemical characterization of RRM2 segments was completed by EG and HT. Peptide crystals were grown by EG and HT. DB collected electron diffraction data for the ^{247}DLIIKGISVHI^{257} crystals. Structural determination of ^{248}LIIKGI^{253} and ^{247}DLIIKGISVHI^{257} was done by EG, HT, MS and DC. EG prepared the ^{247}DLIIKGISVHI^{257} fibril sample for cryo-EM, which PG collected data for and solved the structure. Further structural analysis of the unique interfaces and backbones as well as model generation was done by EG, with the help of MS.TG and HZ served as collaborating principal investigators, sharing resources and time with us. DE served as the lead investigator, offering critical commentary and feedback throughout the research project. The manuscript was written by EG, MS and DE.

Chapter 2 is a draft manuscript for submission by Elizabeth Guenther, Hamilton Trinh, Michael Hughes, Michael Sawaya, Duilio Cascio and David Eisenberg. Biochemical characterization and crystal growth were completed by EG and HT. Data collection, processing and structure determination of ^{300}GNNQGSN^{306} was done by HT and MS. Data collection and structure determination of ^{370}GNNSYS^{375}, ^{396}GFNGGFG^{402}, ^{312}NFGAFS^{317}, and ^{312}NFGTFS^{317} and ^{312}NFGEFS^{317} was done by EG and MS. DC contributed to the data processing for ^{312}NFGEFS^{317}. Models for the role of stress granules in pathological aggregation were created by EG. DE was the principal investigator for this project, providing endless and invaluable feedback along the way. The manuscript was written by EG and DE.

Chapter 3 is a reprint of the published manuscript, Magdalena Ivanova, Stuart Sievers, Elizabeth Guenther, Lisa Johnson, Duane Winkler, Ahmad Galaleldeen, Michael Sawaya, John Hart and David Eisenberg. Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS. *PNAS*. 2017 111:1 197–201,doi: 10.1073/pnas.1320786110.The project was designed by MI,, SS, DW, AG., MS, PH, and DE. designed research. MI, MS and SS solved the peptide structures. Biochemical characterization of segments was completed by MI, while EG completed the prolin mutation characterization on full length SOD1. LJ synthesized peptide samples. PH helped analyze data and the manuscript was written by MI, SS, LJ, DW, MS, PH and DE. All authors offered commentary on the manuscript.

Chapter 4 is a reprint of a recently accepted manuscript, Smriti Sangwan, Anni Zhao, Katrina L. Adams, Christina K. Jayson, Michael R. Sawaya, Elizabeth Guenther, Albert C. Pan, Jennifer Ngo, Destaye Moore, Angela B. Soriaga, Thanh D. Do, Lukasz Goldschmidt, Rebecca Nelson, Michael T Bowers, Carla M. Koehler, David E. Shaw, Bennett G. Novitch, David S. Eisenberg. Atomic structure of a toxic, oligomeric segment of SOD1 linked to amyotrophic lateral sclerosis (ALS). *PNAS* (2017). Project was designed by SS, RN and DE. SS and MS solved the corkscrew structure. AZ, KA, and SS performed toxicity experiments on neuronal cell lines. CJ completed the zebrafish work. EG and SS designed and completed the time course assay for fibril formation and toxicity. LG, RN, TD, AB, DM and JN contributed to some experimental work and offered experimental feedback. The manuscript was written by SS, MS and DE with all authors offering commentary.

Chapter 5 is a reprint of the article, Jose A. Rodriguez, Magdalena I. Ivanova, Michael R. Sawaya, Duilio Cascio, Francis Reyes, Dan Shi, Smriti Sangwan, Elizabeth L. Guenther, Lisa M.

Johnson, Meng Zhang, Lin Jiang, Mark A. Arbing, Brent Nannega, Johan Hattne, Julian Whitelegge, Aaron S. Brewster, Marc Messerschmidt, Sébastien Boutet, Nicholas K. Sauter, Tamir Gonen, David Eisenberg. Structure of the toxic core of α-synuclein from invisible crystals. *Nature*. 525: 486-490. doi:10.1038/nature15368. M.I.I. characterized the α-synuclein segments and crystals. EG completed the aggregation kinetics experimens for NACore and SubNACore. M.I.I. and S.S. conducted the toxicity assays. L.M.J. synthesized and purified NACore peptide. M.A.A. prepared the N-terminally acetylated a-synuclein. S.S. and M.Z. prepared wild-type α-synuclein. J.W. performed the mass spectrometry analyses of a-synuclein. M.I.I. and L.M.J. crystallized NACore. E.G. grew crystals of SubNACore. E.G., M.I.I. and M.R.S. collected and processed the data and solved the structure of SubNACore. L.J. and J.A.R. identified and crystallized PreNAC. J.A.R., D.S., B.L.N. and T.G. collected MicroED data on PreNAC and NACore nanocrystals. J.A.R., F.E.R., J.H., T.G., L.J., M.R.S., and D.C. processed theMicroED data and solved the structure of PreNAC and NACore. J.A.R.,M.R.S., D.C., M.M. and S.B. collected XFEL diffraction fromNACore nanocrystals. A.S.B. and N.K.S. processed the XFEL data.M.R.S and L.J. built the structure model of A53T a-synuclein protofibril. J.A.R., M.I.I., M.R.S., D.C., S.S. and E.G. prepared the figures. J.A.R.,M.I.I., M.R.S., D.C., T.G. and D.S.E. wrote the paper, and all authors commented.

In addition to these collaborators I also need to thank the many members of the Eisenberg Lab. There are far too many of you to name but special thanks to Magda for you mentorship, Duilio and Mikey for your endless patience and help in structure determination and Pascal for your friendship. Thank you to Michael Collazo at the Crystallization Core and the staff at the APS NECAT 24-ID-E beamline for providing the critical support needed for each chapter of this paper.

VITA

2007-2011	BS in Chemistry, Boston College
2007-2009	Beckman Undergraduate Research Fellow, Bruner Group, Boston College
2012-2015	UCLA-NIH Chemistry and Molecular Biology Training Grant
2012	UCLA Hanson Dow Teaching Assistant Award
2012-2013	Teaching Assistant, Department of Chemistry, UCLA
2015	GSA Jeffrey L. Hanson Distinguished Service Award Finalist, UCLA
2015	Business of Science Center Venture Competition Winner, UCLA
2016-2017	Teaching Assistant, Anderson School of Business, UCLA

Select Publications and Presentations

Sangwan S, Zhao A, Adams KL, Jayson CK, Sawaya MR, **Guenther EL**, Pan AC, Ngo J, Moore D, Soriaga AB, Do TD, Goldschmidt L, Nelson R, Bowers MT, Koehler CM, Shaw DE, Novitch BG, Eisenberg DS. *PNAS*. Accepted.

Rodriguez, J.A., Ivanova, M.I., Sawaya, M.R., Cascio, D., Reyes, F.E., Shi, D., Sangwan, S., **Guenther, E.L.**, Johnson, L.M., Zhang, M., et al.. Structure of the toxic core of α-synuclein from invisible crystals. *Nature*. **2015.** Epub ahead of print.

Ivanova MI, Sievers SA, **Guenther EL**, Johnson LM, Winkler DD, Galaleldeen A, Sawaya MR, Hart PJ, Eisenberg DS. Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS. *PNAS*. **2014** 111, 197

Cooke H.A, **Guenther E.L**, Shen B., Bruner S.D. Molecular basis of substrate promiscuity for the SAM-dependent O-methyltransferase NcsB1, involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin. *Biochemistry* **2009** 48, 9590.

Guenther EL, Hughes MP, Trinh H, Eisenberg DS. Reversible Versus Pathological Amyloid Formation by TAR DNA Binding Protein-43 (TDP-43). Molecular Mechanisms and Physiological Consequences of Protein Aggregation. June 21-26, 2015. West Palm Beach, Florida.

Guenther EL. Ivanova M., Sangwan S., Trinh H, Sievers S., Johnson L., Eisenberg DS. Proline substitutions reveal aggregation-triggering segments of SOD1 and support a common pathway for familial and sporadic ALS. Seaborg Symposium. October 26, 2013. UCLA.

Guenther EL. Ivanova M. Eisenberg DS. Crystallization and Characterization of Segments from the Amyloidogenic Protein, α-synuclein. Molecular Mechanisms and Physiological Consequences of Protein Aggregation. June 23-28, 2013. Big Sky, Montana.

CHAPTER 1

Molecular Structures of a Key Segment of Amyloid-forming Protein TDP-43 Reveal Prodigious Packing Polymorphism

INTRODUCTION

The formation of elongated, unbranched fibrillar protein aggregates, termed amyloid, has been linked to neurodegenerative diseases including Alzheimer's Diseases, Parkinson's Disease, prion diseases and Amyotrophic Lateral Sclerosis (ALS) among others (1-4). For decades scientists have sought molecular structures for amyloid aggregates and have attempted to understand the range of polymorphic fibrillar structures formed by proteins in the amyloid state (5-7). Contributions to understanding amyloid structure have come in recent years from diffraction methods (8-10), solid-state NMR (11-14), and electron cryo-microscopy (15-17). These studies have shown that amyloid fibrils consist of packed β-sheets that run parallel or roughly parallel to the fibril axes. Each β-sheet adheres to its neighboring sheets because the sidechains of each β-strand project roughly perpendicular to the fibril axis, toward the neighboring sheet. Because the sidechains of mating sheets can interdigitate like the teeth of a zipper, this interaction has been called a steric zipper (8). Steric zippers can form between identical or non-identical β-sheets. A common property of β-sheets in the amyloid state is that they can pack in a variety of ways to form several different fibril structures. Such polymorphic amyloid assemblies are well known. Research on amyloid-β has demonstrated how conformational variations can lead to distinct assemblies of amyloid aggregates (5, 6). Different polymorphs of the same amyloid protein can exhibit different biophysical or biomedical properties. In the case of prion protein, distinct strains can confer different toxic effects as well as transmission capabilities based on the fibrillar structure the protein has adopted (18). Since

polymorphs can determine transmission and severity of disease, understanding polymorph structure has medical implications.

Structural studies of amyloid at the atomic level have offered evidence for two basic models to explain the frequent observation that a single protein produces amyloid fibrils of varying morphologies or polymorphs. In the first model, termed segmental polymorphism, the segment forming the zipper spine differs between amyloid fibrils (19, 20). Segmental polymorphism has been observed in a number of proteins to date, including islet amyloid polypeptide (IAPP), in which various six- and eleven-residue segments of IAPP are each capable of forming zippers, independently (21,22). In a second model, termed packing polymorphism, a single zipper-forming segment packs differently in different amyloid fibrils(19, 20). Outside of amyloid research, this type of packing is commonly referred to as modal polymorphism. We have observed this polymorphism by the amyloid-β segment, ^{16}KLVFFA^{21}, in which three distinct zipper interfaces were revealed by three crystal forms (23). Both of these types of polymorphism presume homotypic zipper interfaces: two mated sheets both composed of the same segment. However, analogous polymorphisms are likely also to exist for heterotypic zippers, where the two mated sheets are derived from sequentially distant segments (19, 20) as in solid-state NMR structures of α-synuclein and amyloid-β (11-13). The abundance of examples of segmental and packing variations observed between distinct fibrils raises the question of whether these same variations play a role within a single fibril, a phenomenon termed positional polymorphism. If so, to what extent does positional polymorphism expand the repertoire of amyloid polymorphs accessible to a given sequence?

Currently, one of the proteins in the forefront of research on amyloid disease is TAR DNA Binding Protein 43 (TDP-43) owing to the presence of its aggregates in patients with ALS and Frontotemporal Lobar Degeneration (FTLD) (24, 25). Recent findings have shown that aggregates of TDP-43 are also found in patients with Alzheimer's, Parkinson's and Huntington's Diseases (26-29). Distinct from other amyloid and amyloid-like proteins, TDP-43 is viewed as

an essential protein for cellular homeostasis. It has been shown that knockdown of the expression of TDP-43 in mice results in embryonic lethality (30-32). In addition, other studies have demonstrated that changes in TDP-43 expression levels leads to motor deficits and loss of body fat in mouse models (30, 33). This has led to the hypothesis that TDP-43 aggregation can confer toxicity through a loss of function mechanism. In this case, TDP-43 is redistributed from the nucleus to the cytoplasm as seen in neuronal cell culture experiments (34, 35). The effects of the nuclear clearance of TDP-43 is still being investigated, however research has shown that loss of TDP-43 function results in the inclusion of cryptic exons in transcripts, improper regulation of microprocessing and dysregulation of Rho family GTPases (33, 36, 37). These findings have made understanding the mechanism by which TDP-43 aggregates an essential focus of research in ALS and other neurodegenerative diseases. By understanding how the protein aggregates, inhibitors against the aggregation can be developed and protein function can be restored, as demonstrated with p53 in ovarian carcinomas (38).

Here, we apply three structural methods, x-ray diffraction, electron diffraction, and cryoEM, to determine three different assemblies of a key segment in the second RNA Recognition Motif (RRM2) of TDP-43. In addition to segmental and packing polymorphisms, we reveal that positional polymorphism contributes an important alternative route to amyloid fibril assembly. We propose such positional polymorphism is not limited to TDP-43, but a widespread phenomenon among amyloid proteins in general.

RESULTS

Segments from the RRM2 of TDP-43 exhibit characteristic amyloid properties

We decided to target the RRM2 of TDP-43 for structural studies based on three previous findings. First, researchers have shown that patients with ALS and FTLD have aggregates in the brain and spinal cord that are primarily composed of C-terminal fragments ranging in size from 15-35kDa (35, 39, 40). The longer fragments include part of the first RNA Recognition Motif

(RRM1), the RRM2 and the C-terminal tail while the shorter segments include only part of the RRM2 and the C-terminal tail. Second, a study showed that inclusion of a truncated RRM2 at residue 208 increased the number of TDP-43 aggregates present in cells and also increased cellular death (41). Finally, a study illustrated that the RRM2 of TDP-43 adopts an intermediate state during folding and that the presence of the RRM1 aids the RRM2 in folding into its final conformation (42). These findings together indicated that the unfolding or destabilization of the RRM2 by cleavage could expose an aggregation prone region that was previously protected and thus facilitate amyloid aggregation of TDP-43.

Since amyloid spines are typically composed of segments ranging from 6-11 residues, we used the 3D Profile Method, ZipperDB, to narrow our region of focus in the RRM2 to a 13-residue segment ^{247}DLIIKGISVHISN^{259}(Figure 1.1A) (43). Within this region, 4 of the top 10 potential zipper forming segments of TDP-43 are located based on Rosetta energy. This is significant as there are 408 possible 6-residue zippers that the 3D Profile Method evaluates for the 414 residue TDP-43. We decided to characterize biochemically 11 segments within this region, ranging from 6 to 11 residues in length with average of hydropathicity values ranging from 0.07 to 2.17 (Figure 1.1A, Supplementary Table 1.1) (44).

To determine if these segments exhibit amyloid properties, we tested their propensity to aggregate, their stability under denaturing conditions, and their fibril diffraction pattern. We find that all 11 segments form fibrils when shaken for one week (Figure 1.1B). Based on previous studies demonstrating that amyloid fibrils are stable when incubated with SDS and heat, we subjected the fibrils to a 2% SDS solution at 70°C for 15 minutes (45). We find that under these extremely harsh denaturing conditions, the fibrils exhibit moderate to high stability. Segments ^{247}DLIIKGISV^{255}, ^{247}DLIIKGISVH^{256}, ^{248}LIIKGISVHI^{257}, and ^{252}GISVHI^{257} all exhibit the highest stability illustrating almost no dissociation of fibrils (Figure 1.1C and Supplementary Figure 1.1). Finally, we checked whether the dried fibril samples exhibit the characteristic cross-β diffraction pattern indicative of amyloid fibrils (46). For the nine samples on which we were able to study

diffraction, we found that all exhibit reflections at spacings of 4.6-4.7Å and 8.8-11.5Å, which supports the presence of a typical steric-zipper-based amyloid fibril (Supplementary Figure 1.2).

Segments ^{248}LIIKGI^{253} and ^{247}DLIIKGISVHI^{257} form two distinct steric zipper structures.

Following biochemical characterization of peptide segments from the 13-residue range within the RRM2 we attempted crystallization of all 11 segments. We successfully determined the structures of two constructs, ^{248}LIIKGI^{253} and ^{247}DLIIKGISVHI^{257}, through the use of two methods: x-ray diffraction and microelectron diffraction (MicroED) respectively.

Crystals of ^{248}LIIKGI^{253} were grown in hanging drop 96-well trays and were visible by light microscopy (Figure 1.2A). Fibril diffraction pattern revealed rings at 4.6Å and 11.5Å, which are consistent with amyloid structure (Figure 1.2B). Initial processing of the diffraction data from a micro-focus x-ray beamline provided unit cell dimensions of 11.54Å, 9.59Å, and 21.18 Å indicating an antiparallel β-sheet. Molecular replacement and refinement revealed a face=back, up-up, in-register antiparallel Class 7 steric zipper (Figure 1.2C, Table 1.1) (9). The hydrogen bonding network is such that the sheet comprises only the first four residues, LIIK (Figure 1.2D), thus leaving two-residue overhangs and a ragged appearance to the sheet edges. Due to its ragged edges, the buried surface area is relatively low at 117Å^2, however the shape complementarity is high at 0.78.

In contrast to the microcrystals formed by segment ^{248}LIIKGI^{253}, crystals of ^{247}DLIIKGISVHI^{257} were not visible by light microscopy owing to their small size. The nanocrystals of ^{247}DLIIKGISVHI^{257}, which we term the RRMcore, were formed by shaking in pH 8.5 buffer, and at ~100nm in width were visible only by electron microscopy (Figure 1.2E). We decided to use the cryoEM method, MicroED, on the basis of its previous success in determining key structures from three other amyloid proteins, α-synuclein, tau and IAPP (21, 47-49). With this technique, an electron microscope is used to collect data in diffraction mode. The structure of ^{247}DLIIKGISVHI^{257} was determined to be a parallel, face-to-back Class 2 steric zipper (Figure

1.2G and 1.2H and Table 1.2) (9). The steric zipper interface is formed by the inter-digitation of residues DLIIKGIS from one strand with KGISVHI from the mating strand. The tight interface results in 350 Å^2 buried surface area and a shape complementarity of 0.67. Fibril diffraction of the sample illustrated characteristic amyloid reflections at spacings of 4.72Å and 10.4Å (Figure 1.2F).

The structures of ^{248}LIIKGI^{253} and ^{247}DLIIKGISVHI^{257} reveal two possible ways that segments from the RRMcore can associate to form an amyloid spine. The heterotypic nature of the two structures suggests that the RRMcore can form polymorphic assemblies.

CryoEM Structure of ^{247}DLIIKGISVHI^{257}Fibril Reveals a 27-Filament Helical Fibril

During biochemical characterization of segments within the RRMcore we noticed that aggregates of ^{247}DLIIKGISVHI^{257} grown at pH 8.5 differed in morphology from those grown at pH 7.5. The crystals formed at pH 8.5 revealed the parallel Class 2 zipper described above, while the aggregates grown at pH 7.5 displayed fibrillar morphology, but not crystalline. Fibril diffraction of this sample revealed cross-β diffraction at 4.65Å and 10.5Å, indicating the presence of steric zipper interfaces formed by parallel β-sheets (Figure 1.3B). By electron microscopy, the fibrils exhibited a regular helical twist (Figure 1.3A and Supplementary Figure 1.3). Following studies that used cryoEM to determine structures of helical fibrils of amyloid-β and light chain amyloid (15, 16), we applied this method to our fibrils

CryoEM experiments on the ^{247}DLIIKGISVHI^{257} sample revealed 8 classes of fibrils including cylinders, twists, sheets and a 3-start helix (Supplementary Figure 1.4). The most dominant species, comprising ~60% of the sample, was the 3-start helix, a common organization pattern of biological fibrils (50). Subsequent class averaging of the 3 start-helix allowed us to visualize multiple faces of the fibril, generating enough data for single particle reconstruction (Supplementary Figure 1.5).

Using single particle reconstruction and accounting for helical symmetry, a model of the structure was determined at 3.7Å resolution (Figure 1.3C and Supplementary Figure 1.6). The

^{247}DLIIKGISVHI^{257} 3-start helix appears as a many-filament parallel β-sheet fibril, exhibiting 3-fold screw symmetry. A single layer cross-section of the fibril comprises 27 peptide strands, with nine strands in the asymmetric unit (Figure 1.3D). Each of the nine strands displays a kinked, straight, or curved conformation, contributing to the complex polymorphism of the structure (Supplementary Figure 1.7). The asymmetric units are related by a left handed three-fold screw with a rise of 1.60Å and a rotation of 120.4° per step. The individual protofilament β-sheets, 27 in total, exhibit a left-handed twist, as is common for β-sheets. The density map indicates that the β-sheets at the center and outer perimeter of the fibril have disordered termini (Figures 1.3E and 1.3F). Thus, we have modeled only the six ordered residues for these.

This complex helical fibril structure illustrates a tight packing of the ^{247}DLIIKGISVHI^{257} segment into a polymorphic structure, the first amyloid molecular structure of its kind. The packing offers insight into how the RRMcore of TDP-43, and amyloid proteins in general, can use the flexibility of their peptide backbones to adopt multiple conformations and form diverse steric zipper interfaces, even in a single structure. This expanded repertoire of geometric shapes reveals the capacity of protofilaments to fit together to fill space, exclude water, and bury hydrophobic surfaces, as discussed below (Supplementary Figure 1.8).

Structures from the RRMcore Offer Insight into Polymorphic Features of Amyloid Formation

Our detailed analysis of the RRMcore polymorphs reveals that there are seven distinct interfaces that are formed among the three structures (Figure 1.4 and Supplementary Table 1.2). For the cryoEM structure we considered only interfaces that have at least 3 inter-digitating side chains. Each distinct steric zipper interface features different interacting amino acid sidechains. Some of the interfaces feature homotypic interactions, whereas others feature heterotypic interactions. This creates interfaces that have varying values of buried surface area and shape complementarity, as well as different symmetries. These seven interfaces can belong to five classes of steric zippers. Two of these classes were revealed by the x-ray and electron diffraction

structures. The ^{248}LIIKGI^{253} x-ray structure is an antiparallel Class 7 zipper (the two outer faces of the pair of β-sheet are equivalent, and both β-sheets have the same strand edge facing up) (Figure 1.4A). The ^{247}DLIIKGISVHI^{257} microED structure is a parallel Class 2 zipper (the two b-sheets are face to back, again with the same strand edge facing up) (Figure 1.4B). This steric zipper has the greatest surface area buried at 350Å^2 (Figure 1.4B). The cryoEM structure accounts for the other five distinct interfaces that belong to three classes of zippers: 1, 3 and 4 (Figure 1.4C-E). Class 3 zippers had been postulated (sheets face-to-face, with the strand edge facing up in one sheet and down in its mate), but not previously observed; now we find them in our cryoEM structure at the interfaces of ISVHI/LIIKGIS and DLIIKG/KGISVH (Figure 1.4C). These various distinct interfaces are achieved due to the flexibility of the RRMcore backbone, which is displayed as an overlay (Figure 1.4F).

Within the cryoEM structure we see that the β-sheet interfaces exhibit both up-up and up-down backbone directionality (51). This is achieved by a kinked peptide flipping the direction of its backbone at the central isoleucine, amino acid eight of the 11 residue segment (Supplementary Figure 1.9). That is, the carbonyl of the glycine residue faces down, and at the kink, the carbonyl of the adjacent isoleucine also faces down. This creates a strand in which half of the segment exhibits “up” directionality and the other half exhibits “down” directionality. Here, the ^{247}DLIIKGISVHI^{257} strand can have both an up-up interface on one side and an up-down interface on the other side. Because the RRMcore can bend back on itself, the interactions of strands 6’, 2 and 5 illustrate how only three copies of the peptide segment are needed to create two entirely distinct interfaces (Figures 1.5A and 1.5B). This enables lateral growth of the fibril, perpendicular to the fibril axis and suggests how a propagating fibril could be composed of two distinct steric zipper cores within a localized region.

We also analyzed the flexibility of the ^{247}DLIIKGISVHI^{257} backbone. To do this we applied the LSQ superimposition feature from the program COOT on eight distinctly different b-strands of RRMcore in two of our structures (52). We compared the backbone conformations of

the seven full length strands from the cryoEM structure, strands 1-7, and the backbone of the microED structure, strand 10 (Figure 1.4F). When analyzing RMSD values we find that none of the strands perfectly overlay (Supplementary Table 1.3). In the cryoEM structure, we see four kinked segments in the asymmetric unit, all of which have distinct backbones and RMSD values ranging from 0.41 to 0.72 Å. The other four strands of ^{247}DLIIKGISVHI^{257} are more linear in conformation but also all have distinctive backbone angles, with RMSD values ranging from 0.59 to 2.34 Å. We find that the eight strands can be divided into three subclasses: kink, straight, and curve. We represent the eight strands as a pseudo-phylogenic tree to show the structural relationships of the distinct backbones (Figure 1.5C).

The last features we investigated were the possible modes of formation of this complex fibril from 27 distinct strands. To do this we first computed the solvation energies of the surfaces of three possible protofibrils, each of 9 protofilaments, to determine the asymmetric unit that best represents the protofibril (Supplementary Figure 1.10) (53). From three potential collections of nine protofilaments, we found the lowest solvation energy to be 3,460 cal/mol, corresponding to the asymmetric unit displayed in Figure 3 (Supplementary Figure 1.11). A pseudo two-fold axis runs between strands 1, 2, 5 and 8 and 3, 4, 6 and 9 (Supplementary Figure 1.12). To visualize this we have displayed all kinked strands as red, straight strands as blue, partial strands as yellow and curved strands as green (Figure 1.5D). In this representation, one can see how two red strands associate with a blue and a yellow strand to form a core component of the asymmetric unit that is then repeated through two fold symmetry. The green, curved strands serve to stabilize the asymmetric unit or protofibril, resulting in three-fold symmetry of the mature fibril.

The data from all three structures provide us with information for how the RRMcore can initiate aggregation. The ^{247}DLIIKGISVHI^{257} backbone can adopt at least eight different conformations. This allows the segment to associate with seven unique interfaces comprising five different classes of steric zippers. Ultimately the polymorphic features of this region can

offer insight into how core segments of other amyloid proteins may initiate and accelerate fibril formation.

DISCUSSION

Nucleation of complex amyloid fibrils

The combination of the diverse interfaces and the flexibility of the ^{247}DLIIKGISVHI^{257} backbone offers the opportunity to formulate models for the nucleation of amyloid fibrils. We have designed four alternative mechanistic models, supported by the structures in this paper, to explain the seeding and propagation of amyloid fibrils (Figures 1.5E-H). In the first two mechanisms we focus on growth of a single fibril, while the last two mechanisms display both lateral extension and secondary nucleation, respectively. In the first mechanism, two individual β-sheets form independently (Figure 1.5E). They subsequently mate together to form the steric zipper interface. Elongation follows, allowing for growth of the fibril. This is likely how the microED structure and the kinked interfaces of the cryoEM structure associate. In the second mechanism, an individual β-sheet forms first and serves as a seed or template for the additional sheet (Figure 1.5F). This forces a specific backbone conformation on the mating peptide strand and dictates the type of interface that forms. Elongation subsequently follows the initial templating of the new strand. We expect this process to occur in the cryoEM structure. The presence of nine distinct strands makes it unlikely that they all form independently and come together in a perfectly symmetrical manner. Rather, we envision that the most common interface, the kinked interface, forms independently and dictates new interfaces and backbone conformations for additional strands, such as the blue, yellow and green strands. This is explained further below through the lateral growth and secondary nucleation models.

In a third model, we propose how a fibril could grow laterally with two different interfaces (Figure 1.5G). The first fibril forms and serves as a template for an additional sheet on the back side, or non-zipper side, of the templating sheet. This arrangement is found in our fibril

structure at the interactions of strands 2, 5 and 6'. Here, the prime symbol represents the corresponding strand on the adjacent asymmetric unit. Strand 2 could form the initial fibril with strand 5 and then serve as a seed for the growth of strand 6' (Figure 1.5A). In the fourth model, we illustrate how this lateral growth could lead to secondary nucleation (Figure 1.5H). Once lateral fibril growth has occurred, it is possible for one of the sheets to dissociate and create a new steric zipper interface with a different strand. This is represented by the orange and green arrows. From our cryoEM structure, one can envision strand 6' dissociating from the protofilament of strand 2 and forming a new interface with strand 9' (Supplementary Figure 1.11). This creates two independent filaments that were nucleated by one. Each of these models offers one explanation of how amyloid proteins might form polymorphic assemblies in distinct cellular environments.

Packing polymorphism of RRMcore

Understanding how amyloid proteins form polymorphic assemblies is a critical aspect of neurodegenerative research. Here, we present evidence that an amyloid-forming segment can display two distinct types of packing polymorphism: positional and modal, concepts introduced by Caspar and Cohen (54). Positional polymorphism is defined as the different disposition of identical parts within a single assembly. In contrast, modal polymorphism is defined the different dispositions of identical parts in different modes of assembly-the customary usage of the term packing polymorphism in amyloid structures. The structures of RRMcore reported here show both Positional Polymorphism and Modal Polymorphism. Positional Polymorphism, for the first time in amyloid structure research, is evident in the 27-β-sheet fibril, where the seven β-sheets composed of RRMcore assume several distinctly different structures, as discussed above. An additional different RRMcore β-sheet is the straight polymorph of the microED structure (Figure 1.5C,D) which qualifies RRMcore also as exhibiting Modal Polymorphism. This remarkable variation in conformation of the β-sheets formed by RRMcore

leads to prodigious packing polymorphism, which now includes positional, in additional to modal polymorphism.

Although the entire TDP-43 protein cannot be modeled into the 27- sheet fibril determined here by cryoEM, the structure offers an example at the molecular scale of the prodigous polymorphism of which amyloid assemblies are capable, and the great range of structures that amyloid fibrils can form. This extreme polymorphism has implications for research in neurodegeneration. Distinct amyloid polymorphs may seed distinct "strains" of amyloid, which in a disease context can lead to different clinical conditions. Each polymorph may require its own drug. Previous work from our lab has shown that fibril inhibition of one specific amyloid polymorph can inhibit fibril formation, and in some cases toxicity (38, 55, 56). In cases where multiple polymorphs form, inhibiting one structure may not be sufficient to inhibit disease formation or progression. Thus, understanding amyloid polymorphism is likely to be important in the process of drug development against amyloid diseases. We suggest that the structures presented in this paper offer insight not only about TDP-43 aggregation, but also about the challenges presented by amyloid aggregation in general.

MATERIALS AND METHODS

Selection and Generation of TDP-43 Peptide Targets

All targeted peptide segments were chosen based on the Rosetta Energy 3D Profile Method, Zipper DB (43). Following selection of peptide targets ranging in size from 6-11 residues, all peptides were purchased from Genscript at a purity of 95% or higher.

Peptide Fibril Formation

For initial biochemical characterization, all peptide segments were dissolved in phosphate buffered saline (PBS) at pH 7.5. Peptide segments ^{247}DLIIKG^{252}, ^{247}DLIIKGISV^{255}, ^{247}DLIIKGISVH^{256}, ^{248}LIIKGISVH^{256}, ^{248}LIIKGISVHI^{257}, ^{250}KGISVHISN^{259}, ^{252}GISVHI^{257} and ^{252}ISVHISN^{259} were dissolved) to a final concentration of 20mM. Peptide segments ^{248}LIIKGI^{253}

and ^{253}SVHISN^{259} were dissolved to a final concentration of 40mM due to their high solubility. The last peptide segment, ^{247}DLIIKGISVHI^{257}, was dissolved to a final concentration of 1mM due to its low solubility. All peptides were then incubated at 37°C in a Torrey Pine Scientific shaker at level 9 for 1 week. Samples were examined by TEM following fibril formation.

Negative Stain Transmission Electron Microscopy (TEM)

Negatively stained samples for TEM were prepared by applying 5 μL of sample on hydrophilic 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The fibrils were allowed to adhere for 4 min and washed once with 2% uranyl acetate. The samples were then stained for 2 min with 2% uranyl acetate, leaving a few nanoliters of stain on the grid. Grids were allowed to dry for 5 minutes before storage. Each grid was inspected on a T12 (FEI) microscope.

Fibril Denaturation Studies

Peptide fibrils were generated as described above. Untreated and treated samples were then prepared. For untreated samples, 6μL of water was added to 54μL of fibril preparation. For treated samples, 6μL of 20% SDS solution were added to 54μL of fibril preparation for a final 2% SDS concentration. The treated samples were then quiescently incubated at 70°C for 15 minutes. Turbidity experiments were subsequently performed as untreated and treated samples were aliquoted into black 384-well plates (Nunc, Rochester, NY) covered with optical tape. Values were recorded at a wavelength of 340nm in a Spectramax M5 (Molecular Devices, Sunnyvale, CA). Experiments were performed in triplicate. Data was normalized against untreated fibril samples so the starting value was 1.00. Error bars represent standard deviation.

Fibril Diffraction

Fibrils prepared in PBS were spun down at 15,000 rpm for 30 minutes in a tabletop micro-centrifuge and washed with water a total of three times. Fibrils of ^{247}DLIIKGISVHI^{257} prepared in water did not undergo this step. Concentrated fibrils were diluted in 20μL of water. 3 μL of fibril sample was applied to the middle of two glass capillaries for alignment and allowed

to dry overnight. For data collection, the glass capillaries, with fibril sample, were mounted on brass pins. Diffraction was collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA with a wavelength of 0.9792 Å. All patterns were collected at a distance of 300 mm and analyzed using the Adxv software package (57).

Crystallization

^{248}LIIKGI^{253} was dissolved in 10mM LiOH to a final concentration of 18.7 mg/mL. Crystals were grown by hanging drop vapor diffusion over a reservoir containing 0.1M CHES, pH 9.5 and 20% w/v PEG 8000. Each hanging drop contained 1µL of peptide solution mixed with 1µL of reservoir solution. Crystals were mounted in loops and stored in liquid nitrogen before data collection.

^{247}DLIIKGISVHI^{257} was dissolved in CHES buffer at pH 8.5 to a final concentration of 1mM. Crystals were formed by shaking the peptide at 37°C for 1 week in a Torey Pine Scientific Shaker at level 9.

Xray Data Collection and Structure Determination

The x-ray diffraction data for segment ^{248}LIIKGI^{253} was collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA. Detector ADSC HF -4M was used. Data was collected at a wavelength of 0.9792Å, temperature of 100K, at 5° oscillations and 175mm detector distance. Indexing was done using XDS, integration was done with DENZO and data sets from two crystals were merged in P1 and scaled with SCALEPACK to 1.4Å resolution (58). The data was imported using the CCP4 suite (59). Molecular replacement was done using Phaser (60). Initial models of different idealized polyalanine β-strands did not pack correctly in the cell. The final model used was a sheet composed of two anti-parallel strands with the sequence, AIIAGI, based on the structure of VEAVYL. The model was manually adjusted over 13 cycles to the correct sequence and rotamer orientation in Coot (52). The atomic model was refined by Refmac with a final Rfactor/Rfree of 20.7/22.6 (61).

MicroED Data Collection and Structure Determination

The microED data for segment ^{247}DLIIKGISVHI^{257} was collected using the same protocol as Rodriguez et al. with minor differences detailed below (49, 62, 63). Data was collected at HHMI Janelia Research Campus using the Tecnai F20 TEM operating at 200 kV recorded using a bottom mount TVIPS F416 CMOS camera. Individual image frames were exposed for 3s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.3° per second. Diffraction data was collected on 41 individual crystals, but only 13 were used for data processing. Indexing and integration was done using XDS and data sets from 13 crystals were merged in P1 and scaled with SCALEPACK to 1.4Å resolution (64). The data was imported using the CCP4 suite (59). Molecular replacement was done using Phaser (59). An idealized 11 residue β-strand model of DLIIKGISVHI was used for molecular replacement. The model was further refined in refmac, accounting for twinning operations and adjusting rotamer orientations in Coot (52, 61). The twinning operator is –H, K, K-L with a fraction of 0.5. The final model had an Rfactor/Rfree of 26.2/30.6

Cryo electron microscopy (cryoEM) and reconstruction.

^{247}DLIIKGISVHI^{257} peptide was dissolved in water to a final concentration of 1mM. The sample was shaken for 1 week at 37°C in a Torrey Pine shaker on level 9. The fibril sample was subsequently spun down at 15,000 rpm for 30 min. The supernatant, with contaminant monomer, was disposed of and the fibril sample was resuspended in 20µL of water.
Each aliquot of 2.5 µL of the fibril sample was applied onto a “baked” (65) Quantifoil 1.2/1.3 µm, 200 mesh grid, blotted for 4 seconds at force 1, then vitrified in liquid ethane at liquid nitrogen temperature in a Vitrobot Mark IV (FEI). CryoEM data were collected in an FEI Titan Krios microscope equipped with a Gatan Quantum LS / K2 Summit direct electron camera system (operated with a 300kV acceleration voltage and a slit width 20 eV). Data collection was in counting mode driven by Leginon automation software package (66). Defocus values were controlled with Leginon by applying a single 3.0 µm target defocus. Dose-fractionation movies

were recorded at a frame duration of 200 ms and total duration of 10 s. The total dosage was 60 e/A^{-2}, with each frame 1.2 e/A^{-2} (6 $e/A^{-2}s$) [measured on Digital Micrograph (Gatan) software]. Frames were aligned as described (67), except that an iterative alignment scheme as previously described (68) was implemented after the original alignment method. We used the summary of all frames to determine defocus and particle locations, and 3rd to 20th frames for data processing. The focal parameters of the data were determined by CTFFIND4 (69). We selectively included images within a defocus range of 2.5 – 4 μm.

D11I fibrils were manually marked in EMAN (70) *helixboxer*. A total of 1,904 individual filaments were selected. These filaments were segmented by 90% overlap scheme into 33,758 boxes of 288x288 pixels. A series of Relion (71) 2D classifications were done to the dataset for the following purposes: 1. all boxes were subjected to a general (tau_fudge = 2) 2D classifications to eliminate bad particles, and to separate different morphologies (See Text for morphological distributions). We chose the group of fibrils with the most abundant morphology (20,138 boxes) for the next run (2. tau_fudge = 4) which was to further eliminate bad particles and misclassified particles. The second 2D classification resulted in 18,887 boxes. 3. We used a last 2D classification with a very high tau_fudge factor (16) to extend the working resolution to beyond 4.8 Å, making it possible to catch the layer line corresponding to the helical arrangement of β-strands.

A total of 18,818 boxes were included in the 3D analyses. We use EMAN based IHRSR[72] to reconstruct the 3D structure of the fibril. The resolution was slightly better than 4 Å, making it possible to trace the backbone and register the 11-peptide sequence. We built an atomic model for an asymmetric unit (with 9 strands) with Coot (52). We built 18 helically related copies of the asymmetric unit to account for the helix. We did a preliminary Fourier space pseudo-crystallographic model refinement (72). We then carried out another model refinement with the phenix.real_space_refine command of the Phenix package (73) using default settings.

We further refined the 3D model of the fibril in Relion (with IHRSR implementation (74) by Class3D with one class. (Due to the fine feature of the β-strands, Refine3D was unable to adequately regulate working resolution to refine with high enough resolution factors.) We tested the resolution by even-odd division of the dataset according to movie numbers. The FSC crosses 0.5 at 3.8 Å and 0.143 at 3.2 Å. The local resolution was also accessed by ResMap (75) to be 3.5 Å. The side chain features of the map is consistent with a resolution of 3.5-3.6 Å. We further refined the atomic model in Phenix with phenix.real_space_refine command with 30 asymmetric units.

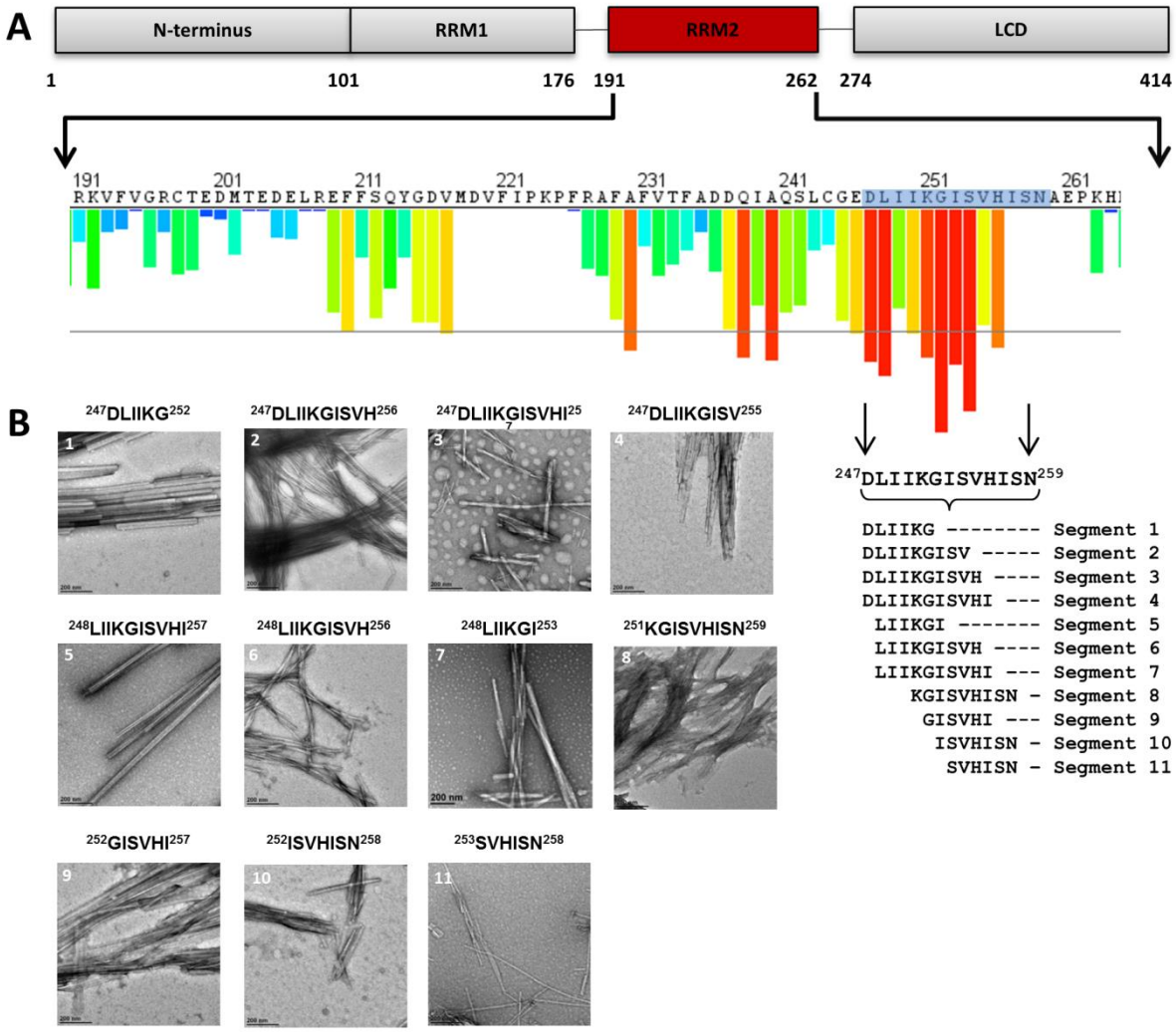
Analysis of Structural Features

All images displaying the interfaces of the RRM2 structures were created in Pymol (76). Area buried and shape complementarity were calculated based on established protocols (77, 78). Backbone RMSD values were determined using the LSQ superimpose feature in COOT (52). The pseudo-phylogeny tree comparing these RMSD value was generated using the T-REX interface (79) at URL: http://www.trex.uqam.ca/.

ACKNOWLEDGEMENTS

We thank Todd Yeates, Feng Guo, James Bowie, Qin Cao, Pascal Krotee, Smriti Sangwan, Paul Seidler and Lorena Saelices for insightful discussions. We thank the Advanced Photon Source (APS) staff at Northeastern Collaborative Access Team beamline 24-ID-E for support during data collection; specifically M. Capel, K. Rajashan- kar, N. Sukumar, J. Schuermann, I. Kourinov and F. Murphy. The APS is funded by National Institutes of Health grants P41 RR015301 and P41 GM103403 and Department of Energy under Contract Number DE-AC02-06CH11357. We also thank Michael Collazo and the UCLA-DOE Macro-Molecular Crystallization Core Technology Center for setting up crystallization screens as well as Ivo Atanasov and the Electron Imaging Center for NanoMachines (EICN) of California NanoSystems Institute (CNSI) at UCLA for support on the use of electron microscopes. The

UCLA-DOE Macromolecular Crystallization Core Technology Center is partially funded by the Department of Energy grant DE-FC0302ER63421. We thank the DOE, Howard Hughes Medical Institute (HHMI), and NIH grant R01-AG029430 for support to the Eisenberg Lab. Elizabeth Guenther was funded by the UCLA CMB Training Program (Ruth L. Kirschstein NIH GM00785).


A
N-terminus
RRM1
RRM2
LCD
1
101
176
191
262
274
414
191 201 211 221 231 241 251 261
RKVFVGRCTEDMTEDELREFFSQYGDVMDVFIPKPFRAFAFVTFADDQIAQSLCGEDLIIKGISVHISNAEPKH
B
247DLIIKG252
247DLIIKGISVH256
247DLIIKGISVHI25
247DLIIKGISV255
247DLIIKGISVHISN259
DLIIKG -------- Segment 1
DLIIKGISV ----- Segment 2
DLIIKGISVH ---- Segment 3
DLIIKGISVHI --- Segment 4
LIIKGI ------- Segment 5
LIIKGISVH ---- Segment 6
LIIKGISVHI --- Segment 7
KGISVHISN - Segment 8
GISVHI --- Segment 9
ISVHISN - Segment 10
SVHISN - Segment 11
248LIIKGISVHI257
248LIIKGISVH256
248LIIKGI253
251KGISVHISN259
252GISVHI257
252ISVHISN258
253SVHISN258
200 nm

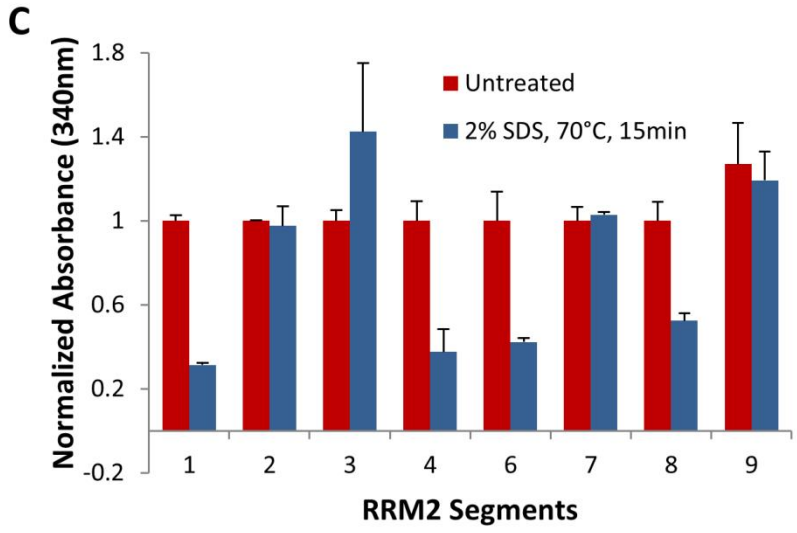

C
Normalized Absorbance (340nm)
1.8
1.4
1
0.6
0.2
-0.2
Untreated
2% SDS, 70°C, 15min
1 2 3 4 6 7 8 9
RRM2 Segments

Figure 1.1: Segments from the RRM2 domain of TDP-43 exhibit characteristic amyloid properties. (A) Schematic of TDP-43 domains showing the sequence of RRM2 and amyloid-promoting segments predicted by ZipperDB (red histogram bars) (43). Four of the top ten potential zipper-forming peptide segments lie between residues 247-259, the RRMcore. The eleven segments chosen for further characterization are displayed below the ZipperDB predictions. (B) Negative stain EM images of 5mM fibril samples of segments in the ^{247}DLIIKGISVHISN^{259} region grown in PBS. Segments range in length from six to eleven residues and span the entire 13-residue RRMcore. Fibril morphology differs for many of the segments, indicating the potential for polymorphic assembly in this region. (C) Segments from the RRM2 exhibit moderate to high stability under heat and SDS denaturation. Fibril samples were first grown by shaking at 37°C at 20mM concentration in PBS, pH 7.5. Following fibril formation, samples were treated with 2% SDS and heated to 70°C for 15 minutes. The data are normalized with untreated samples at an initial value of 1. Error bars represent standard deviation.

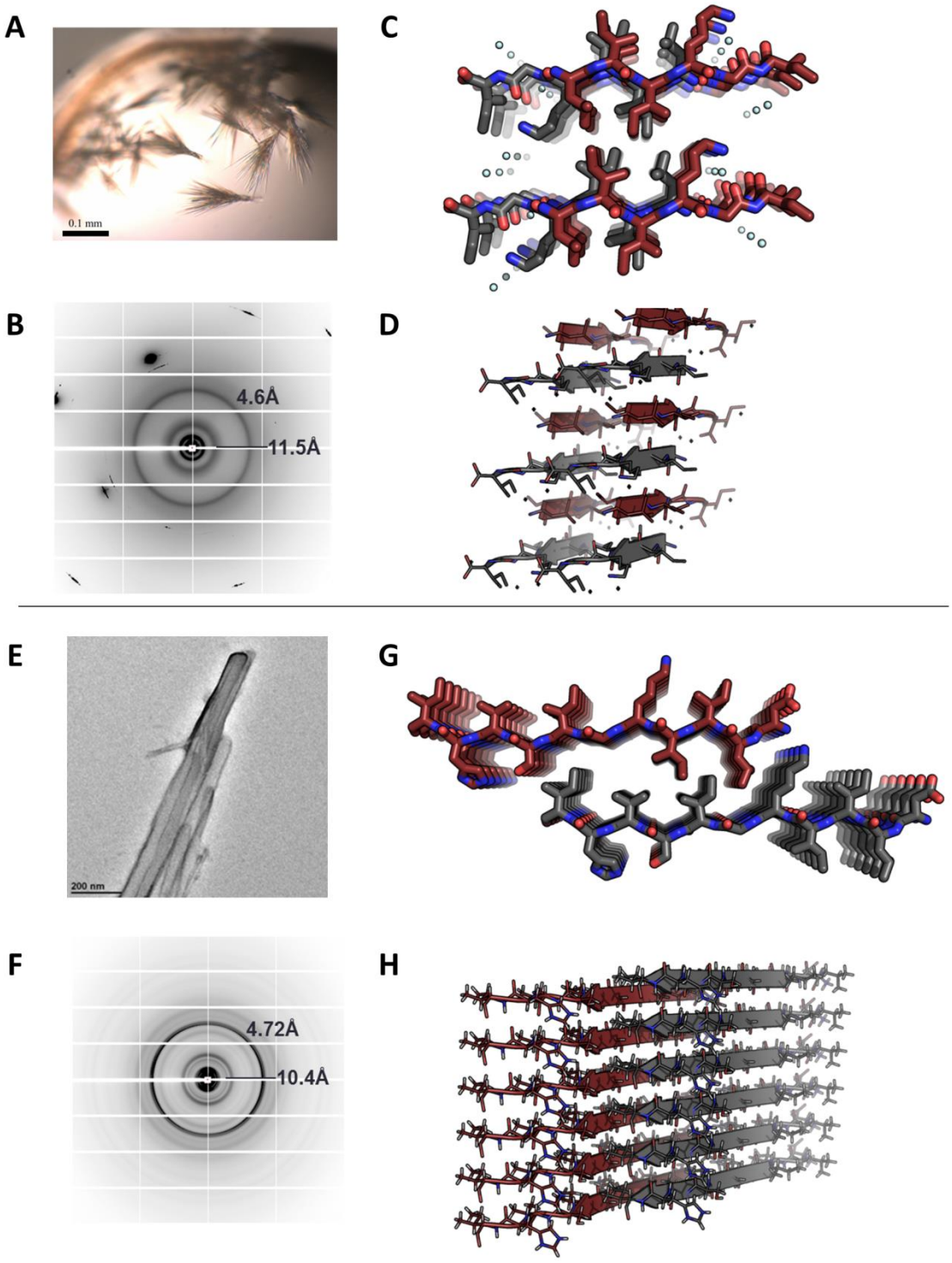


Figure 1.2: Atomic structures of ^{248}LIIKGI^{253} and ^{247}DLIIKGISVHI^{257}, determined by diffraction methods, illustrate different steric zipper interfaces (A) Microcrystals of ^{248}LIIKGI^{253} were grown in hanging drop experiments at a pH of 8.5. Crystals were visible by

light microscopy. (B) Fibril diffraction of the ^{248}LIIKGI^{253} sample exhibits reflections consisted with the characteristic cross-β pattern with 4.6Å and 11.5Å diffraction rings. (C) The structure of ^{248}LIIKGI^{253} was determined from x-ray diffraction data taken at the Advanced Photon Source. The segment forms an anti-parallel, face=back (Class 7) steric zipper interface (Sawaya et al. 2007), viewed here down the fibril axis. (D) View of ^{248}LIIKGI^{253}, perpendicular to the fibril axis, illustrates the anti-parallel stacking of the strands. (E) Nanorystals of the ^{247}DLIIKGISVHI^{257} were grown in a shaker at 37°C at a pH of 8.5. Crystals were invisible by light microscopy and successfully visualized using an electron microscope. (F) Fibril diffraction of the ^{247}DLIIKGISVHI^{257} segment exhibits reflections consistent with the characteristic cross-β pattern with 4.7Å and 10.4Å diffraction rings. (G) The structure of ^{247}DLIIKGISVHI^{257} was determined by micro-electron diffraction at HHMI Janelia Research Campus. The segments form a parallel face-to-back Class 2 steric zipper interface viewed here down the fibril axis. (H) Orthogonal view of the ^{247}DLIIKGISVHI^{257} structure illustrates the stacking of β-strands.

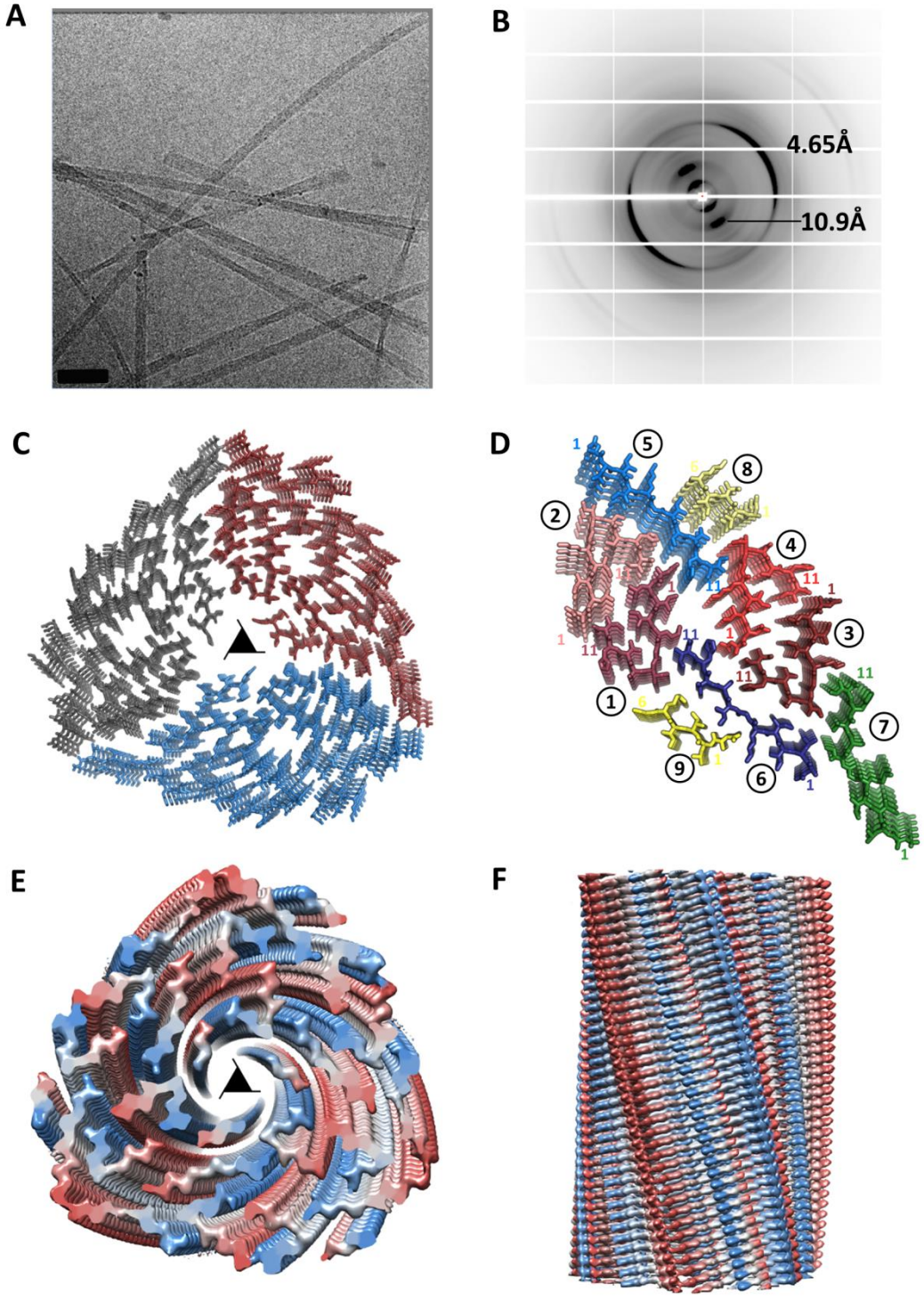


Figure 1.3: Cryo EM Structure of the ^{247}DLIIKGISVHI^{257} fibril illustrates a 27-filament, polymorphic 3-start helix. (A) CryoEM image of the ^{247}DLIIKGISVHI^{257} fibril sample grown in water at 1mM. Sample was frozen in liquid ethane. Scale bar is 60nm. (B) X-

ray fibril diffraction of the ^{247}DLIIKGISVHI^{257} sample grown in water exhibits a characteristic cross-β pattern with 4.65Å and 10.9Å reflections. (C) The structure of the ^{247}DLIIKGISVHI^{257} fibril is described as a 3-start helix with a left-handed twist. The fibril is composed of 27 β-sheets running the full length of the fibril, with 9 β-sheets per asymmetric unit, related by three-fold symmetry. This model displays the fibril looking down the fibril axis, through the center pore of the helical fibril. The three asymmetric units are colored gray, blue and red. The three fold screw axis is represented by the triangle. (D) The asymmetric unit of this structure is composed of nine β-sheets, each exhibiting a left-handed twist moving down the fibril axis. Each sheet is labelled by a number 1-9. The β-sheets are classified by their relative conformation; kinked = red, straight = blue, curve = green and partial = yellow. Each strand is represented by a unique color within its designated family with the first and last amino acid residues indicated by number in corresponding color. Two of the β-sheets, at the center and on the periphery of the helix, show density for only the first six residues: ^{247}DLIIKG^{252}. (E) Top view of the density of the 3-start helix viewed down the fibril axis. (F) Density of the fibril, viewed perpendicular to the fibril axis.

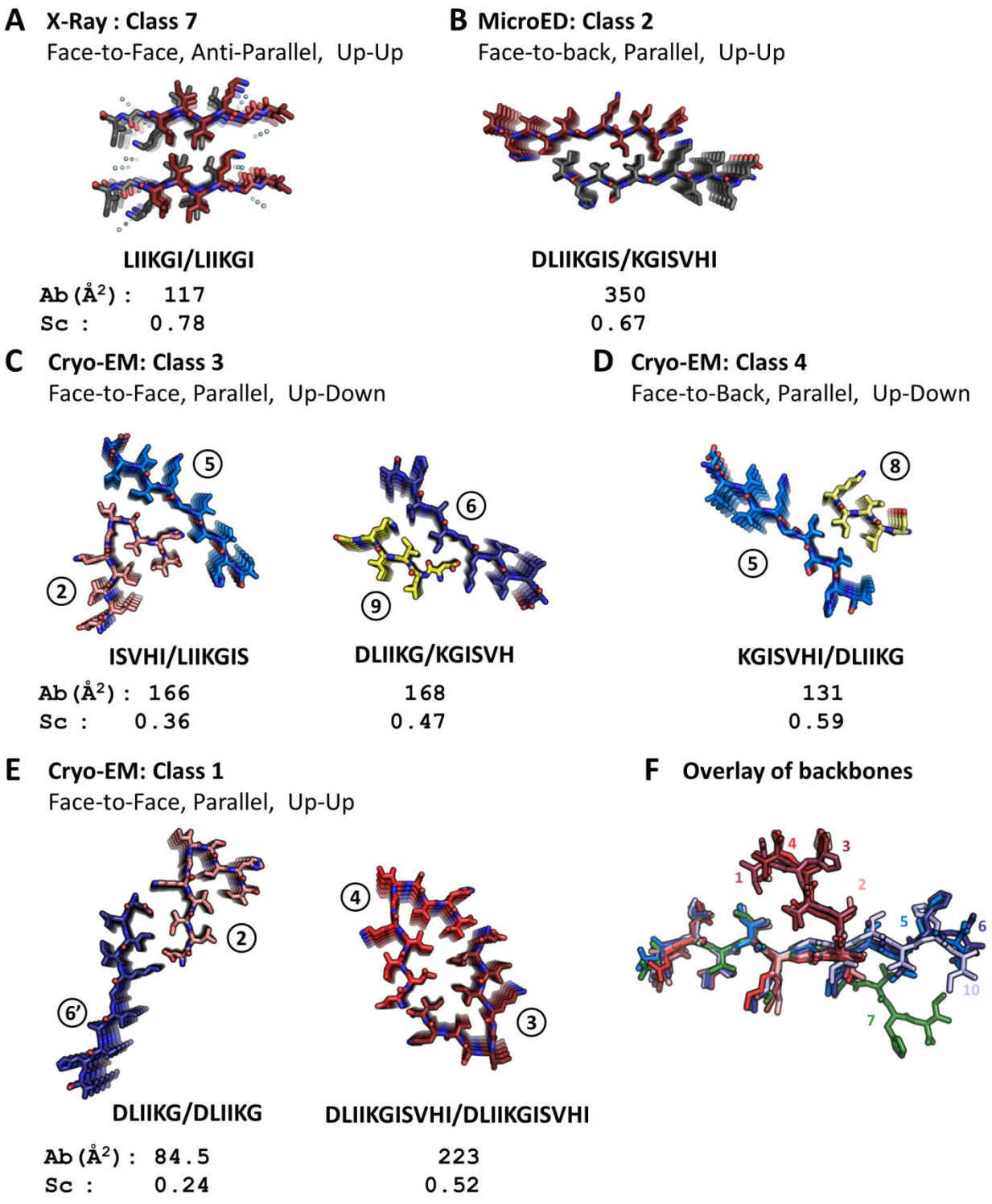
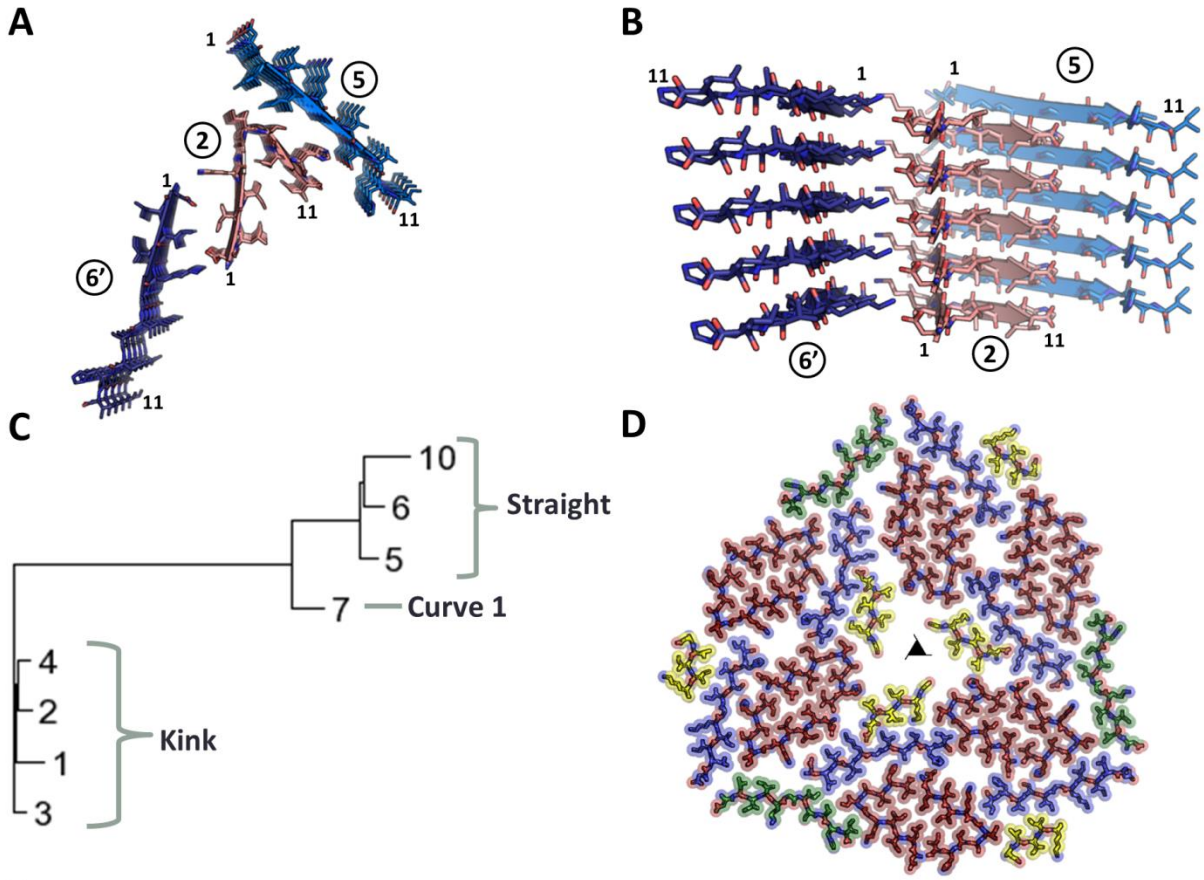
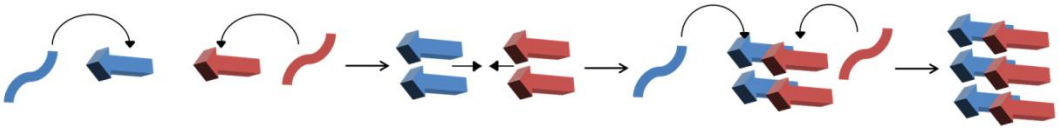


Figure 1.4: **The three atomic structures reported here display seven unique sheet-to-sheet interfaces, representing five classes of amyloid steric zippers** (9) The buried solvent-accessible surface area (Ab) and steric complementarity (Sc) are shown for each

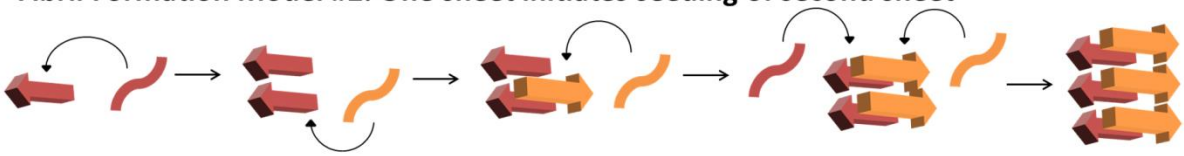
interface. (A) The ^{248}LIIKGI^{253} structure, determined by x-ray diffraction, is the only example from all seven interfaces of an anti-parallel zipper. This Class 7 structure exhibits the highest shape complementarity at 0.78. (B) The ^{247}DLIIKGISVHI^{257} electron diffraction structure illustrates a Class 2 steric-zipper interface. This structure exhibits the tightest packing of all DLIIKGISSVHI interfaces, with a buried surface area of 350Å^2. (C) Two interfaces from the ^{247}DLIIKGISVHI^{257} cryoEM structure exhibit Class 3 morphology. These are the first two examples of Class 3 zippers determined at atomic resolution. (D) The third interface in the ^{247}DLIIKGISVHI^{257} cryoEM structure is a Class 4 steric zipper formed at the periphery of the 3-start helix. (E) Two interfaces from the ^{247}DLIIKGISVHI^{257} cryoEM structure exhibit Class 1 morphology. The kinked interface is one of the most distorted interfaces found to date for amyloid steric zippers. The segment β-strand appears to pack against itself as well as against a mating sheet. The DLIIKG/DLIIKG interface represents the weakest interface with the smallest area buried 97Å^2 and the poorest shape complementarity at 0.17. (F) The backbones of all full RRMcore strands have been superimposed for residues LIIK to illustrate their differences. Only the 7 sheets formed from full length ^{247}DLIIKGISVHI^{257} strands of the cryoEM structure are assigned numbers. Strand 10 is from the microED structure.



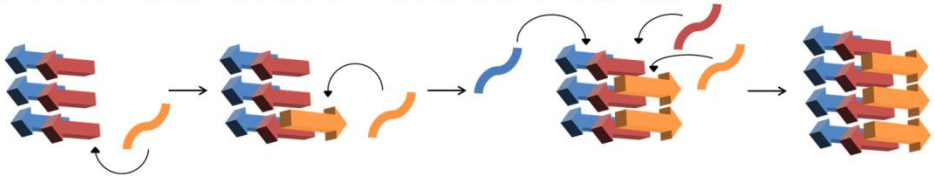
E **Fibril Formation Model #1: Individual sheets self-seed**



F **Fibril Formation Model #2: One sheet initiates seeding of second sheet**



G **Fibril Formation Model #3: Lateral Fibril Extension**



H **Fibril Formation Model #4: Secondary nucleation**

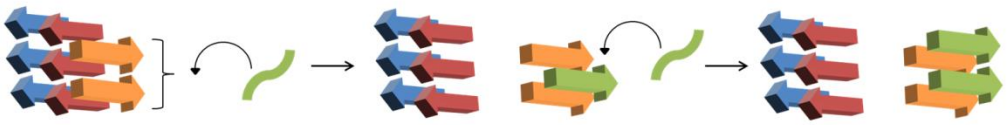


Figure 1.5: The conformational flexibility of the ^{247}DLIIKGISVHI^{257} backbone provides insight into amyloid fibril formation and polymorphism (A) Enlargement looking down the fibril of two distinct steric-zipper interfaces formed by β-sheets 2, 5 and 6' (left to right). Here the ' symbol indicates the corresponding strand on the adjacent asymmetric unit (i.e 6' is strand 6 on the adjacent asymmetric unit moving counterclockwise, or up the fibril axis). This arrangement of DLIIKG/DLIIKG and ISVHI/LIIKGIS interfaces is possible because the ^{247}DLIIKGISVHI^{257} segment kinks on itself. The N- and C- termini of the segment are marked by the numbers 1 and 11 respectively. (B) View perpendicular to the fibril axis of the arrangement of Panel A, showing 6 of the 15,000+ layers of the fibril (C) The structural relationships of the eight distinct β-sheet backbones illustrated by a pseudo-phylogenetic tree. The two principal branches are kinked and linear sheets. The linear branch forks into two sub-branches: straight and curve. (D) A view down the fibril axis of one layer of the fibril represented by sticks and spheres. The strands are colored by relative conformation: kink = red, straight = blue, curve = green and partial = yellow. (E) A model of fibril nucleation in which individual sheets form before mating to form a fibril. Each sheet is represented by blue or red. The schematic illustrates that the monomer of the amyloid protein is added along the hydrogen bonding axis followed by the mating of two distinct sheets. Elongation of the fibril follows the mating step. (F) A second model of fibril formation in which one sheet nucleates the second sheet. This schematic illustrates how one sheet may form first and then dictates the conformation a monomer of the amyloid protein will nucleate the mating sheet. The mating sheet continues to grow and the fibril is elongated. (G) A third model of fibril formation in which lateral extension of the fibril is shown. The schematic illustrates that an additional, distinct fibril interface (red/orange) can form off of a preexisting fibril (red/blue). (H) A fourth model in which secondary nucleation is displayed. The schematic illustrates how once the lateral extension occurs from model 3, the orange strand can dissociate and form a distinct and

independent filament with a new strand, green. Now two independent filaments exist from the seed of one fibril.

Table 1.1: Statistics of x-ray diffraction data collection and atomic refinement

Sample	LIIKGI
Data Collection	
Beam Line	APS NECAT 24-ID-E
Detector	ADSC HF-4M
Date	3/6/15
PDB accession code	
Space group	P1
Unit cell dimensions	
a,b,c (Å)	11.54, 9.59, 21.18
α,β,γ(°)	95.63, 98.21, 76.50
Reflections observed	7208
Unique reflections	1561
Wavelength (Å)	0.9792
Resolution (Å) [a]	20.91-1.40(1.45-1.40)
R_{sym} (%) [b]	12.7(36.1)
CC(1/2)	0.979 (0.879)
I/σ	11.5(5.05)
Completeness (%)	94.5(94.6)
Wilson B value (Å^2)	9.57
Refinement	
Resolution (Å)	20.91 – 1.40
Resolution (Å) last shell	1.43-1.40
Reflections in Working Set	1241
Reflections in Test Set	160
R_{work}(%)/Rfree(%)[c]	20.7(27.2)/22.6(37.1)
Protein Molecules in Asymmetric Unit	2
Number of non-H atoms	
Protein	92
Non-protein	6
RMS deviations	
Bond lengths (Å)	0.006
Bond angles (°)	2.087
Average *B*-factor (Å^2)	9.95
Protein atoms	9.55
Non-protein atoms	16.14
Ramachandran plot regions [d]	
Favored	100
Allowed	0
Outliers	0

[a] Highest resolution shell shown in parenthesis.

[b] $Rsym = 100\,x\sum (I - I(mean))^2 / \sum I^2$

where *I* is the observed intensity of the reflection HKL and the sum is taken over all reflections HKL.

[c] $Rfactor = 100\,x\sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \sum \left| F_{obs} \right|$

F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement

[d]Percentage of residues in Ramachandran plot regions were determined using Molprobity

Table 1.2: Statistics of microED data collection and atomic refinement

Sample	DLIIKGISVHI
Excitation Voltage (kV)	200
Wavelength (Å)	0.0251
Total dose per crystal (e/ Å^2)	3.4
Frame rate (frame/s)	0.265
Rotation rate (°/s)	0.3
# crystals used	13
Total angular rotation collected (°)	38.4-77.6*
Merging Statistics	
Space group	P1
Unit cell dimensions	
a, b, c (A°)	24.81, 4.73, 15.83
α, β, γ (°)	80.88, 86.37, 89.78
Resolution (Å) [a]	15.60-1.39 (1.43-1.39)
Rmerge[b]	20.3 (145.1)
# of reflections	9353
Unique reflections	1037
Completeness	73.4 (62.1)
Multiplicity	9.0 (6.7)
I/sigma	3.34 (0.92)
$CC_{1/2}$ (Diederichs, 2013)	99.2 (33.3)
Refinement Statistics	
Reflections in working set	944
Reflections in test set	92
R_{work}	26.2 (40.9)
R_{free}	30.6 (53.5)
RMSD bonds (Å)	0.009
RMSD angles (°)	1.211
Ramachandran (%)[d]	
Favored	100
Allowed	0
Outliers	0
PDB ID code	
EMDB ID code	

[a] Highest resolution shell shown in parenthesis.

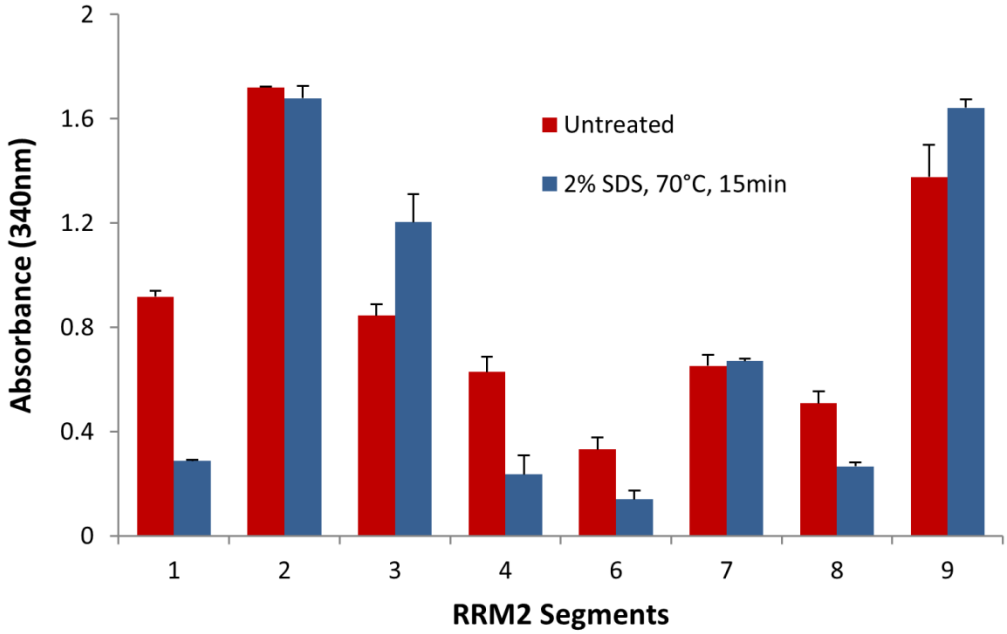
[b] $Rsym = 100\,x\sum (I - I(mean))^2 / \sum I^2$

where I is the observed intensity of the reflection HKL and the sum is taken over all reflections HKL.

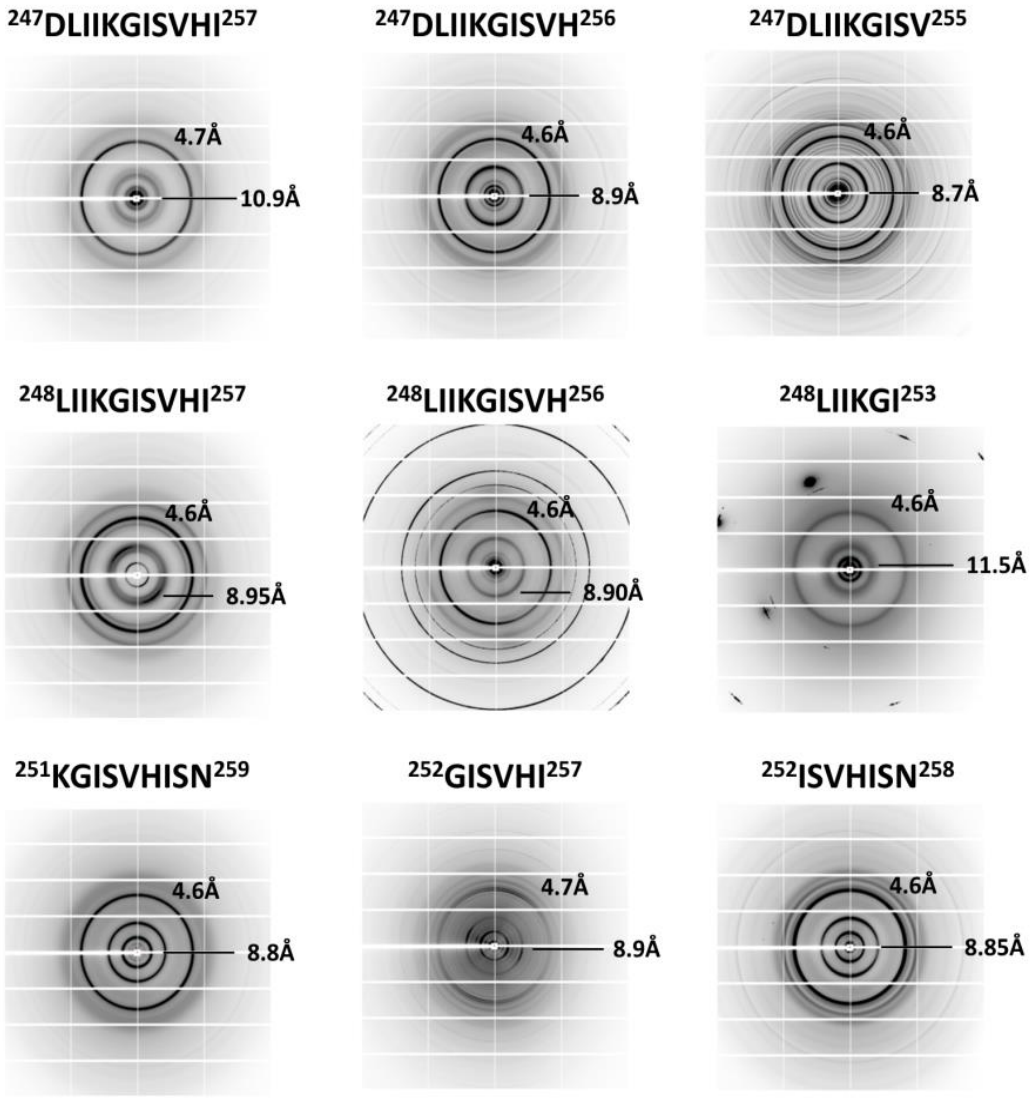
[c] $Rfactor = 100\,x\sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \sum \left| F_{obs} \right|$

F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement

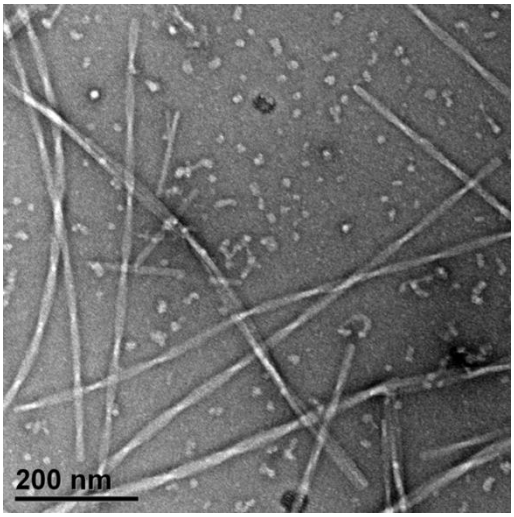
[d]Percentage of residues in Ramachandran plot regions were determined using Molprobity



Supplementary Figure 1.1: RRM2 segments demonstrate high to moderate stability against treatment with SDS and heat. Fibril samples were first grown by shaking at 37°C at 20mM concentration in PBS, pH 7.5. Following fibril formation, samples were treated with 2% SDS and heated to 70°C for 15 minutes. The data shows the raw absorbance values as monitored by the Spectramax.

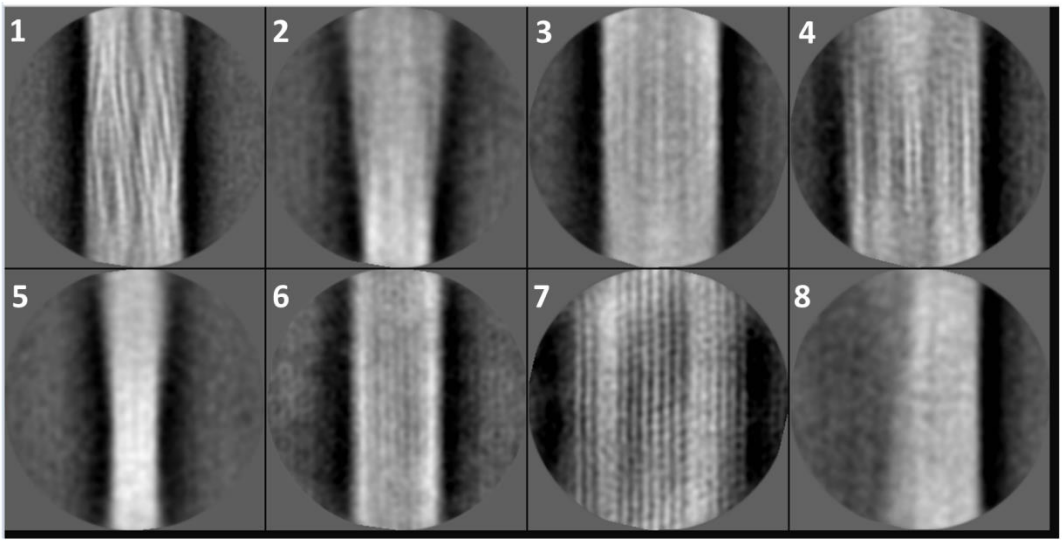


Supplementary Figure 1.2: Fibril diffraction of the RRM2 fibril segments shows classic amyloid diffraction. Fibril diffraction was collected at the 5um beam at the Advanced Photon Source. All nine samples exhibit the classic 4.6-4.8Å ring indicative of the stacking of β-strands and the 8-12Å ring indicative of mating sheets.



Supplementary Figure 1.3: Negative stain EM image of ^{247}DLIIKGISVHI^{257} cryo-EM sample. ^{247}DLIIKGISVHI^{257} fibrils were grown by shaking for one week at 1mM in water. Sample was stained with 2% uranyl acetate and visualized on the FEI T12 transmission electron microscope. Fibril samples illustrate multiple morphologies, including twisted helical assemblies.

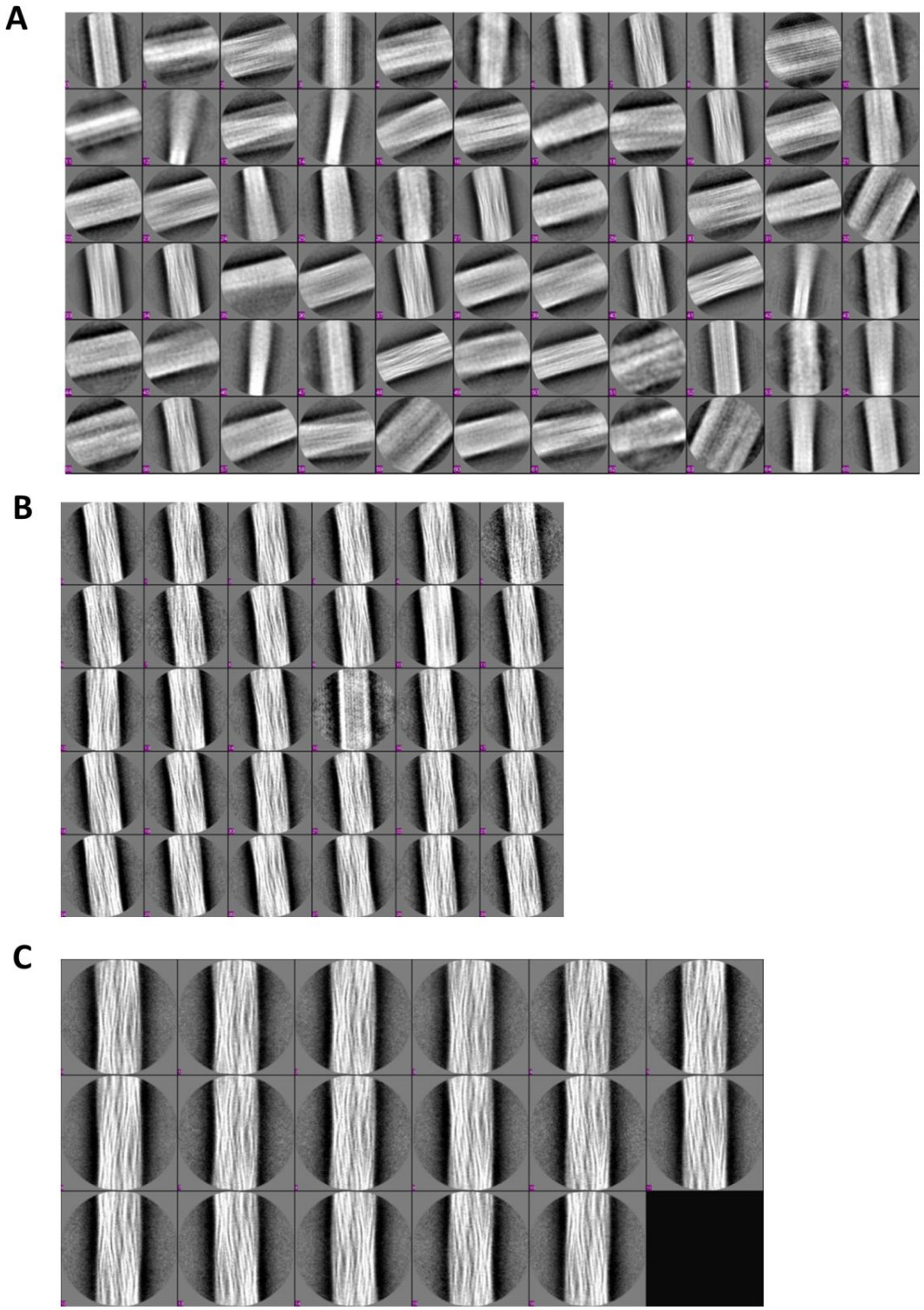
A



B

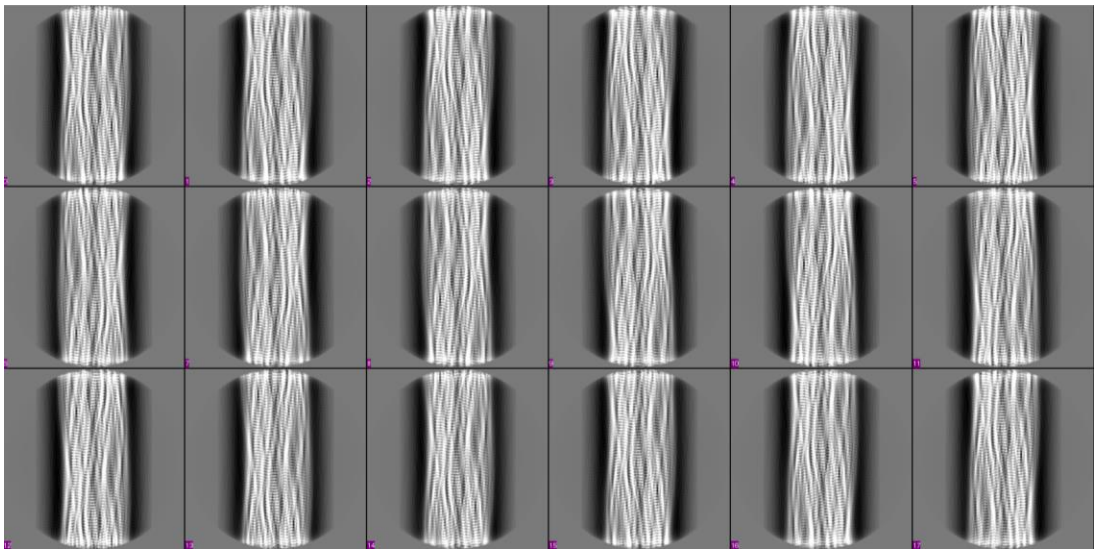
	Morphology	Abundance (%)
1	3-Start Helix	59.7
2	Wide Twist	5.3
3	Cylinder #1	8.4
4	Cylinder #2	4.4
5	Narrow Twist	2.9
6	Cylinder #3	0.6
7	Cylinder #4	0.5
8	Sheet	1.0

Supplementary Figure 1.4: Eight unique fibril morphologies are observed in the ^{247}DLIIKGISVHI^{257} cryo-EM sample. (A) Illustration of the eight different types of fibrils observed by cryo-EM. The eight fibrils demonstrate four different conformations of amyloid fibrils: helices, twists, cylinders, and sheets. (B) The relative abundance of the eight unique fibrils are displayed as a table. The 3-start helix represents the most abundant species, comprising approximately 60% of the sample.

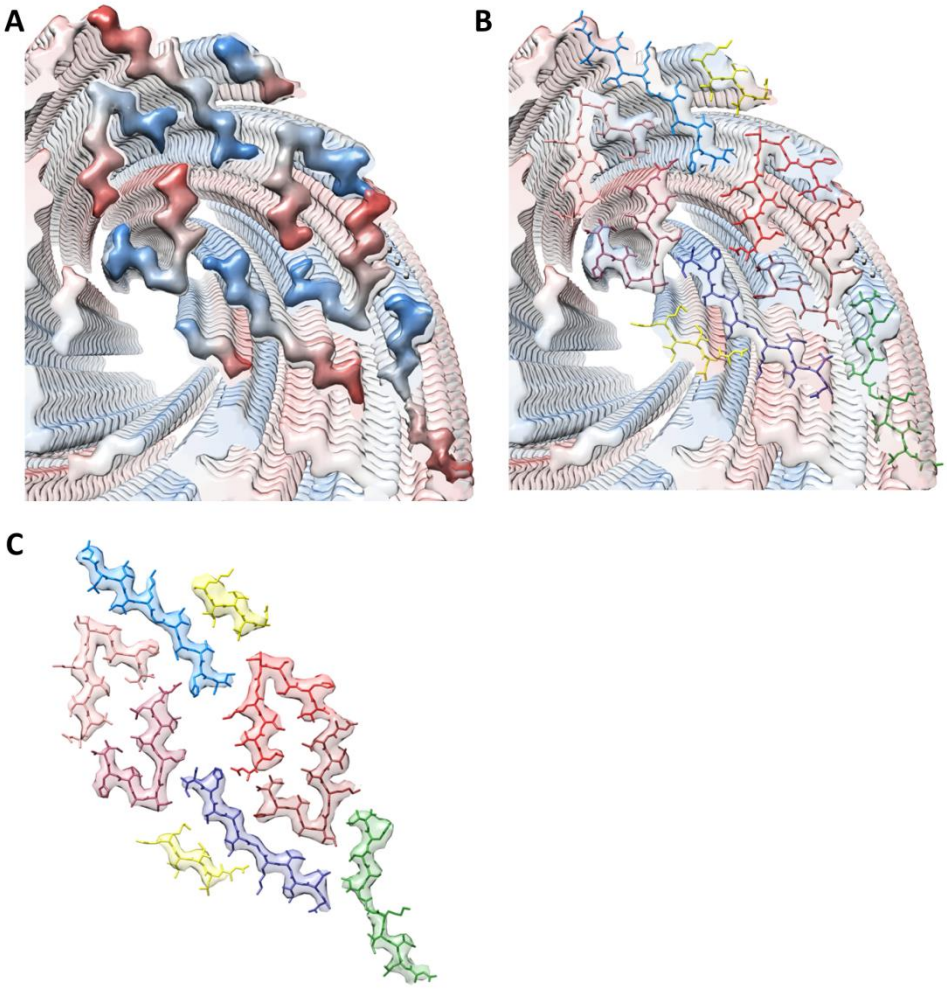


Supplementary Figure 1.5: Classification of images for the ^{247}DLIIKGISVHI^{257} cryo-EM sample. (A) Round 1 classification illustrates the polymorphic assemblies observed

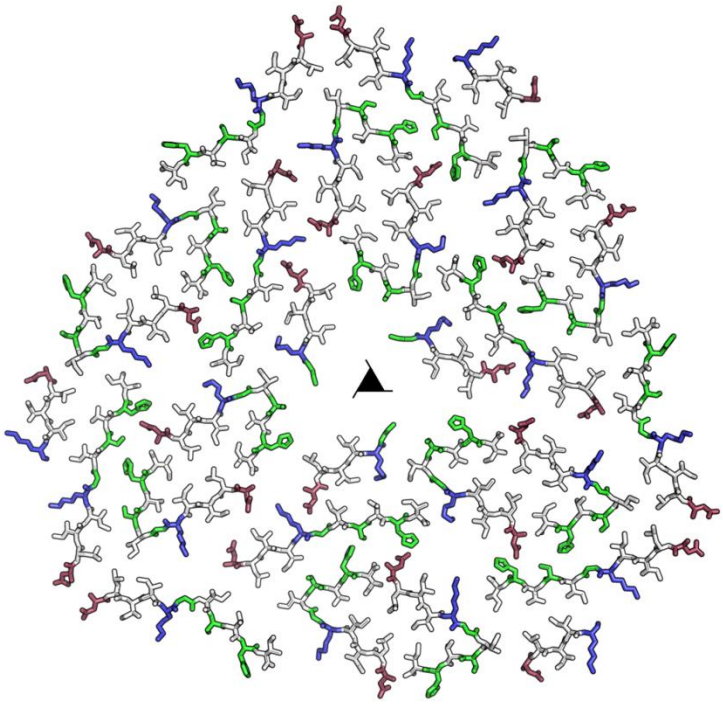
in the fibril sample. (B and C) Round 2 and 3 class averages focus on fibrils exhibiting the 3-start helix morphology. Each panel illustrates a unique face of the fibril as it twists along the EM grid. 5-7Å features of the structure become apparent during these rounds of class averaging.



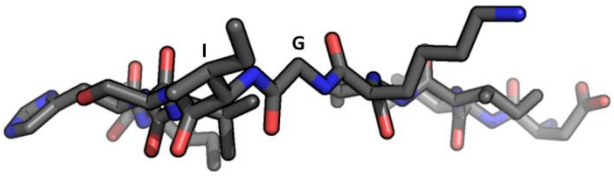
Supplementary Figure 1.6: 2D Projection of the final 3D model of the ^{247}DLIIKGISVHI^{257} cryo-EM fibril structure. The panels illustrate a side view of the fibril, perpendicular to the fibril axis. Each panel represents an 8° azimuthal rotation of the fibril. The hydrogen bonding network is visualized by the stacking of β-strands.



Supplementary Figure 1.7:The asymmetric unit of the 3-start helix is composed of 9 strands (A) The electron density of the asymmetric unit is displayed. Here, the N-terminus is displayed in red and the C-terminus is displayed in blue. (B,C) The peptide strands are displayed within the electron density. The β-sheets are classified by their relative conformation; kinked = red, straight = blue, curve = green and partial = yellow. Each strand is represented by a unique color within its designated family.

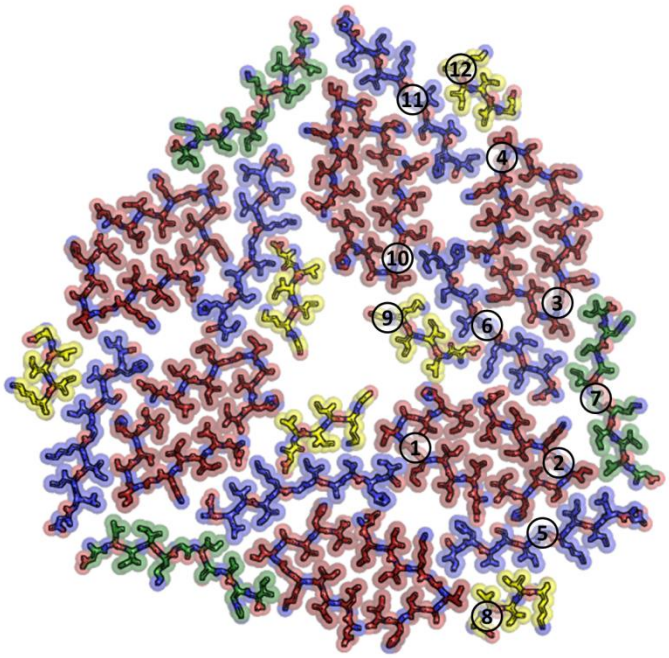


Supplementary Figure 1.8: The protofilaments of the 3-start helix fill space, exclude water, and bury hydrophobic surfaces. Here, glycines and apolar residues are white, polar residues are green, positively charged residues are blue, and negatively charged residues are red. The three fold screw axis is indicated by the triangle.



Supplementary Figure 1.9: The ^{247}DLIIKGISVHI^{257} kinked backbone illustrates the flipping of the backbone orientation. The figure below illustrates the kinking of the ^{247}DLIIKGISVHI^{257} backbone allows for the carbonyl of the glycine and isoleucine to both face down. The glycine and isoleucine residues have been marked by one letter abbreviations.

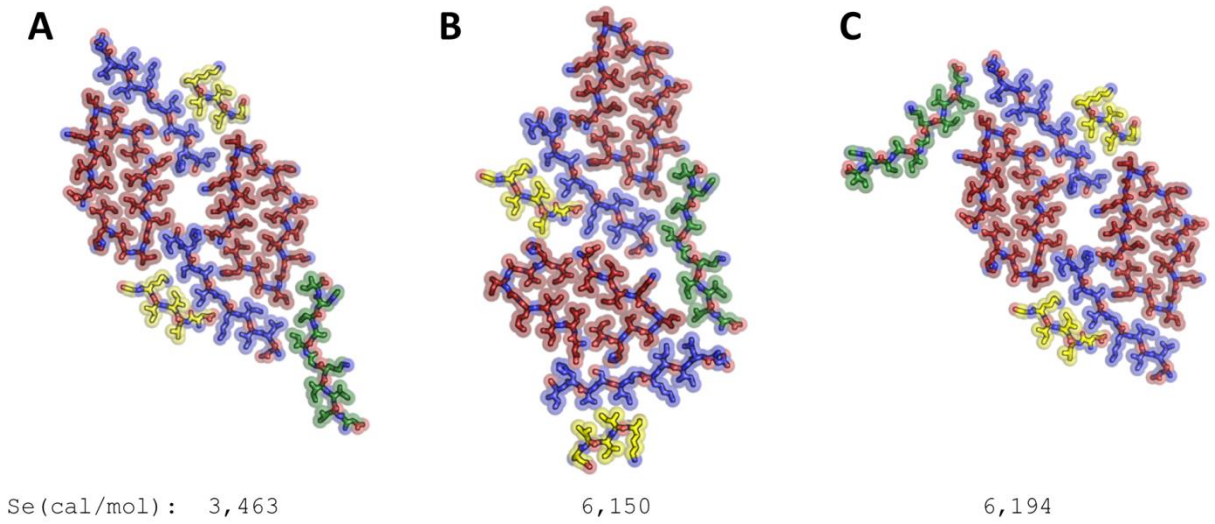
A



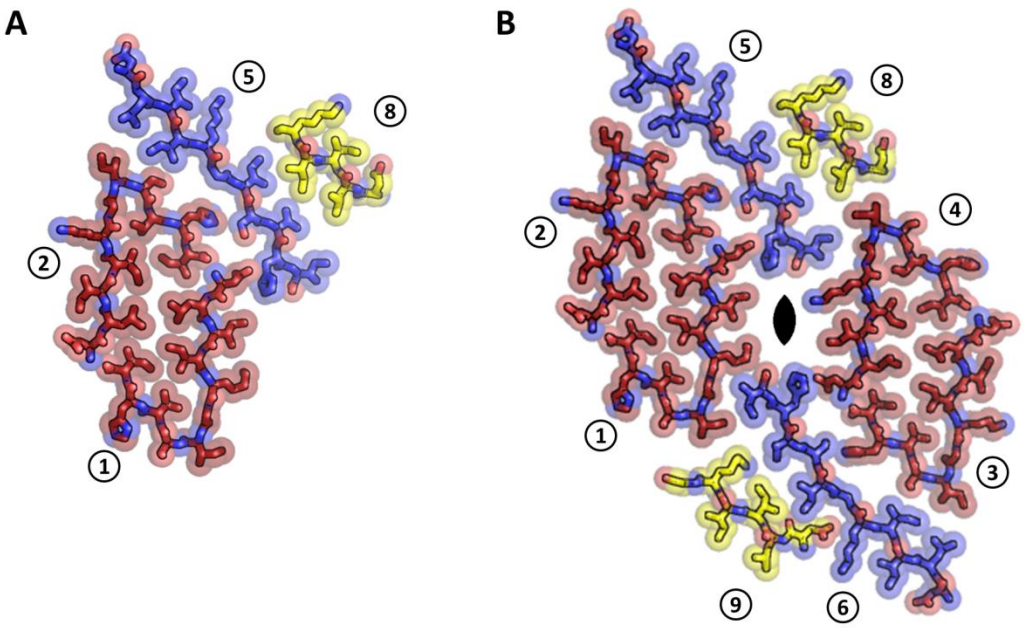
B

Strand	1	2	4	6	7	8	10	11
1		1241						
3			1336	1238	1591			
4				56				870
5	140	1737			354	1571		
6		188			881			
7		752						
9	437			1359			496	
10				949				151
12			761					

Supplementary Figure 1.10: Solvation energies of all unique interfaces. (A) One layer of the 27 strand fibril is illustrated with 12 strands labelled. The interactions of these strands account for all interfaces in the full fibril structure. (B) The solvation energy values in cal/mol are included for all 19 distinct interactions that occur in this structure. Not all interactions are steric zippers.



Supplementary Figure 1.11: Solvation energies for three possible asymmetric units. (A-C) The figure displays the solvation energies in cal/mol for the three possible combinations of the nine strands in an asymmetric unit. The organization in Panel A resulted in the lowest solvation energy, 3,463 cal/mol.



Supplementary Figure1.12: Pseudo-two fold symmetry is observed in the asymmetric unit (A-B) Strands 1, 2, 5 and 8 form a complex which exhibits two fold symmetry with strands 3, 4, 6 and 9. The oval shape indicates two fold symmetry.

Segment	Sequence	# of Charges	Net Charge	GRAVY
1	DLIIKG	4	0	0.83
2	DLIIKGISV	4	0	1.43
3	DLIIKGISVH	5	1	0.97
4	DLIIKGISVHI	5	1	1.29
5	LIIKGI	3	1	2.17
6	LIIKGISVH	4	2	1.47
7	LIIKGISVHI	4	2	1.77
8	KGISVHISN	4	2	0.07
9	GISVHI	3	1	1.47
10	ISVHISN	3	1	0.70
11	SVHISN	3	1	0.07

Supplementary Table 1.1: Charge and hydropathicity values for the RRM2 segments. Total number of charges and net charges are displayed for fibrils prepared in PBS at pH 7.5. GRAVY value represents the average of hydropathy values as calculated with EXPASY Protparam.

Interface	Segment	Ab (Å²)	Average Ab (Å²)	Sc
ISVHI/LIIKGIS	2	175	166	0.36
	5	157		
DLIIKG/KGISVH	9	164	168	0.47
	6	172		
KGISVHI/DLIIKG	5	127	131	0.59
	8	135		
DLIIKG/DLIIKG	6’	87	85	0.24
	2	82		
D11I (kink)	4	226	223	0.52
	3	220		

Supplementary Table 1.2: Area buried and shape complementarity values for the 5 interfaces formed in the cryo-EM, 3-start helix structure. Area buried is represented by Ab and shape complementarity is represented by Sc.

Strand	1	2	3	4	5	6	7	10
1	0.00							
2	0.67	0.00						
3	0.64	0.50	0.00					
4	0.72	0.41	0.47	0.00				
5	6.38	6.20	6.08	6.16	0.00			
6	6.54	6.34	6.23	6.29	0.59	0.00		
7	5.60	5.36	5.25	5.29	2.00	2.00	0.00	
10	6.96	6.76	6.65	6.71	1.13	1.06	2.34	0.00

Supplementary Table 1.3: RMSD values the ^{247}DLIIKGISVHI^{257} backbone illustrates that all strands are unique. The RMSD values in Å^{2}illustrate that none of the backbones overlay perfectly. Strands 1-4 represent the kinked morphologies while strands 5-8 represent the linear morphologies.

REFERENCES

1. Chiti, F. & Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).

2. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297,** 353–356 (2002).

3. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15,** 384–96 (2014).

4. Greenwald, J. & Riek, R. Biology of amyloid: Structure, function, and regulation. *Structure* **18,** 1244–1260 (2010).

5. Goldsbury, C., Frey, P., Olivieri, V., Aebi, U. & Müller, S. A. Multiple assembly pathways underlie amyloid-?? fibril polymorphisms. *J. Mol. Biol.* **352,** 282–298 (2005).

6. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci.* **105,** 18349–18354 (2008).

7. Wiltzius, J. J. W. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16,** 973–8 (2009).

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

9. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–7 (2007).

10. Stroud, J. C. The zipper groups of the amyloid state of proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 540–545 (2013).

11. Tuttle, M. D. *et al.* Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat. Struct. Mol. Biol.* **23,** 1–9 (2016).

12. 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,** E4976-84 (2016).

13. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic A??42 Amyloid Fibrils. *J. Am. Chem. Soc.* **138,** 9663–9674 (2016).

14. Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16742–7 (2002).

15. Schmidt, A., Annamalai, K., Schmidt, M., Grigorieff, N. & Fändrich, M. Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril. *Proc. Natl. Acad. Sci.* **113,** 201522282 (2016).

16. Schmidt, M. *et al.* Peptide dimer structure in an Aβ(1-42) fibril visualized with cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 11858–11863 (2015).

17. Bai, X. C. *et al.* An atomic structure of human γ-secretase. *Nature* **525,** 212–7 (2015).

18. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. Conformational variations in an infectious protein determine prion strain differences. *Nature* **428,** 323–8 (2004).

19. Riek, R. & Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature* **539,** 227–235 (2016).

20. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148,** 1188–1203 (2012).

21. Krotee, P. *et al.* Atomic Structures of Fibrillar Segments of hIAPP Suggest Tightly Mated β-Sheets Are Important for Cytotoxicity. *Elife* 1–26 (2016). doi:10.7554/eLife.19273

22. Soriaga, A. B., Sangwan, S., MacDonald, R., Sawaya, M. R. & Eisenberg, D. Crystal structures of IAPP amyloidogenic segments reveal a novel packing motif of out-of-register beta sheets. *J. Phys. Chem. B* **120,** 5810–5816 (2016).

23. Colletier, J.-P. *et al.* Molecular basis for amyloid-beta polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 16938–43 (2011).

24. Manuela Neumann1, 11 *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science (80-.).* (2006). doi:10.1126/science.1134108

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

26. Schwab, C., Arai, T., Hasegawa, M., Yu, S. & Mcgeer, P. L. Colocalization of Transactivation-Responsive DNA-Binding Protein 43 and Huntingtin in Inclusions of Huntington Disease. *J Neuropathol Exp Neurol* **67,** 1159–1165 (2008).

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

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

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

30. Kraemer, B. C. *et al.* Loss of Murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol.* **119,** 409–419 (2010).

31. Sephton, C. F. *et al.* TDP-43 is a developmentally regulated protein essential for early embryonic development. *J. Biol. Chem.* **285,** 6826–6834 (2010).

32. Wu, L.-S. S. *et al.* TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *Genesis* **48,** n/a-n/a (2016).

33. Chiang, P. *et al.* Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16320–16324 (2010).

34. Zhu, L. *et al.* An ALS-mutant TDP-43 neurotoxic peptide adopts an anti-parallel ??-structure and induces TDP-43 redistribution. *Hum. Mol. Genet.* **23,** 6863–6877 (2014).

35. Igaz, L. M. *et al.* Expression of TDP-43 C-terminal fragments in vitro recapitulates

pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* **284,** 8516–8524 (2009).

36. Iguchi, Y. *et al.* TDP-43 Depletion Induces Neuronal Cell Damage through Dysregulation of Rho Family GTPases. *J. Biol. Chem.* **284,** 22059–22066 (2009).

37. Di Carlo, V. *et al.* TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation. *Mol. Neurobiol.* **48,** 952–963 (2013).

38. Soragni, A. *et al.* A Designed Inhibitor of p53 Aggregation Rescues p53 Tumor Suppression in Ovarian Carcinomas. *Cancer Cell* **29,** 90–103 (2016).

39. Nonaka, T., Kametani, F., Arai, T., Akiyama, H. & Hasegawa, M. Truncation and pathogenic mutations facilitate the formation of intracellular aggregates of TDP-43. *Hum. Mol. Genet.* **18,** 3353–3364 (2009).

40. Li, Q., Yokoshi, M., Okada, H. & Kawahara, Y. The cleavage pattern of TDP-43 determines its rate of clearance and cytotoxicity. *Nat. Commun.* **6,** 6183 (2015).

41. Wang, Y. T. *et al.* The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* **288,** 9049–9057 (2013).

42. Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R. & Zitzewitz, J. A. Folding of the RNA Recognition Motif (RRM) domains of the Amyotrophic Lateral Sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem.* **289,** 8264–8276 (2014).

43. 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,** 3487–92 (2010).

44. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

45. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149,** 753–767 (2012).

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

47. Liu, S. *et al.* Atomic resolution structure determination by the cryo-EM method MicroED. *Protein Sci.* **26,** 8–15 (2016).

48. de la Cruz, M. J. *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14,** 399–402 (2017).

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

50. Sept, D., Baker, N. a & McCammon, J. A. The physical basis of microtubule structure and stability. *Protein Sci.* **12,** 2257–2261 (2003).

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

52. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

53. Eisenberg, D., McLachlan, A. D., Eisenberg, D., McLachlan, A. D. & McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **319,** 199–203 (1986).

54. Caspar, D. L. D. & Cohen, C. in *Nobel Symposium 11. Symmetry and Function of Biological Systems at the Macromolecular Level* 393–414 (1969).

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

56. Saelices, L. *et al.* Uncovering the mechanism of aggregation of human transthyretin. *J. Biol. Chem.* **290,** 28932–28943 (2015).

57. Arvai, A. No Title. *Adxv - A Program to Display X-ray Diffraction Images*. http://www.scripps.edu/tainer/arvai/adxv.html (2015).

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

59. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67,** 235–242 (2011).

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

61. Vagin, A. A. *et al. REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2184–2195 (2004).

62. Hattne, J. *et al.* MicroED data collection and processing. *Acta Crystallogr. Sect. A Found. Adv.* **71,** 353–360 (2015).

63. Shi, D. *et al.* The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11,** 895–904 (2016).

64. Kabasch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr* **26,** 795– 800 (1993).

65. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).

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

67. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).

68. Banerjee, S. *et al.* 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science (80-.).* **351,** 871–875 (2016).

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

70. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated Software for High-

Resolution Single-Particle Reconstructions. *J. Struct. Biol.* **128,** 82–97 (1999).

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

72. Ge, P. & Zhou, Z. H. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proc. Natl. Acad. Sci.* **108,** 9637–9642 (2011).

73. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

74. Clemens, D. L., Ge, P., Lee, B. Y., Horwitz, M. A. & Zhou, Z. H. Atomic structure of T6SS reveals interlaced array essential to function. *Cell* **160,** 940–951 (2015).

75. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2013).

76. DeLano, W. L. The PyMOL Molecular Graphics System. *Schrödinger LLC wwwpymolorg* **Version 1.,** http://www.pymol.org (2002).

77. Lawrence, M. & Colman, P. Shape complemntarity at protein/protein interfaces. *J Mol Biol* **234,** 956–950 (1993).

78. Anonymous. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* **50,** 760–763 (1994).

79. Alix, B., Boubacar, D. A. & Vladimir, M. T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40,** 573–579 (2012).

CHAPTER 2

Atomic Structures of segments within the Low Complexity Domain of TDP-43 Offer Insight into its Dual Functions of Reversible and Pathogenic Aggregation

INTRODUCTION

Over the last several decades, there has been immense interest in understanding the role of low complexity domains (LCDs) in proteins linked to neurodegenerative diseases. These LCDs are regions of proteins that exhibit biased compositions of amino acids and frequently exhibit no secondary structure (1-3). In some cases the LCDs are the result of nucleotide expansions that cause duplications of short sequences, as in Huntington's Disease (4-6). In other cases the LCDs are intrinsic to the protein and believed to have a functional role as in the RNA binding proteins: TAR DNA binding protein 43 (TDP-43), Fused in sarcoma (FUS) and hnRNPA1 (7). Recent findings have illustrated that the LCDs of these RNA-binding proteins play a critical role in a process known as stress granule formation (8-11).

Stress granules (SGs) are aggregates of RNA and RNA binding proteins that form in the cytoplasm when the cell is experiencing some sort of stress, whether it is physical, mechanical or chemical (12-14). This process is often initiated by a halt in translation and puts the cell in a hibernating state until the stress can be remedied (15). These aggregates are reversible, meaning that when the stress is relieved, phosphorylation of SG-proteins result in disaggregation (10, 16, 17). The proteins and RNA are then free to perform their normal functions in the cell. Over the last decade, researchers have studied how this process occurs and what biophysical properties these aggregates exhibit. Recent findings have demonstrated that these SGs exist in a liquid-like state, although their mechanism of assembly and the role of RNA are still being investigated (10, 18-20). Studies on FUS and hnRNPA1 have demonstrated that the LCDs are sufficient for liquid-liquid phase separation (LLPS) (10, 18). Other studies

involving the stress granule protein, G3BP-1, have shown that the SGs exist in a biphasic state with the mRNA packed at the center and the RNA-binding proteins on the periphery (21). Because this research is in its infancy, questions remain about the structures of the aggregates and the structural aspects of stress granule assembly.

In addition to studies aimed at determining the structure of stress granule assemblies, researchers have focused on the role of SG-associated proteins in neurodegenerative diseases. One of the major hallmarks of neurodegenerative diseases, is the deposition of proteinaceous fibrils, or amyloids, in cells (22-24). Different from stress granule assembly, aggregation of these proteins into the amyloid state in disease seems a virtually irreversible process (25, 26). Variants found in TDP-43, FUS and hnRNPA-1 have all been linked to Amyotrophic Lateral Sclerosis, ALS, among other diseases (27). Of these proteins, TDP-43 stands out for four key reasons. First, TDP-43 has over 40 variants that are linked to ALS and other neurodegenerative diseases (27-29). Second, the majority of the variants that cause disease are localized to the LCD (30, 31). Third, the majority of the aggregates found in the cytoplasm of cells of patients are cleavage products inclusive of the LCD (32-34). Finally, TDP-43 is found to aggregate in 97% of sporadic cases and 45% of familial case of ALS (35, 36). This makes TDP-43 an intriguing candidate for understanding how the LCD participates in both reversible, stress granule assembly and irreversible, pathogenic assembly.

TAR DNA Binding Protein (TDP-43) is 43kD protein composed of four domains: an N-terminal domain, two RNA recognition motifs and a C-terminal low complexity domain (LCD) (37-39). This protein is involved in a number of cellular functions including mRNA shuttling, transcription, miRNA processing and alternative splicing among others (35, 40-42). Over the past decade, research into TDP-43 has found aggregates of the protein in patients with ALS, Alzheimer's Disease, Parkinson's Disease, Frontotemporal Lobar Degeneration and Huntington's Disease(43-46). These aggregates are irreversible, ubiquitinated, hyper-phosporylated, C-terminal cleavage products ranging in size from 25-35kD (47). Thus it is

hypothesized that the LCD of TDP-43 plays a role in both this irreversible disease aggregation and the reversible stress granule aggregation.

Through structure determination, we can learn which segments of TDP-43 may drive aggregation and how familial variants affect this process. Here we use x-ray diffraction and micro-electron diffraction (microED) to determine the atomic resolution structures of six peptide segments, four in the reference sequence and two variants, from the C-terminus, or LCD, of TDP-43. We show that the LCD is capable of forming typical steric-zipper β-sheet structures characteristic of pathogenic aggregation. However, we also show that a key segment, 312-317, forms a kinked, reversible structure characteristic of proteins in stress granules that is perturbed by the familial mutants A315T/E.

RESULTS

Three segments from the LCD of TDP-43 form steric zippers

We targeted segments throughout the entire C-terminus because there is not a consensus of which region is the amyloidogenic core. The LCD of TDP-43 is an unstructured domain composed of approximately 140 amino acid residues. Researchers have demonstrated that this region plays a role in aberrant mislocalization and aggregation of TDP-43 to the cytoplasm in mice and neuroblastoma cell lines (48-50). It was initially suggested that the entire LCD, residues 277-414, served as a prion like domain and was the minimum fragment required for aggregation (51). However, additional studies suggested smaller regions within the LCD are sufficient for aggregation and neurodegeneration. In the first study, segment 286-331 was identified and shown to form amyloid fibrils and confer neurotoxicity on primary cortical neurons (52). A second study identified residues 311-360, illustrating through NMR and CD studies that this segment undergoes a helix to β-sheet transition that initiates aggregation and cytoplasmic inclusion (53, 54). A third study suggested that segment 342-366 is the minimum toxic core that transitions from a random coil to a β-sheet (55). Ultimately, these studies

illustrated that there are several regions of the LCD that can be responsible for amyloid-like aggregation.

Previous findings have demonstrated that steric-zipper adhesive segments of amyloid fibrils, are composed of segments ranging from 6-11 residues (56, 57). For this reason we decided to target such peptide segments in the C-terminus for structural determination. Because the C-terminus of TDP-43 is a prion-like sequence we focused our efforts on regions of the LCD that contain glutamine and asparagine residues which are found to be important for prion aggregation (58, 59). We applied ZipperDB, (URL: https://services.mbi.ucla.edu/zipperdb) to identify segments with adhesive energies more than 22 kcal/mol-per segment (60). With these criteria, we targeted over ten, 6-11 residue segments in the C-terminus and were able to determine the structure of three segments: ^{300}GNNQGSN^{306}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402} (Fig. 1A-D, Table 1). All three structures form steric zippers which are composed of parallel, in-register β-sheets. Segments ^{300}GNNQGSN^{306} and ^{370}GNNSYS^{375} both form Class 1 steric zippers in which the β-sheets are face-to-face with the same edges of the sheet oriented up the fibril (Fig. 1B) (57). Both structures have two distinct interfaces, as commonly observed in steric zippers structures. Segment ^{300}GNNQGSN^{306} has a hydrated interface with a buried surface area of 212Å^2 and a dry interface with a buried surface area of 184Å^2 (Fig. 1B). The interfaces of segment ^{370}GNNSYS^{375} are both dry with buried surface areas of 253Å^2 and 166Å^2 (Fig. 1C). Segment ^{396}GFNGGFG^{402} forms a Class 4 steric zipper with face-to-back β-sheets and up-down orientation (Fig. 1D) (57). Because the structure is face-to-back there is only one unique interface which has a buried surface area of 222Å^2.

In addition to determining the structure of these segments, we also wanted to confirm that they exhibit typical amyloid characteristics. We aggregated the samples by shaking and observed as expected that all three segments form fibrils by electron microscopy (Fig. 1F). Because amyloid fibrils have been shown to be resistant to treatment with heat and SDS, we also subjected the fibrils to a 2% SDS solution at 70°C for 15 minutes (17). We found that fibrils of

^{370}GNNSYS^{375} and ^{396}GFNGGFG^{402} were both resistant to this treatment as determined by turbidity. This indicates a strong steric zipper interaction (Fig. 1E). The biochemical properties of these segments, along with the structures that we determined provide the first atomic resolution information on segments of the LCD of TDP-43 that may be responsible for pathogenic aggregation of the protein.

Segment 312-317 of the LCD forms a kinked structure that is disrupted by the A315T/E familial mutants

To further explore the role of the LCD in the aggregation of TDP-43, we focused on segments that have the potential to contribute to stress granule formation. The McKnight laboratory discovered that glycine residues flanked by aromatic side chains are features of hydrogel formation of the LCD of the protein FUS (17). The hydrogels formed by the LCD of FUS were re-solubilized by treatment with SDS and heat, unlike pathogenic amyloid fibrils. Deletion of tyrosine residues in the LCD disrupted hydrogel formation. This led to the hypothesis that aromatic residues play a role in phase separation and can be important for the LLPS found during stress granule assembly. For this reason we focused our structural and biochemical studies on a segment of the LCD, ^{312}NFGAFS^{317}, that contains a glycine residue and two proximal aromatic phenylalanines. Of additional interest is that two ALS familial variants are located in this region: A315T and A315E (52, 61). Both of these variants have been shown to increase redistribution of TDP-43 to the cytoplasm and increase neuronal death as compared to the reference sequence (62-65). We proceeded to biochemically and structurally characterize segments ^{312}NFGAFS^{317}, ^{312}NFGTFS^{317}, and ^{312}NFGEFS^{317}.

For biochemical characterization we first investigated the segments' ability to form fibrils and the stability of those fibrils under denaturing conditions. All three peptide segments formed fibrils when shaken in phosphate-buffered saline (PBS) as observed by electron microscopy (Fig. 2A). Next we tested the stability of the fibrils when treated with 2% SDS at 70°C for 15 minutes (Fig. 2B). We found that the reference sequence, ^{312}NFGAFS^{317}, was sensitive to this treatment

and that the fibrils dissociated as observed by turbidity. Because we targeted this sequence due to its potential role in LLPS and reversible aggregation, this finding provided initial support for our hypothesis that this sequence and aromatics may be important for reversible aggregation of TDP-43. The mutant segment ^{312}NFGTFS^{317} was also sensitive to the treatment and resulted in fibril dissociation, however the ^{312}NFGEFS^{317} segment was stable under denaturing conditions (Fig. 2B). This shows that the A315E variant changed the native structure and biochemical properties of the reference ^{312}NFGAFS^{317} segment.

Following biochemical characterization, we determined the structures of all three segments, revealing that each sequence forms a distinct structure. The reference ^{312}NFGAFS^{317} segment forms a kinked structure, similar to conformations from the LCDs of FUS and hnRNPA1 (Fig. 3B, Table 2) (66). Phenylalanine-5, the residue at the kink, has Ramachandran angles of a left-handed α-helix. The kink places phenylalanine at position 5 on the same side of the sheet as phenylalanine at position 2. In contrast, a canonical β-sheet places the odd numbered sidechains on one face and the even numbered sidechains on the opposite face (56). In the ^{312}NFGAFS^{317} structure, residues 2, 4, and 5 are all on the same face. The subsequent packing of mating strands creates two different interfaces. Here we see that all of the aromatic residues pack near each other creating a hydrophobic, aromatic interface. The other interface is held together by the interactions of the asparagine side chains. This novel kinked structure provides us with an atomic-resolution insight into how aromatics can create a stable, but also labile fibril. We have termed structures that adopt similar kinked backbones, LARKS (Low complexity, Amyloid-like, Reversible, Kinked Segments).

Following structural characterization of the ^{312}NFGAFS^{317} segment, we determined the structure of the ^{312}NFGTFS^{317} variant, which exhibits similar susceptibility to denaturation as ^{312}NFGAFS^{317} (Fig. 3C). The ^{312}NFGTFS^{317} structure is similar to the ^{312}NFGAFS^{317} structure in having an aromatic, hydrophobic interface, however there are three significant differences (Table 2). First, the ^{312}NFGTFS^{317} segment is kinked in a different manner. In ^{312}NFGTFS^{317} the

peptide kinks at the central glycine, residue 3, which has Ramachandran angles of a left-handed α-helix similar to the phenylalanine in the ^{312}NFGAFS^{317} structure. The phenylalanine sidechains are again on the same side of the β-sheet, however this is because residues 2, 3, and 5 face the same direction. The second difference is the hydration of the ^{312}NFGTFS^{317} structure. In the ^{312}NFGAFS^{317} structure there are no water molecules, however, in the ^{312}NFGTFS^{317} structure the non-aromatic, second interface is heavily hydrated (Fig. 3B). Lastly, the buried surface area differs greatly between the structures. While ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317} have an aromatic interface with similar buried surface area values (185Å^2 and 190Å^2 respectively), the buried surface area of interface 2 is very different. Interface 2 in the ^{312}NFGAFS^{317} structure has an area buried of 233 Å^2, while the area buried in ^{312}NFGTFS^{317} interface 2 is 169Å^2 (Fig. 3B, C). The different manner in which ^{312}NFGTFS^{317} kinks creates a less tight interface that is hydrated, likely making a fibril that is more labile than the ^{312}NFGAFS^{317} fibril. Therefore, the A315T variant may induce TDP-43 aggregation through destabilization of the LARKS structure.

The third such structure we determined was the ^{312}NFGEFS^{317} segment, which, unlike ^{312}NFGAFS^{317} demonstrated high stability against denaturing conditions. The backbone of ^{312}NFGEFS^{317} also kinks, however in this case the kink is at the glutamate, residue 4 (Fig. 4D, Table 3). This kink again positions phenylalanines 2 and 5 on the same face of the β-strand. However, despite the presence of a kink, the packing of the strands is significantly different than in ^{312}NFGAFS^{317}. Here the aromatic residues of adjacent strands do not pack together. Instead, there is a face-to-back packing that creates one interface with a buried surface area of 243Å^2. This buried surface area offers some insight into why the ^{312}NFGEFS^{317} fibril is more stable than ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317} fibrils. Considering the two interfaces that form, the buried surface area of ^{312}NFGAFS^{317} is 418Å^2, while it is 359Å^2 for ^{312}NFGTFS^{317}. Because ^{312}NFGEFS^{317} is composed of only one unique interface, the buried surface area is a doubling of that value, 486Å^2. This illustrates that the packing of ^{312}NFGEFS^{317} is significantly tighter than the

^{312}NFGAFS^{317} segment, which may explain its stability against denaturing conditions and offer insight into how the A315E variant induces TDP-43 aggregation.

In summary we find that the ^{312}NFGAFS^{317} segment of TDP-43 forms a kinked β-sheet structure that is amyloid-like, but reversible. We find that the familial variants within this segment, A315T and A315E, perturb the reference sequence structure in different ways. The A315T mutant creates a less stable kinked structure, whereas the A315E variant creates a more stable, steric-zipper-like structure. It is important to note that we believe it is likely that there are other LARKS throughout the C-terminus that have yet to be identified, as evidenced by the fact that 3D Profiling indicates the presence of 14 putative LARKS in TDP-43 (Fig. S1). Our structures offer a starting point for better understanding the regions of TDP-43 responsible for reversible amyloid-like aggregation.

DISCUSSION

The structures determined in this paper suggest how the LCD of TDP-43 can drive aggregation of the protein in either of two directions: into intracellular bodies such as stress granules and also into pathogenic amyloid fibrils. In Figure 4 we present three models that incorporate our findings to explain how LLPS or stress granule assembly can lead to pathogenic assembly. In Fig. 4A we illustrate how the wildtype protein can form pathogenic amyloid assemblies. The soluble protein begins the process of stress granule assembly, LLPS(states 1 to 2). The LARKS of the LCD, such as ^{312}NFGAFS^{317}, can stabilize the stress granule (States 2 to 3). Because we have shown that the C-terminus contains denaturing-resistant, steric-zipper segments (^{300}GNNQGSN^{306}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402}), it is then possible for the LARKS-mediated liquid to progress to pathogenic amyloid assembly (States 2 to 4 and 3 to 4). Our models do not clarify what prevents LARKS and steric zippers from spontaneously assembling in other cellular conditions. However, it is clear that nucleation of either structure is stochastic and rare. Higher concentrations and longer times favor formation of both

intracellular bodies such as stress granules and pathogenic aggregation because the elements of both LARKS and steric zippers are present in the LCD.

The models of Figures 4B and C are variations of the first model that illustrate how familial variants in the LCD can contribute to pathogenic amyloid formation. Fig. 4B illustrates how the A315T variant destabilizes LARKS assembly and pushes the equilibrium towards pathogenic assembly. The interface of ^{312}NFGTFS^{317} is smaller than the ^{312}NFGAFS^{317} interface. This weakens the LARKS assembly and suggests that the protein will spend less time in State 3, pushing the protein back to State 2 LLPS or into pathogenic amyloid assembly. Figure 4C illustrates how the A315E variant creates a new steric-zipper forming segment which pushes the protein into pathogenic amyloid assembly. Here the ^{312}NFGEFS^{317} segment no longer forms a reversible LARKS, but instead forms a steric zipper. By creating this new zipper, the mutation removes a stable reversible segment and increases the likelihood that an irreversible steric zipper will form.

These structure-based models suggest how LLPS and stress granule assembly may contribute to pathogenic amyloid formation. We have structurally characterized two familial variants in this paper, but it is our hypothesis that the other 40+variants found in the TDP-43 LCD may operate by similar mechanisms: either destabilizing LARKS assemblies or creating new steric zipper segments. We suggest these structural concepts can apply to other RNA binding stress granule proteins such as FUS and hnRNPA1. In conclusion, we believe this study offers the first atomic-level insight into how the LCD of TDP-43 functions in both reversible stress granule and irreversible pathogenic assemblies.

MATERIALS AND METHODS

Selection and Generation of TDP-43 Peptide Targets

Putative steric zipper forming segments were selected based on the presence of glutamine and asparagine residues (58, 59). Targets were narrowed down using the 3D Profile

Method, ZipperDB, selecting for segments with energies below -22 kcal/mol (67). ^{312}NFGAFS^{317} was selected based on the presence of aromatic residues and familial variants located within the segment (17).

Peptide Fibril Formation

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

Negative Stain Transmission Electron Microscopy (TEM)

Samples for TEM were prepared by applying 5 μL of sample on hydrophilic 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The fibrils were allowed to adhere for 4 min and washed twice with water. The samples were then stained for 2 min with 2% uranyl acetate, leaving a few nanoliters of stain on the grid. Grids were allowed to dry for 1 minute before storage. Each grid was inspected on a T12 (FEI) microscope.

Fibril Denaturation Studies

Peptide fibrils were generated as described above. Untreated and treated samples were then prepared. For untreated samples, 6μL of water was added to 54μL of fibril preparation. For treated samples, 6μL of 20% SDS solution were added to 54μL of fibril preparation for a final 2% SDS concentration. The treated samples were then quiescently incubated at 70°C for 15 minutes. Turbidity experiments were subsequently performed as untreated and treated samples were aliquoted into black 384-well plates (Nunc, Rochester, NY) covered with optical tape. Values were recorded at a wavelength of 340nm in a Spectramax M5 (Molecular Devices, Sunnyvale, CA). Experiments were performed in triplicate. Data was normalized against untreated fibril samples so the starting value was 1.00. Error bars represent standard deviation.

Crystallization

All crystals, except ^{312}NFGEFS^{317}, were grown by hanging drop vapor diffusion. ^{300}GNNQGSN^{306} was dissolved in water to a final concentration of 10 mg/mL. The reservoir

solution was optimized around the initial hit of 0.2M sodium sulfate, 0.1M Bis tris propane pH 7.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein: reservoir.

^{370}GNNSYS^{375} was dissolved in water to a final concentration of 10mg/mL. Crystals were grown over a reservoir solution containing 0.2M sodium nitrate, 0.1M Bis tris propane pH 8.5, and 20% (w/v) PEG 3350. Hanging drops contained a 2:1 solution of protein: reservoir.

^{396}GFNGGFG^{402} was dissolved in water to a final concentration of 12 mg/mL. Crystals formed with the reservoir solution: 800 mM Sodium phosphate monobasic/ 1200 mM Potassium phosphate dibasic, 100 mM Sodium acetate/ Acetic acid pH 4.5. Hanging drops contained a 2:1 solution of protein: reservoir.

^{312}NFGAFS^{317} was dissolved in water to a final concentration of 11mg/mL. Crystals formed with a reservoir solution of 0.1M phosphate/citrate pH 4.2, 40% w/v ethanol and 5% w/v PEG 1,000. The drops contained a 2:1 ratio of protein solution: reservoir.

^{312}NFGTFS^{317} was dissolved in water to a final concentration of 11mg/mL. Crystals formed with a reservoir solution of 0.2M ammonium acetate, 0.1M sodium acetate pH 4.6, and 30% w/v PEG 4,000. The drops contained a 2:1 ratio of protein solution: reservoir.

^{312}NFGEFS^{317} was dissolved in PBS (pH 7.5) to a final concentration of 20mM. The peptide solution was shaken at 37°C in a Torrey Pines shaker at level 9 for 80 hours.

X-ray Diffraction Data Collection and Processing

All x-ray diffraction data was collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA. Detector ADSC Q315 was used. For NFGAFS, data was collected at a wavelength of 0.9791Å, temperature of 100K, at 5° oscillations and 130mm detector distance. Indexing and integration was done using XDS in P212121 and scaled with SCALEPACK to 1.1Å resolution (68). The data was imported using the CCP4 suite (69). Molecular replacement was done using Phaser (70). Initial models of different idealized polyalanine β-strands did not yield a structure. The final model used was a kinked strand with the sequence NFGAAS, which was based on the truncated structure of NNFGAIL (71). The

model was manually adjusted over 3 cycles to the correct sequence and rotamer orientation in COOT (72). The atomic model was refined by Phenix with a final Rfactor/Rfree of 13.3/16.0 (73).

For NFGTFS, data was collected at a wavelength of 0.9791Å, temperature of 100K, at 5° oscillations and 125mm detector distance. Indexing and integration was done using XDS in P21 and scaled with SCALEPACK to 1.0Å resolution (68). The data was imported using the CCP4 suite (69). The structure was solved by direct methods using the program SHELX (74). The model was manually adjusted over 3 cycles to the correct sequence and rotamer orientation in COOT (72). The atomic model was refined by Refmac with a final Rfactor/Rfree of 8.7/9.3 (75).

For GNNQGSN, data was collected at a wavelength of 0.9791Å, temperature of 100K, at 5° oscillations and 125mm detector distance. Indexing and integration was done using XDS in C2 and scaled with SCALEPACK to 1.8Å resolution (68). The data was imported using the CCP4 suite (69). Molecular replacement was done using Phaser with an idealized β-strand of the sequence GNNAGA (70). The model was manually adjusted over 38 cycles to the correct 7 residue sequence and rotamer orientation in COOT (72). The atomic model was refined by Refmac with a final Rfactor/Rfree of 26.0/26.0 (75).

For GFNGGFG, data was collected at a wavelength of 0.9791Å, temperature of 100K, at 5° oscillations and 150mm detector distance. Indexing and integration was done using Denzo in P1 and scaled with SCALEPACK to 1.25Å resolution (76). The data was imported using the CCP4 suite[69]. Molecular replacement was done using Phaser (70). Initial models of different idealized polyalanine β-strands did not yield a structure. The final model used was two kinked strands with the sequence GFAGG, which was based on the truncated structure of SYSGYS (66). The model was manually refined using Phenix over 4 cycles to the correct 7 residue sequence and rotamer orientation in COOT (72, 73). The atomic model has a final Rfactor/Rfree of 14.4/16.1.

For GNNSYS, data was collected at a wavelength of 0.9791Å, temperature of 100K, at 3° oscillations and 125mm detector distance. Indexing and integration was done using Denzo in

P212121 and scaled with SCALEPACK to 1.00Å resolution (76). The data was imported using the CCP4 suite (69). Molecular replacement was done using Phaser with an idealized 5 residue polyalanine β-strand (70). The model was manually refined using Phenix over 2 cycles to the correct 6 residue sequence and rotamer orientation in COOT (72, 73). The atomic model has a final Rfactor/Rfree of 15.5/18.1.

MicroED Data Collection and Processing

The microED data for segment NFGEFS was collected using the same protocol as Rodriguez et al. with minor differences detailed below (77-79). Data was collected at HHMI Janelia Research Campus using the Tecnai F20 TEM operating at 200 kV recorded using a bottom mount TVIPS F416 CMOS camera. Individual image frames were exposed for 3s per image. During each exposure, crystals were continuously rotated within the beam at a rate of 0.2° per second. Diffraction data was collected on 15 individual crystals, but only 4 were used for data processing.

Indexing and integration was done using XDS and data sets from 4 crystals were merged in P21212 and scaled with SCALEPACK to 1.0Å resolution (68). The data was imported using the CCP4 suite (69). The structure was solved by direct methods using the program SHELX (74). The model was manually adjusted over 4 cycles to the correct sequence and rotamer orientation in COOT (72). The atomic model was refined by Refmac with a final Rfactor/Rfree of 23.7/28.6 (75).

Structure Analysis

All images of crystallographic structures were generated in Pymol. Area buried and shape complementarity values were calculated based on previous literature (80, 81).

ACKNOWLEDGEMENTS

We thank Todd Yeates, Feng Guo, James Bowie, Pascal Krotee, and Smriti Sangwan for insightful discussions. We thank the Advanced Photon Source (APS) staff (M. Capel, K.

Rajashan- kar, N. Sukumar, J. Schuermann, I. Kourinov and F. Murphy) at the Northeastern Collaborative Access Team beamline 24-ID-E for support throughout two years of data collection.. We also thank Michael Collazo and the UCLA-DOE Macro-Molecular Crystallization Core Technology Center for setting up crystallization screens and Ivo Atanasov and the Electron Imaging Center for NanoMachines (EICN) of California NanoSystems Institute (CNSI) at UCLA for support on the use of T12 FEI electron microscope. The APS is funded by National Institutes of Health grants P41 RR015301 and P41 GM103403 and Department of Energy under Contract Number DE-AC02-06CH11357. The UCLA-DOE Macromolecular Crystallization Core Technology Center is partially funded by the Department of Energy grant DE-FC0302ER63421.. We thank the Howard Hughes Medical Institute (HHMI),DOE and NIH grant R01-AG029430 for support to the Eisenberg Lab. Elizabeth Guenther was f'unded by the UCLA CMB Training Program (Ruth L. Kirschstein NIH GM00785).

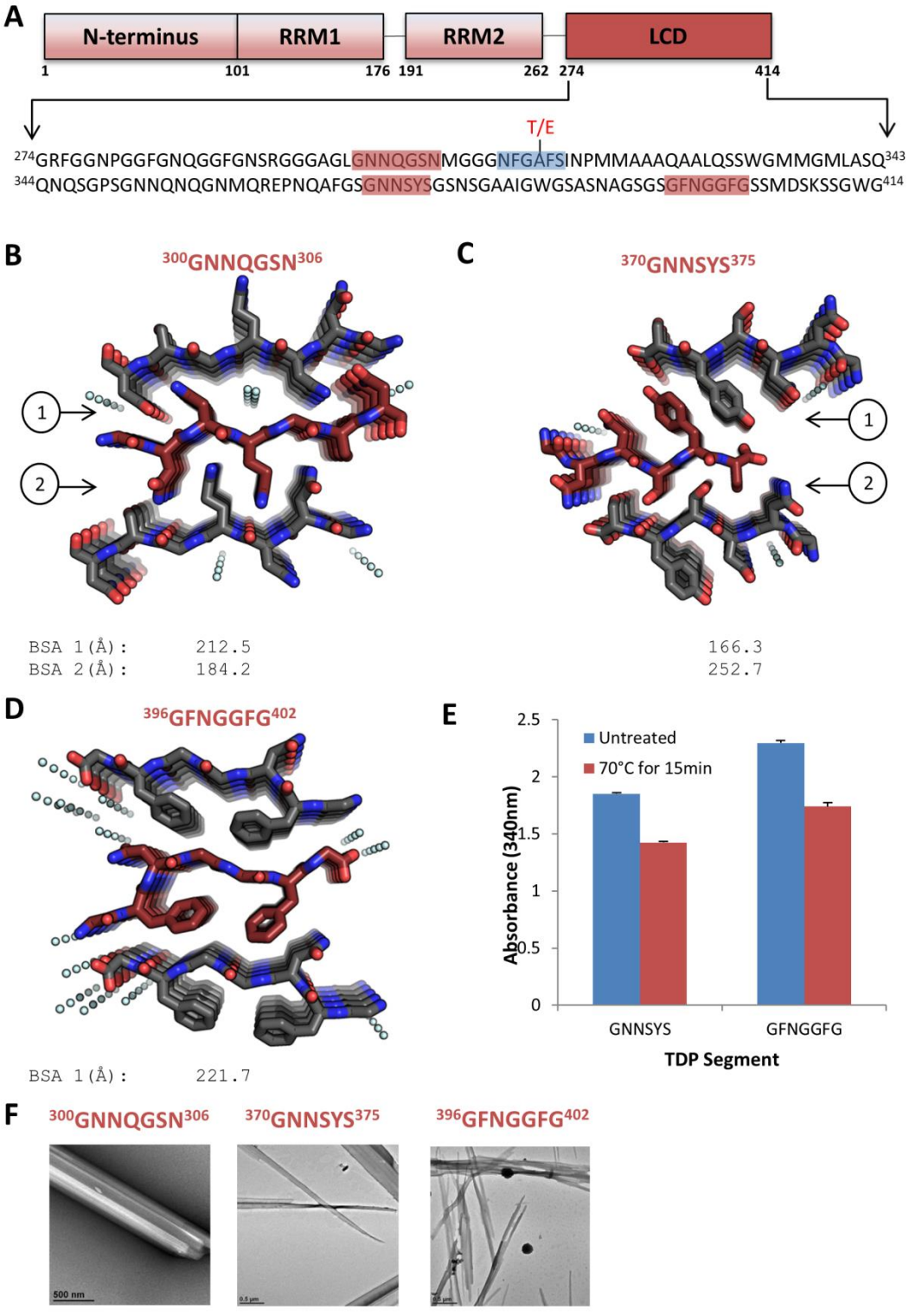


Figure 2.1: Segments from the C-terminus form steric zippers and exhibit high stability against heat denaturation. (A) The sequence of the LCD C-terminus of TDP-43.

The three segments that form steric zippers are highlighted in red, while the LARKS segment, ^{312}NFGAFS^{317} is highlighted in blue. The position of familial variants A315T and A315E are indicated in red font. (B) The structure of ^{300}GNNQGSN^{306} was determined from x-ray diffraction data taken at the Advanced Photon Source. The segment forms a parallel, face-to-face, Class 1 steric zipper interface, viewed here down the fibril axis. (C) The structure of ^{370}GNNSYS^{375} was determined from x-ray diffraction data taken at the Advanced Photon Source. The segment also forms a parallel, face-to-face, Class 1 steric zipper interface, viewed here down the fibril axis. (D) The structure of ^{396}GFNGGFG^{402} was determined from x-ray diffraction data taken at the Advanced Photon Source. The segment also forms a parallel, face-to-back, Class 4 steric zipper interface, viewed here down the fibril axis. (E) Segments, ^{370}GNNSYS^{375} and ^{396}GFNGGFG^{402} exhibit high stability under heat and SDS denaturation consistent with their steric-zipper structures. Fibril samples were first grown by shaking at 37°C at 20mM concentration. Following fibril formation, samples were treated with 2% SDS and heated to 70°C for 15 minutes.

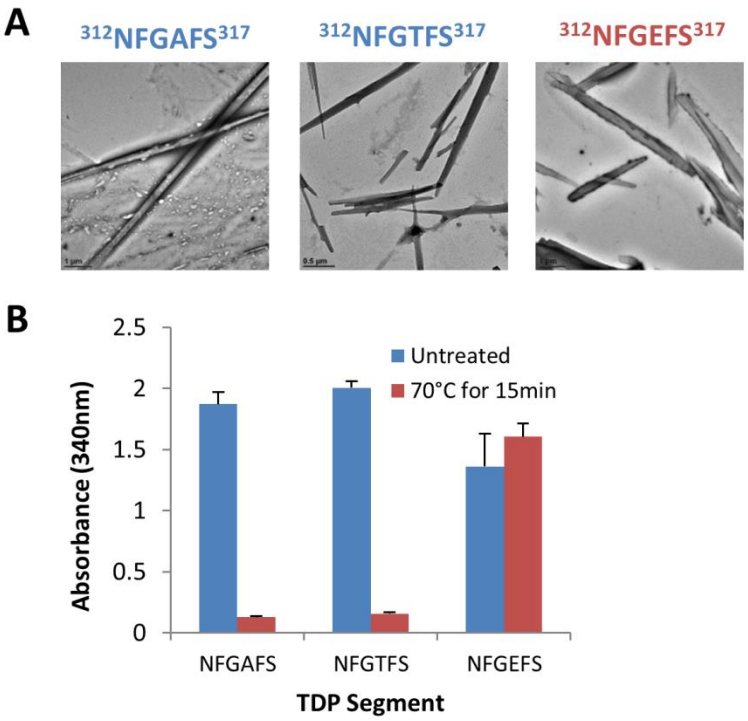


Figure 2.2: The ^{312}NFGAFS^{317} segment of TDP-43 forms labile fibrils that are stabilized by the familial variant A315E and destabilized by A315T. (A) Negatively stained EM images of 20mM fibril samples of the reference and familial variant sequences. Notice that these fibrils resemble fibrils formed by characteristic amyloid proteins. (B) ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317} dissociate under denaturing conditions, while ^{312}NFGEFS^{317} exhibits stability comparable to that of steric zippers. Fibril samples were first grown by shaking at 37°C at 20mM concentration. Following fibril formation, samples were treated with 2% SDS and heated to 70°C for 15 minutes.

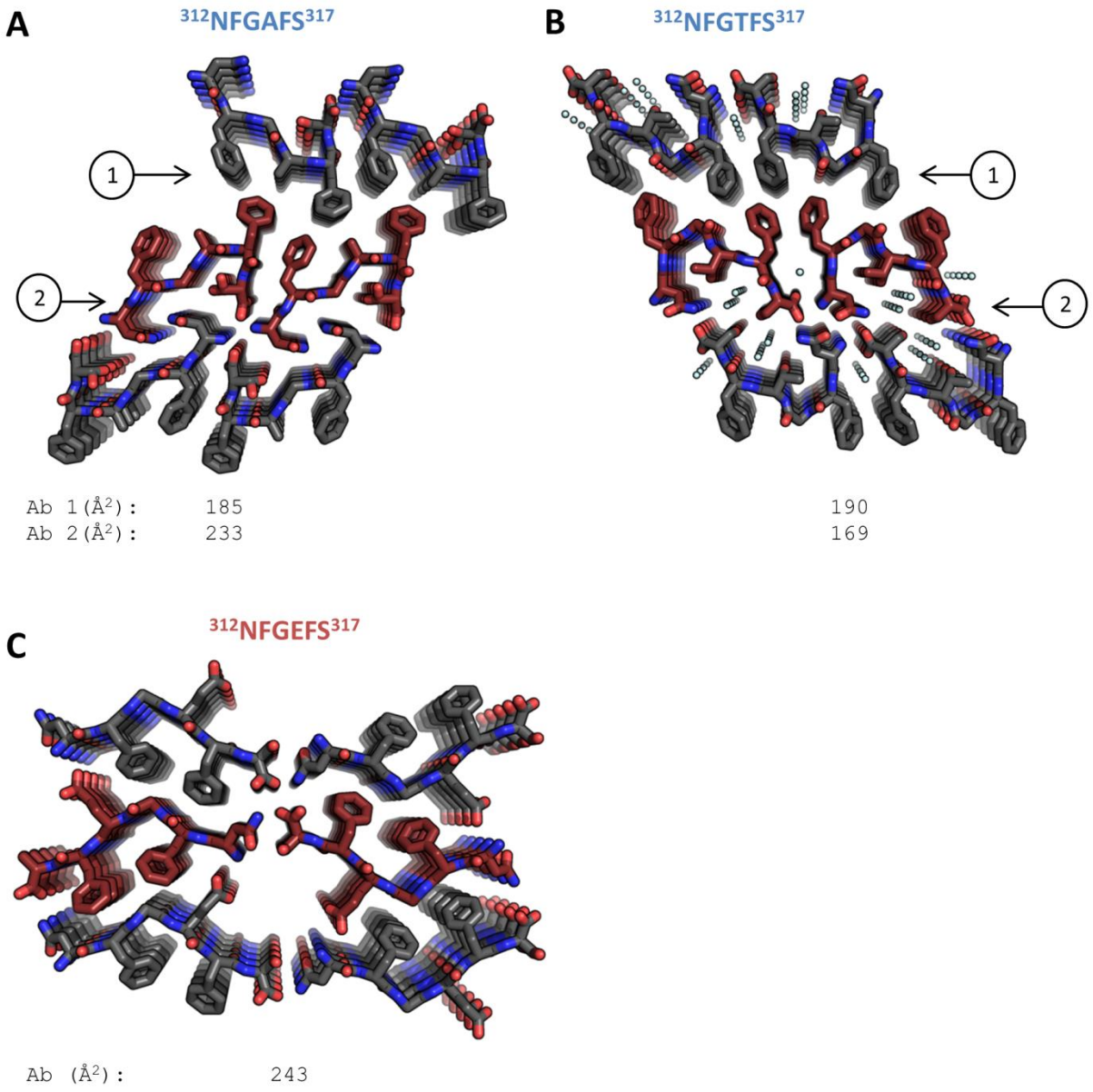


Figure 2.3: The ^{312}NFGAFS^{317} segments forms a kinked β-sheet that is disrupted by familial variants A315T/E (A) The structure of ^{312}NFGAFS^{317} was determined from x-ray diffraction taken at the Advanced Photon Source. The segment forms a kinked structure which places residues 2, 4 and 5 on the same face of the β-sheet. An aromatic, hydrophobic interface 1 mediates the interaction of mating sheets. Here we view the structure down the fibril axis. (B) The structure of ^{312}NFGTFS^{317} was determined from x-ray diffraction taken at the Advanced

Photon Source. The segment also kinks, but at the central glycine, placing residues 2, 3, and 5 on the same face of the β-sheet. The structure packs less tightly than ^{312}NFGAFS^{317} and is hydrated. The structure is displayed looking down the fibril axis. (C) The structure of ^{312}NFGEFS^{317} was determined by electron diffraction at Janelia Research Campus. This segment kinks at the glutamate, placing residues 2 and 5 on the same face of the β-sheet. The mating sheets pack face-to-back and do not have the aromatic interface observed in ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317}. The area buried for this structure is the largest of all three structures, all of which are viewed down the fibril axis.

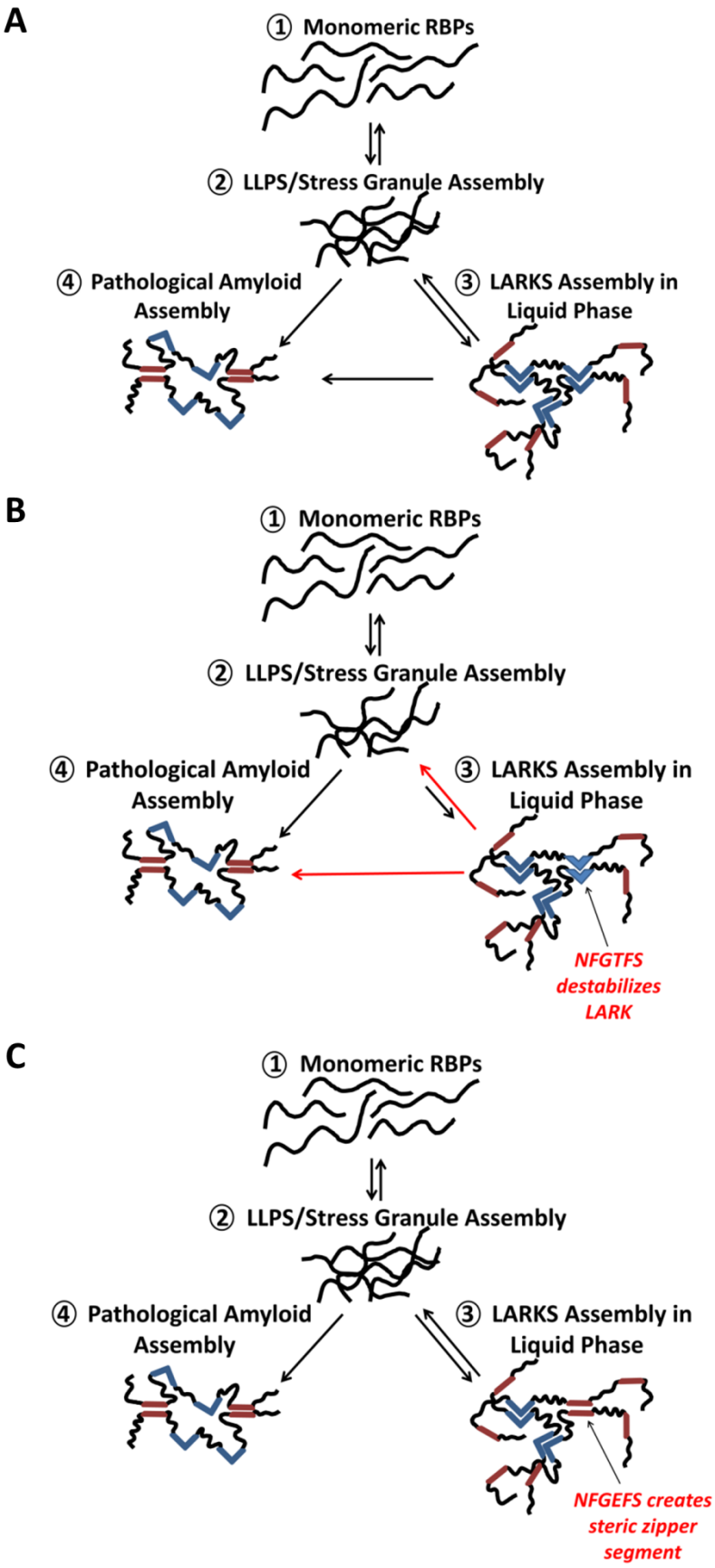

A
① Monomeric RBPs
② LLPS/Stress Granule Assembly
④ Pathological Amyloid Assembly
③ LARKS Assembly in Liquid Phase
B
① Monomeric RBPs
② LLPS/Stress Granule Assembly
④ Pathological Amyloid Assembly
③ LARKS Assembly in Liquid Phase
NFGTFS destabilizes LARK
C
① Monomeric RBPs
② LLPS/Stress Granule Assembly
④ Pathological Amyloid Assembly
③ LARKS Assembly in Liquid Phase
NFGEFS creates steric zipper segment

Figure 2.4: Models illustrating how the LCD of TDP-43 can participate in dynamic intracellular bodies such as stress granules through LARKS (﹀) and in pathogenic amyloid fibrils through steric zippers (▬) (A) By concentration, LLPS of protein can promote pathogenic amyloid aggregation. The soluble protein begins LLPS (states 1 to 2). LARKS can then associate and stabilize the stress granule (states 2 to 3). High concentration of protein favors association of steric zipper-forming segments leading to pathogenic aggregates (states 2 to 4 and 3 to 4). (B) A variant sequence such as A315T can destabilize the LARKS association, pushing the aggregate towards irreversible, pathogenic fibril assembly. Here, because state 3 is destabilized, the protein will move back to LLPS (state 2) or pathogenic fibrils (state 4). (C) A familial variant sequence such as A3115E can convert a LARKS into a new steric-zipper interface. In this case, the new interface promotes pathogenic aggregation and pulls the protein into an irreversible amyloid fibril.

Table 2.1: Statistics of x-ray diffraction data collection and atomic refinement for ^{300}GNNQGSN^{306}, ^{370}GNNSYS^{375}, and ^{396}GFNGGFG^{402}

	GNNQGSN	GFNGGFG	GNNSYS
Data Collection			
Beam Line	APS NECAT 24-ID-E	APS NECAT 24-ID-E	APS NECAT 24-ID-E
Detector	ADSC Q315	ADSC Q315	ADSC Q315
Date	8/13/2014	6/19/2014	4/18/2015
PDB accession code			
Space group	C2	P1	P212121
Unit cell dimensions			
a,b,c (Å)	49.67, 4.77, 14.99	4.81, 16.44, 22.67	4.79, 15.62,40.68
α,β,γ(°)	90, 101.66, 90	90.63, 86.37, 91.26	90, 90, 90
Reflections observed	1,177	7,394	9,975
Unique reflections	385	1,802	1,837
Wavelength (Å)	0.9791	0.9791	0.9791
Resolution (Å)	24.32-1.80	22.53-1.24	20.34-1.00
Highest Resolution Shell (Å) [a]	2.08-1.80	1.29-1.25	1.06-1.00
R_{sym} (%) [b]	32.5(57.1)	16.1(40.6)	16.1(50.0)
CC(1/2)	95.5(86.3)	98.5(69.0)	99.2(66.8)
I/σ	2.75(1.71)	7.10(1.95)	9.2(2.1)
Completeness (%)	94.1(97.8)	94.4(95.4)	95.4(70.0)
Wilson B value (Å^2)	6.22	2.43	2.69
Refinement			
Resolution (Å)	24.32-1.80	22.53-1.24	20.34-1.00
Resolution (Å) last shell	1.85-1.80	1.29-1.24	1.04-1.00
Reflections in Working Set	345	1,614	1,667
Reflections in Test Set	39	180	169
R_{work}(%)/Rfree(%)[c]	26.0(36.6)/26.0(61.0)	14.4(26.2)/16.1(39.0)	15.5(28.9)/18.1(28.4)
Protein Molecules in Asymmetric Unit	1	2	1
Number of non-H atoms			
Protein	49	94	45
Non-protein	2	8	1
RMS deviations			
Bond lengths (Å)	0.019	0.006	0.005
Bond angles (°)	1.70	1.04	1.18
Average *B*-factor (Å^2)			
Protein atoms	15.42	3.3	3.46
Non-protein atoms	22.59	7.4	15.34

Ramachandran plot regions [d]			
Favored	100	100	100
Allowed	0	0	0
Outliers	0	0	0

[a] Highest resolution shell shown in parenthesis.

[b] $Rsym = 100\,x \sum (I - I(mean))^2 / \sum I^2$

where I is the observed intensity of the reflection HKL and the sum is taken over all reflections HKL.

[c] $Rfactor = 100\,x \sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \sum \left| F_{obs} \right|$

F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement

[d]Percentage of residues in Ramachandran plot regions were determined using Molprobity

Table 2.2: Statistics of x-ray diffraction data collection and atomic refinement ^{312}NFGAFS^{317} and ^{312}NFGTFS^{317}

	NFGAFS	NFGTFS
Data Collection		
Beam Line	APS NECAT 24-ID-E	APS NECAT 24-ID-E
Detector	ADSC Q315	ADSC Q315
PDB accession code		
Space group	P212121	P21
Unit cell dimensions		
a,b,c (Å)	13.82, 4.85, 46.74	15.33, 4.84, 23.57
α,β,γ(°)	90, 90, 90	90.00, 96.91, 90.00
Reflections observed	12,008	
Unique reflections	1,447	
Wavelength (Å)	0.9791	
Resolution (Å)	13.25-1.10	
Highest Resolution Shell (Å) [a]	1.14-1.10	
R_{sym} (%) [b]	17.2 (72.6)	
CC(1/2)	99.1 (82.3)	
I/σ	9.65 (2.52)	
Completeness (%)	94(75.9)	87.6(39.2)
Wilson B value (Å^2)	4.55	2.54
Refinement		
Resolution (Å)	13.25-1.10	13.52-1.00
Resolution (Å) last shell	1.14-1.10	1.04-1.00
Reflections in Working Set	1,300	1,635
Reflections in Test Set	145	183
R_{work}(%)/Rfree(%)[c]	13.3(21.9)/16.0(18.3)	
Protein Molecules in Asymmetric Unit	1	
Number of non-H atoms		
Protein	50	48
Non-protein	0	3
RMS deviations		
Bond lengths (Å)	0.006	0.020
Bond angles (°)	0.95	1.73
Average *B*-factor (Å^2)		
Protein atoms	4.55	2.39
Non-protein atoms	0	14.75
Ramachandran plot regions [d]		
Favored	100	100
Allowed	0	0
Outliers	0	0

[a] Highest resolution shell shown in parenthesis.

[b] $Rsym = 100\,x \sum (I - I(mean))^2 / \sum I^2$

where I is the observed intensity of the reflection HKL and the sum is taken over all reflections HKL.

[c] $Rfactor = 100\,x \sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \sum \left| F_{obs} \right|$

F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement

[d] Percentage of residues in Ramachandran plot regions were determined using Molprobity

Table 2.3: Statistics of microED data collection and atomic refinement for ^{312}NFGEFS^{317}

Sample	NFGEFS
Excitation Voltage (kV)	200
Wavelength (Å)	0.0251
Total dose per crystal (e/ Å^2)	3.4
Frame rate (frame/s)	0.3-0.5
Rotation rate (°/s)	0.2
# crystals used	4
Total angular rotation collected (°)	72
Merging Statistics	
Space group	P21212
Unit cell dimensions	
a, b, c (A°)	42.77, 17.42, 4.90
α, β, γ (°)	90, 90, 90
Resolution (Å) [a]	13.54-1.00
Rmerge	1.03-1.00
# of reflections	28.3(99.3)
Unique reflections	18,942
Completeness	2,004
Multiplicity	88.7(78.0)
I/sigma	9.45(4.81)
$CC_{1/2}$ (Diederichs, 2013)	4.64(1.07)
Refinement Statistics	99.2(36.0)
Reflections in working set	
Reflections in test set	1,851
R_{work}	206
R_{free}	23.7
RMSD bonds (Å)	28.6
RMSD angles (°)	0.008
Ramachandran (%)[d]	0.69
Favored	
Allowed	75
Outliers	25
PDB ID code	0
EMDB ID code	

[a] Highest resolution shell shown in parenthesis.

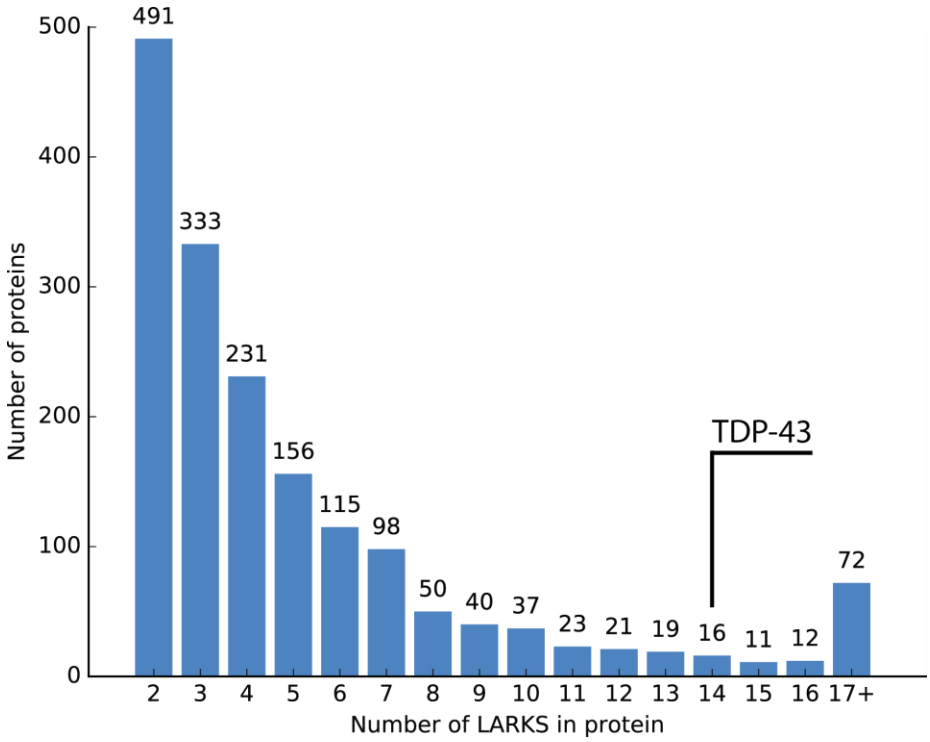
[b] $Rsym = 100\,x\sum (I - I(mean))^2 / \sum I^2$

where *I* is the observed intensity of the reflection HKL and the sum is taken over all reflections HKL.

[c] $Rfactor = 100\,x\sum \left\| F_{obs} \right| - \left| F_{calc} \right\| / \sum \left| F_{obs} \right|$

F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{work} refers to the R_{factor} for the data utilized in the refinement and R_{free} refers to the R_{factor} for 5% of the reflections randomly chosen that were excluded from the refinement

[d]Percentage of residues in Ramachandran plot regions were determined using Molprobity



Supplementary Figure 2.1: 14 putative LARKS are found in the LCD of TDP-43. The graph illustrates the number of proteins in the human proteome that contain putative LARKS and the respective number of LARKS. The data were generated by a 3D Profiling algorithm developed in our lab that predicts the likelihood of a peptide segment adopting a kinked backbone conformation. TDP-43 is predicted to have 14 LARKS, which is on the high end of predictions.

REFERENCES

1. Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293,** 151–160 (1999).

2. Golding, G. B. Simple sequence is abundant in eukaryotic proteins. *Protein Sci.* **8,** 1358–61 (1999).

3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).

4. Mangiarini, L. *et al.* Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87,** 493–506 (1996).

5. Zoghbi, H. Y. & Orr, H. T. G Lutamine R Epeats and. *Annu. Rev. Neurosci.* **23,** 217–247 (2000).

6. Zuccato, C. *et al.* Loss of Huntingtin-Mediated BDNF Gene Transcription in Huntington ' s Disease. **293,** 493–498 (2001).

7. March, Z. M., King, O. D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* **1647,** 9–18 (2016).

8. Liu-Yesucevitz, L. *et al.* Tar DNA binding protein-43 (TDP-43) associates with stress granules: Analysis of cultured cells and pathological brain tissue. *PLoS One* **5,** (2010).

9. Bentmann, E. *et al.* Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* **287,** 23079–23094 (2012).

10. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163,** 123–133 (2015).

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

12. Kedersha, N., Gupta, M., Li, W., Miller, I. & Anderson, P. eIF-2 ⎵ to the Assembly of Mammalian Stress Granules. *J. Cell Biol.* **147,** 1431–1441 (1999).

13. 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).

14. Gilks, N. *et al.* Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1 □. **15,** 5383–5398 (2004).

15. Dang, Y. *et al.* Eukaryotic initiation factor 2α-independent pathway of stress granule induction by the natural product pateamine A. *J. Biol. Chem.* **281,** 32870–32878 (2006).

16. Kedersha, N. *et al.* Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *J. Cell Biol.* **151,** 1257–1268 (2000).

17. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149,** 753–767 (2012).

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

19. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by ??-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24,** 1537–1549 (2016).

20. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60,** 208–219 (2015).

21. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *Elife* **5,** 1–25 (2016).

22. Chiti, F. & Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).

23. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15,** 384–96 (2014).

24. Greenwald, J. & Riek, R. Biology of amyloid: Structure, function, and regulation. *Structure* **18,** 1244–1260 (2010).

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

26. Meersman, F. & Dobson, C. M. Probing the pressure–temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochim. Biophys. Acta - Proteins Proteomics* **1764,** 452–460 (2006).

27. Harrison, A. F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* **474,** 1417–1438 (2017).

28. Rutherford, N. J. *et al.* Novel Mutations in TARDBP (TDP-43) in Patients with Familial Amyotrophic Lateral Sclerosis. *PLoS Genet.* **4,** e1000193 (2008).

29. Kabashi, E. *et al.* TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* **40,** 572–574 (2008).

30. Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D. W. TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19,** 46–64 (2010).

31. Pesiridis, G. S., Lee, V. M. Y. & Trojanowski, J. Q. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **18,** 156–162 (2009).

32. Igaz, L. M. *et al.* Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* **284,** 8516–8524 (2009).

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

34. Li, Q., Yokoshi, M., Okada, H. & Kawahara, Y. The cleavage pattern of TDP-43 determines its rate of clearance and cytotoxicity. *Nat. Commun.* **6,** 6183 (2015).

35. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in als and FTD:

Disrupted RNA and protein homeostasis. *Neuron* **79,** 416–438 (2013).

36. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. 1 , 2 1 1. **2,** 2 (2014).

37. Chang, C. ke *et al.* The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem. Biophys. Res. Commun.* **425,** 219–224 (2012).

38. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **20,** 1443–1449 (2013).

39. Saini, A. & Chauhan, V. S. Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir* **30,** 3845–3856 (2014).

40. Polymenidou, M. *et al.* Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Res.* **1462,** 3–15 (2012).

41. Bhardwaj, A., Myers, M. P., Buratti, E. & Baralle, F. E. Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res.* **41,** 5062–5074 (2013).

42. Buratti, E. & Baralle, F. E. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA Biol.* **7,** 420–429 (2010).

43. Schwab, C., Arai, T., Hasegawa, M., Yu, S. & Mcgeer, P. L. Colocalization of Transactivation-Responsive DNA-Binding Protein 43 and Huntingtin in Inclusions of Huntington Disease. *J Neuropathol Exp Neurol* **67,** 1159–1165 (2008).

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

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

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

47. Manuela Neumann1, 11 *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science (80-.).* (2006). doi:10.1126/science.1134108

48. Yang, C. *et al.* The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. *PLoS One* **5,** (2010).

49. Pesiridis, G. S., Tripathy, K., Tanik, S., Trojanowski, J. Q. & Lee, V. M. Y. A 'two-hit' hypothesis for inclusion formation by carboxyl-terminal fragments of TDP-43 protein linked to RNA depletion and impaired microtubule-dependent transport. *J. Biol. Chem.* **286,** 18845–18855 (2011).

50. Chen, A. K. *et al.* Induction of Amyloid Fibrils by the C-Terminal Fragments of TDP-43 in Amyotrophic Lateral Sclerosis. 1186–1187 (2010).

51. Cushman, M., Johnson, B. S., King, O. D., Gitler, A. D. & Shorter, J. Prion-like disorders: blurring the divide between transmissibility and infectivity. *J. Cell Sci.* **123,** 1191–1201 (2010).

52. Guo, W. *et al.* An ALS-associated mutation affecting TDP-43 enhances protein

aggregation , fibril formation and neurotoxicity. **18,** (2011).

53. Jiang, L. L. *et al.* Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* **288,** 19614–19624 (2013).

54. Jiang, L.-L. *et al.* Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Sci. Rep.* **6,** 23928 (2016).

55. Mompeán, M. *et al.* Structural characterization of the minimal segment of TDP-43 competent for aggregation. *Arch. Biochem. Biophys.* **545,** 53–62 (2014).

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

57. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–7 (2007).

58. Budini, M. *et al.* Cellular model of TAR DNA-binding Protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. *J. Biol. Chem.* **287,** 7512–7525 (2012).

59. Fuentealba, R. A. *et al.* Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. *J. Biol. Chem.* **285,** 26304–26314 (2010).

60. 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,** 3487–92 (2010).

61. Gitcho, M. A. *et al.* TDP-43 A315T mutation in familial motor neuron disease. *Ann. Neurol.* **63,** 535–538 (2008).

62. Xu, M. *et al.* Characterization of b -domains in C-terminal fragments of TDP-43 by scanning tunneling microscopy. *J. Struct. Biol.* **181,** 11–16 (2013).

63. Lim, L., Wei, Y., Lu, Y. & Song, J. ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43. *PLoS Biol.* **14,** 1–35 (2016).

64. Estes, P. S. *et al.* Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a Drosophila model of ALS. **20,** 2308–2321 (2011).

65. Walker, A. K. *et al.* ALS-associated TDP-43 induces endoplasmic reticulum stress, which drives cytoplasmic TDP-43 accumulation and stress granule formation. *PLoS One* **8,** 1–12 (2013).

66. Hughes, M. Low-complexity domains adhere by reversible amyloid-like interactions between kinked beta-sheets. *Science (80-.).* **Submitted,** (2017).

67. Wiltzius, J. J. W. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16,** 973–8 (2009).

68. Kabasch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr* **26,** 795– 800 (1993).

69. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*

Crystallographica Section D: Biological Crystallography **67,** 235–242 (2011).

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

71. Wiltzius, J. E. D. J. W. *et al.* Atomic structure of the cross-b spine of islet amyloid polypeptide (amylin). *Insulin* 1467–1474 (2008). doi:10.1110/ps.036509.108.The

72. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

73. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

74. Sheldrick, G. A short history of SHELX. *Acta Crystallogr A* **64,** 112–122 (2008).

75. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **53,** 240–255 (1997).

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

77. Hattne, J. *et al.* MicroED data collection and processing. *Acta Crystallogr. Sect. A Found. Adv.* **71,** 353–360 (2015).

78. Shi, D. *et al.* The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11,** 895–904 (2016).

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

80. Lawrence, M. & Colman, P. Shape complemntarity at protein/protein interfaces. *J Mol Biol* **234,** 956–950 (1993).

81. Anonymous. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* **50,** 760–763 (1994).

CHAPTER 3

Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease that affects motor neurons, often causing death within two to five years. 90% of ALS cases are 'sporadic' (sALS) and their cause is unknown (1). However, the remaining 10% of ALS cases are inherited 'familial' ALS (fALS) and approximately 20% of which are linked to mutations in the SOD1 gene.

Mature SOD1 is a 32 kDa homodimeric metalloenzyme, in which each monomer contains a copper ion, zinc ion, and one intra-subunit disulfide bond (2) (Fig. 1A). SOD1 is one of the most abundant proteins in cells, serving to protect organisms against oxidative damage. The loss of protein function does not necessarily lead to disease because SOD1-deficient mice develop mild impairments that are not observed in ALS (3). Instead, the mutated SOD1 seems to have a toxic gain of function that leads to the pathologies of disease. To date, more than 140 dominant disease-related mutations that span nearly the whole protein sequence have been described (http://alsod.iop.kcl.ac.uk/Als/).

Several studies suggest sALS and fALS have common mechanisms of pathogenesis associated with accumulation of misfolded SOD1 (4). Evidence to support this has shown that insoluble protein aggregates found in both fALS (5) and sALS (6) patients were SOD1 immunoreactive. Scientists have created transgenic mice that express human SOD1 mutations found in fALS. The mice exhibit behavioral and cellular symptoms similar to human ALS (7), such as accumulation of insoluble aggregates (8). In addition, expression in mice of heterozygous wild-type/mutant SOD1 augments disease symptoms relative to homozygous

mutant animals (7), implying that the wild-type protein produced by the allele carrying the normal gene enhances the toxicity of the mutant protein in fALS. It was also shown that transgenic mice over-expressing wild-type SOD1 show ALS-like symptoms and large amounts of aggregated SOD1 in the spinal cord and brain (9). These findings suggest that investigating the mechanism of aggregation of wild-type and mutant SOD1 may further our understanding of the molecular origins of both fALS and sALS.

SOD1-containing pathological inclusions in ALS may be amyloid-like in structure. First, neuronal tissue from a mouse model that expresses the H46R/H48Q mutations can be stained with Thioflavin S, a fluorescent dye whose properties change in the presence of amyloid-like aggregates (10). Second, in vitro, reduction of the disulfide bond and removal of metals from SOD1 and its mutants leads to the formation amyloid-like aggregates (11). Third, in vitro formed SOD1 fibrils share common toxic properties with ALS inclusions, such as the ability to induce inflammation (12) and activate microglial cells (13). And fourth amyloid fibril formation of recombinant SOD1 can be seeded with tissue extracts of SOD1-containing inclusions from ALS transgenic mice (14). All these data suggest the pathologic inclusions share similarities with in vitro formed amyloid fibrils.

It is well established that amyloid fibril formation is driven by the exposure of short aggregation prone-segments. Here, we pinpoint the aggregation-prone regions of SOD1 in order to clarify the molecular origins of ALS. Using a computational (15), we identified four segments in SOD1 with high propensities to form fibrils We experimentally observed that proline substitutions in two of these segments block fibril formation of full-length protein, emphasizing the importance of these segments in amyloid aggregation. Our studies therefore suggest specific segments within the SOD1 sequence that drive the aggregation and perhaps initiate pathology of ALS.

RESULTS

Aggregation-prone segments in SOD1

To find which segments induce aggregation of SOD1, we used the 3D-profile method (15) (Figure 3.1B). In SOD1, we find four segments of 7 to 9 residues with energies favorable for fibril formation: ^{14}VQGIINFE^{21}, ^{30}KVWGSIKGL^{38}, ^{101}DSVISIS^{107} and ^{147}GVIGIAQ^{153}. We confirmed our predictions by determining that all four segments form fibril-like or needle-like aggregates (Figure 3.2).

Most in-segment disease mutations are compatible with aggregate formation

Twenty-four mutations associated with fALS (http://alsod.iop.kcl.ac.uk/Als/) are located in the four aggregation-prone segments. We found that only two of the twenty-four mutations changed fibril formation propensity of the segments from favorable to unfavorable (Supplementary Figure 3.1, Supplementary Table 3.1). To validate the computational predictions we synthesized fourteen peptides with disease-related mutations. We found that thirteen of the fourteen mutant peptides formed aggregates with fibril morphology (Fig. 2). The only peptide, G37R, which did not form fibrils was also predicted to have low fibril-formation propensity.

^{101}DSVISLS^{107}and ^{147}GVIGIAQ^{153}form steric zipper structures

The crystal structures of the two fibril-forming segments ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153} were solved to 1.4 Å and 1.9 Å resolution, respectively. Each segment forms pairs of β-sheets that mate together through complementary surfaces (Figures 3.3A and 3.3B). Such structures are typical for amyloid-forming segments, and are known as steric zippers (16) (Figures 3.3A and 3.3B).

The structures revealed that the steric zippers formed by ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153} differ in arrangement. β-Sheets in the ^{101}DSVISLS^{107} structure interact through their distinct surfaces in the face-to-back packing of class-2 steric zippers (Figure 3.3A). In contrast, the ^{147}GVIGIAQ^{153} steric zipper is formed by identical β-sheet surfaces in face-to-face packing of class-1 steric zippers (Figure 3.3B). In the ^{147}GVIGIAQ^{153} structure, we observed that

β-sheets also interact tightly with their opposite sides in back-to-back packing (Supplementary Figure 3.2A). The area buried and shape complementarity of the two steric zipper interfaces are comparable (Supplementary Table 3.2).

Segments with fALS mutations also form steric zipper structures

To determine how mutations found in fALS change the fibril-forming properties of the wild-type segment, we studied the structure of ^{147}GVTGIAQ^{153} (I149T) and ^{147}GIIGIAQ^{153} (V148I). X-ray fiber diffraction of ^{147}GIIGIAQ^{153} (V148I) showed a cross-β diffraction pattern common to amyloid fibrils (Supplementary Figure 3.3), but we could not obtain a high resolution structure. The structure of ^{147}GVTGIAQ^{153} (I149T) was solved to 1.3 Å (Figure 3.3C and Supplementary Figure 3.2B). The dry interface of ^{147}GVTGIAQ^{153} (I149T) compared to the wild-type segment is additionally stabilized by two hydrogen bonds formed between Thr and Gln (Figures 3.3B and 3.3C). In general the ^{147}GVTGIAQ^{153} (I149T) structure reveals that mutations in the steric zipper can change the spatial arrangement of the β-sheet surfaces forming the steric zipper without diminishing its apparent strength.

The ^{147}GVIGIAQ^{153} segment accelerates fibril formation of apoSOD1^{WT}

We tested whether any of the isolated peptide segments accelerate nucleation or growth of full-length protein fibrils. We hypothesized that if a segment is critical for fibril formation, the presence of that short segment should alter the aggregation kinetics of the full-length protein, possibly by providing a template for fibril nucleation (17). Of the four segments, only ^{147}GVIGIAQ^{153} from the C-terminus accelerated fibril formation of apoSOD1^{WT} in a dose-dependent manner (Figure 3.4A and Supplementary Figure 3.6A). This acceleration was consistently observed, although the required ratio of segment to full-length protein varied with the different preparations of SOD1. These results suggest that ^{147}GVIGIAQ^{153} might initiate fibril formation of full-length protein.

^{147}GVIGIAQ^{153} accelerates the fibril formation of apoSOD1^{G93A}

To test whether the ^{147}GVIGIAQ^{153} segment can nucleate the fibril formation of SOD1 mutants associated with fALS, we followed the aggregation of apoSOD1^{G93A}, one of the best characterized mutants of SOD1 (18). Indeed, similar to apoSOD1^{WT}, fibril formation of apoSOD1^{G93A} was accelerated in the presence of ^{147}GVIGIAQ^{153} (Figure 3.4B). Likewise, we observed variability in the aggregation kinetics and in the quantity of ^{147}GVIGIAQ^{153} needed to nucleate the fibril formation of the apoSOD1^{G93A} mutant from different protein preparations. In electron micrographs of samples collected after the ThT fluorescence signal reached a plateau, we observed both thick rigid fibril-like aggregates of ^{147}GVIGIAQ^{153} together with thin flexible thin fibrils of apoSOD1^{G93A} (Figure 3.4B) The fibrils of full-length protein tend to cluster with needle-like aggregates of ^{147}GVIGIAQ^{153}, further hinting that the full-length aggregation is facilitated by the presence of this C-terminal segment.

DISCUSSION

SOD1 contains four fibril-forming segments

Aided by a computational algorithm (15), we identified four segments likely involved in SOD1 fibril formation. In isolation from the rest of the protein, all these segments formed aggregates with fibril morphology. The existence of several aggregation-prone segments in SOD1 unifies conflicting studies that find different regions from the protein participate in aggregate formation (19, 20).

fALS-linked mutations modulate but do not disrupt aggregation

There are roughly 140 mutations in SOD1 that are associated with fALS. Of the fourteen we tested, thirteen formed aggregates (Figure 3.2). The ability of the segments to form fibrils was unaffected by the mutations that changed their charge, size or hydrophobicity. The structure of the mutant segment I149T revealed molecular details of the changes in the steric zipper organization caused by the mutation. Although, the steric zippers formed by wild-type

and mutant segments have different spatial arrangements, their interaction energies are comparable as judged by the area buried and shape complementarity of the interfaces (Supplementary Table 3.2). Although, many fALS mutations promote aggregation by destabilizing the native protein structure, our results suggest that disease-related mutations generally do not inhibit fibril formation of the aggregation-prone segment and may even enhance it.

Two C-terminal segments dominate fibril formation

Full-length protein with I18P and I35P substitutions formed fibrils similar in morphology to the wild-type protein (Supplementary Figure 3.5), leading us to conclude that the N-terminal segments ^{14}VQGIINFE^{21} and ^{30}KVWGSIKGL^{38} do not form the β-sheet spine of the SOD1 fibrils. Pro substitution at position I104 suppressed fibril formation of wild-type protein, implying that this segment is involved in fibril nucleation and/or growth. We also observed change in fibril morphology of the I149P mutant that blocks the aggregation of ^{147}GVIGIAQ^{153} compared to wild-type fibrils. This Pro mutant formed rod-shaped aggregates that were less than 100nm in length (Figure 3.5), suggesting that the C-terminus of SOD1 takes part in fibril elongation and/or forms the β-sheet spine of the SOD1 fibrils. In the samples of both I104P and T149P mutants we observed wild-type fibrils, suggesting that blocking the aggregation of the C-terminal segments may in turn lead to fibril formation through other segments. The two peptide segments, ^{14}VQGIINFE^{21} and ^{30}KVWGSIKGL^{38} formed aggregates with fibril morphology (Figure 3.2) and high β-sheet content, suggesting they can also participate in aggregation of the full-length protein. Thus for fALS cases in which the fibril-forming segment ^{147}GVIGIAQ^{153} is missing due to early translation termination of SOD1 (21), the remaining three aggregation-prone segments can participate in aggregate formation.

C-terminal segment ^{147}GVIGIAQ^{153} seeds aggregation

Using fibril formation assays, we find that addition of the C-terminal segment ^{147}GVIGIAQ^{153} accelerates the aggregation of wild-type protein (Figure 3.4A). In the native state,

the ^{147}GVIGIAQ^{153} segment is buried in the dimer interface and protected from exposure (Figure 3.1A). Reduction of the disulfide bond results in enhanced mobility of the disulfide loop and of the adjacent ^{147}GVIGIAQ^{153} segment. This in turn weakens the interactions across the dimer interface and destabilizes the β-barrel fold. Indeed, the conditions used in this study favor the monomeric state (22), in which the ^{147}GVIGIAQ^{153} segment is exposed for interaction.

^{147}GVIGIAQ^{153} also accelerated the aggregation of the fALS mutant apoSOD1^{G93A} further suggesting the importance of this segment in SOD1 aggregation. In addition, the wild-type and the mutant protein can template each other's aggregation implies structural similarities between the aggregates. Preformed nuclei of apoSOD1^{G93A} seed the aggregation of wild-type protein (Supplementary Figure 3.8A) and vise-versa (Supplementary Figure 3.8B). Based on this, we can infer that the aggregation of the wild-type protein influences the aggregation of mutant SOD1. Our experiments show that ^{147}GVIGIAQ^{153} segment accelerates the aggregation of both wild-type and mutant protein suggesting that this segment is a common molecular determinant of aggregate formation in sALS and fALS.

Our experiments suggest that fiber formation of SOD1 appears to be initiated by the monomer where the segment ^{147}GVIGIAQ^{153} is unprotected. Although mature SOD1 is a stable dimer, newly translated SOD1 is a copper-free monomer with reduced Cys, in which the aggregation-prone C-terminus is exposed and flexible. Dimer formation is a part of the SOD1 maturation process and facilitated by the copper chaperone for SOD1 (CCS). CCS modifies the nascent SOD1 by inserting copper and oxidizing its intra-subunit disulfide bond; this leads to spontaneous dimer formation (23). Thus any delay in the posttranslational processing of SOD1 by CCS, such as slow protein folding, sequence mutations, or changes in the cellular environment, lead to increased accumulation of unfolded monomeric intermediates (24). Indeed, pathogenic SOD1 proteins derived from either cultured cells or the spinal cords of transgenic mice tend to be metal-deficient (25) and/or lack the disulfide bond (26).

It is possible that various processes, such as abnormally high expression of SOD1 can also result in an increased pool of aggregation susceptible monomers. Normally, the cellular concentration of SOD1 is 30-fold more abundant than that of CCS (at both the RNA and protein levels) (27), so CCS must cycle through the pool of newly translated molecules to activate them. Overexpression of wild type SOD1 in mice caused paralysis (9). The excess SOD1 most likely overwhelms the endogenous CCS system, and the pool of immature wild-type SOD1 molecules aggregate, leading to paralysis which is similar to symptoms in ALS. In short, exposure of the aggregation-prone segments in immature SOD1 could be at the root of both sporadic and familial ALS.

A common pathway for fALS and sALS

Although SOD1-containing aggregates are seen in only a subset of sALS cases, the similarities between SOD1-linked fALS and sALS disease suggest that clarification of the molecular origins of SOD1 aggregation will be informative about ALS cases as a whole. Our hypothesis is that the exposure of one or more fibril-forming segments of SOD1 can lead to aggregation of the full-length protein. Here we identified two segments ^{101}DSVISLS^{107} (located at the Greek Key loop) and ^{147}GVIGIAQ^{153} (located at the C-terminus) that are likely to participate in fibril nucleation and growth of SOD1. fALS mutations located within these segments do not prevent fibril formation, but do alter the fibril morphology and the steric zipper packing. We speculate that *in vivo* exposure of these segments could occur prior to chaperone-assisted maturation of the protein. Overall our data suggest a common molecular origin of fALS and sALS: the exposure and the interaction of aggregation-prone segments at the C-terminus of SOD1.

MATERIALS AND METHODS

Peptide synthesis

GVIGIAQ, VQGIINFE, GSIKGL, and DSVISLS were purchased from CS Bio and Celtek Bioscience Peptides. GIIGIAQ (V148I) and GVTGIAQ (I149T) were purchased from GenScript. The rest of the peptides were synthesized on HMPB-ChemMatrix resin using standard Fmoc solid-phase synthesis conditions (28). 200 μmoles of the first amino acid was coupled to 50μmoles of HMPB-ChemMatrix resin with 200 μmoles coupling reagent 1-(Mesitylene-2-sulfonyl)-3-nitro-1H-1,2,4-triazole (MSNT) and 100μl of the base N-methyl-imidazole dissolved in 3ml of dichloromethane. The reaction was capped and left at room temperature overnight. For a coupling step of the rest of the amino acids, 200 μmoles of Fmoc-protected amino acid (4 eq) and coupling reagent O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU, 4 eq) were weighed into 15ml falcon tube. The amino acid was dissolved in 3.0 ml of 0.1 M 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HOBT) in N-methylmorpholine and 135μl N,N-diisopropylethylamine (DIEA). This mixture was vortexed briefly and allowed to react for at least 1 min. The activated amino acid solution was then added to the fritted/washed polypropylene tube containing the resin. The reaction was heated to 70°C in the Pierce Reacti-Therm heat block for 15min with stirring. After the coupling reaction, the resin was washed three times with dimethyl formamide (DMF). For Fmoc deprotection, 3 ml of 20% piperidine in DMF was added to the resin. The mixture was heated to 70°C in the Pierce Reacti-Therm heat block for 5 min with stirring. After the deprotection reaction, the resin was washed three times with DMF. The cycles of coupling and deprotection were alternately repeated to give the desired full-length peptide. After the final deprotection, the resin was washed three times in DMF followed by three times in dichloromethane and left to dry overnight. Acid-labile side-chain protecting groups were globally removed and peptides were cleaved from the resin by adding 50 μL triisopropylsilane and 50 μL water followed by adding of 4ml trifluoroacetic acid (TFA). The reaction was stirred for three hours. After three hours, the

TFA mixture containing the newly synthesized peptides was dripped into a 50ml tube, and the TFA was evaporated overnight in the fume hood.

Peptide purification

After the synthesis the crude peptides were dissolved in 20% to 30% acetonitrile. Peptides were purified on a preparative C18 column (10µm, 250x20mm, Higgins Analytical) using Reverse Phase HPLC with 9ml/min flow rate. Solvent A was 0.1% TFA dissolved in water. Solvent B was 0.1% TFA dissolved in acetonitrile. The gradient used for peptide purification was 0% to 50% solvent B over 50 minutes. The purity of the synthesized peptides was assessed by C18 analytical column (5µm 250x4.6mm Higgins Analytical) at a flow rate of 1 mL/min and gradient 10-60% B solvent over 50 minutes. Masses of the newly synthesized peptides were verified by MALDI-TOF-MS.

Peptide fibril formation

VQGIINFE, DSVISLS and their corresponding mutants were dissolved in water to a final concentration of 2mM. GVIGIAQ and its corresponding mutants except for RVIGIAQ (G147R) and GVTGIAQ (I149T) were dissolved in water to a final concentration of 1mM. GVTGIAQ (I149T) formed fibril-like aggregates only at 6mM concentration in water. RVIGIAQ (G147R) formed fibrils when dissolved to 1mM in 25mM K phosphate pH 7.0 and 1mM EDTA. Due to the high peptide solubility, KVWGSIKGL and its mutants were dissolved with 0.1M Tris base (pH was not adjusted) to 60 mM. Peptide segments were incubated at 37°C with shaking level 9 in a Torrey Pine Scientific shaker. After 7 to 10 days the samples were examined by TEM.

Crystallization

GVIGIAQ was crystallized using the hanging drop vapor diffusion method: 1 µL of 0.3 mg/mL GVIGIAQ was mixed with 1 µL of reservoir solution. The reservoir solution comprised 1 mL of 1 M sodium acetate pH 4.5 and 1.75 M ammonium sulfate. GVTGIAQ (I149T) was crystallized using the hanging drop vapor diffusion method: 1µl of 6mg/ml peptide dissolved in water was mixed with 2 µl of reservoir solution. The drop was equilibrated over 1 ml of reservoir

solution containing 0.1 M sodium acetate pH 4.5 and 0.7 M hexanediol. DSVISLS was crystallized using the hanging drop vapor diffusion method: after filtering with a 0.2μm filter, 1μL of 3 mg/mL peptide was mixed with 1 μL of reservoir solution, comprising 0.1 M MES pH 6.0, 20% PEG 6000, and 5 mM $ZnCl_2$. Peptide crystals were mounted directly onto the ends of pulled glass capillaries.

X-ray diffraction and data processing

X-ray diffraction data were collected at beamline 24-ID-E of the Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA with a wavelength of 0.8954 Å and at beamline ID13 at the European Synchrotron Radiation Facility in Grenoble, France with a wavelength of 1.0332 Å. Data were collected at 100 K with 5° oscillations.

Indexing of diffraction images was performed using the programs DENZO (29) or XDS (30) Scaling of data was performed using the program SCALEPACK (29). The merged scaled data were imported into CCP4 format with programs from the CCP4 program suite organized under the "CCP4i" interface (31). Molecular replacement solutions were found using the program Phaser (32). In the case of GVIGIAQ the search models consisted of geometrically idealized β-strands with side chain conformations modeled as the most frequently observed rotamers defined in the graphics program COOT (33). The solution of the GVTGIAQ (I149T) was found by Phaser (32) using the structure of GVIGIAQ as a search model. A molecular replacement solution for DSVISLS was determined using an extended beta strand of sequence DSVISAS after expanding dataset with SFTOOLS in the CCP4 suite (31) from P21 to P1. Crystallographic refinements were performed with the programs Refmac (34), Buster, and Phenix (35). Model building was performed with COOT (33) and illustrated with PyMOL (36). Area buried and shape complementarity were calculated with AREAIMOL (37) and SC (38), respectively.

X-ray powder diffraction of GIIGIAQ (V148I)

Aggregates of GIIGIAQ (V148I) grown in 0.1 M sodium phosphate pH 2.5 were pelleted by spinning the sample for 5 minutes at 20,000 x g. The pellet was scooped with a crystallization loop. Data was collected for 5 minutes with 1° oscillations using an in-house Rigaku RAXIS-4++ image plate detector (Rigaku-FRD).

Protein expression, purification and metal removal

Human SOD1WT and SOD1G93A were expressed in the EGy118 strain of Saccharomyces cerevisiae (S. cerevisiae) which lacks the endogenous yeast sod1 gene. The proteins were purified and metals were removed as previously described in (39), except that, in the last step, 25 mM potassium phosphate, pH 7.0, 1 mM EDTA were used as dialysis buffer. After metal removal, apoSOD1WT.and apoSOD1G93A were immediately frozen and stored in -80°C. The protein was used within two weeks of freezing. In Figure 3.5, Supplementary Figure 3.6, Supplementary Figure 3.7 and Supplementary Figure 3.8 are shown assays done with SOD1 purified from S. cerevisiae.

We also used human SOD1 with C6A and C111S substitutions (AS-SOD1) purified from Escherichia coli (E. coli). AS-SOD1 has similar properties to the wild-type protein (40), but lacks the two surface exposed Cys, which can cross link to form larger oligomers (41). The plasmid with AS-SOD1 gene was given by Becky Chan (Prof. Valentine lab). Pro mutations at positions I18P, I35P, I104P and I149P in the AS-SOD1 gene were made using QuickChange Site-Directed mutagenesis Kit (Stratagene).

Proteins (wild-type and mutants) were expressed in the E. coli strain BL21(DE3) in the following way. Cells transformed with pET 22b vector containing AS-SOD1 gene, were grown in LB media containing 100μg/ml ampicillin at 37°C with shaking. Cells were induced at OD600 ~0.6 with 1 mM IPTG. $ZnSO_4$ was added to a final concentration of 0.05mM to aid protein folding. Then, cells were grown at 32°C. After 4 hours, the cells were harvested by centrifugation for 10-15 min at 4600xg.

To release the protein from the periplasm, the cell pellet was resuspended in ice-cold 30 mM Tris-HCl pH 8, 20% sucrose (20-30ml for 1L growth media). Then, cells were stirred slowly at 4°C for 20 min. The cell slurry was spun at 10,000 × g at 4°C for 10 min and the supernatant was discarded. The cell pellet was resuspended in ice-cold water (20-30 ml for 1L growth media) and stirred slowly at 4°C for 20 min. During this step, the periplasmic proteins together with the expressed protein were released into the buffer. The cell suspension was centrifuged at 10,000xg for 10min. Ammonium sulfate was added to the supernatant which contained the expressed protein released by the double osmotic shock (0.326g ammonium sulfate per ml supernatant) and stirred at 4°C for 45min. This solution was centrifuged at 14,000xg for 30min. Then, the supernatant which contained the expressed protein was mounted on a HiTrap Phenyl HP 5ml column (GE Healthcare). Solvent A was 2.5 M ammonium sulfate, 0.15 M sodium chloride, 0.05 M sodium phosphate, pH 7.0. Solvent B was 0.15 M sodium chloride, 0.05 M sodium phosphate, pH 7.0. Proteins were eluted by running 100ml gradients from 0% to 100% at 5ml/min. Finally the protein was loaded on size-exclusion silica G3000 column (Tosoh Bioscience). The column buffer was made of 0.1M sodium sulfate, 25mM sodium phosphate 1mM sodium azide pH 6.5. To remove metals from AS-SOD1 and its mutants we followed the protocol described in (12), except that, in the last step, 25 mM potassium phosphate, pH 7.0, 1 mM EDTA were used as dialysis buffer. Fibril formation assays were done immediately after the metal removal.

Fluorescence assays

Concentrated protein was thawed on ice and filtered through a 0.2µm filter prior to the fibril formation assay. Fibril formation assays were performed with 17 µM apoSOD1WT and apoSOD1G93A in 25 mM potassium phosphate buffer, pH 7.0, 1 mM EDTA, 35 mM Tris(2-carboxyethyl)phosphine (TCEP) and 10 µM ThioflavinT. Under the conditions used, the SOD1 dimer is dissociated, so the molar concentrations of wild-type apoSOD1WT and apoSOD1G93A were calculated using the molecular weight of the monomer. Stock solutions of 0.4 mM GVIGIAQ were made by dissolving the lyophilized peptide in 10mM potassium phosphate pH

7.0, 1mM EDTA. Stock solutions of 0.8mM VQGIINFE, GSIKGL, and DSVISLS were made by dissolving the lyophilized peptide in 10mM potassium phosphate pH 7.0, 1mM EDTA. Peptides were added in the molar ratios indicated in Figure 3.5A and B and Supplementary Figure 3.6 and 3.7. Aggregation assays of apoAS-SOD and its Pro mutants were performed with 35 μM protein in 25mM sodium citrate pH 3.5, 1mM EDTA, 30 mM TCEP and 10 μM ThioflavinT.

Seeding experiments were done using previously formed fibrils (prepared with 25μM protein; see above). 50μl of the pre-formed fibrils were pipetted in a glass NMR tubes and sonicated in a water bath for 15minutes. After sonication the fibrils were diluted twice and 2μl were added for every 100μl reaction volume to make 2% seed solutions (v/v).

All assays were performed in black Nunc 96-well optical bottom plates (Thermo Scientific). Teflon balls (1/8 inch in diameter) were distributed into each well of the 96-well plate. Then, 150-200 μL of solution (four replicates per sample) were pipetted into each. The plate was agitated at 300 rpm with a 3-mm rotation diameter in a Varioskan microplate reader (Thermo) at 37 °C. Fluorescence measurements were recorded every 10-15 min by using λex = 444 nm, λem = 482 nm, with an integration time of 200 μs. Description of data normalization and plotting can be found in the Supporting Information.

Normalization of the fluorescence data

For simplicity, in Figure 3.5 the mean values of ThT fluorescence are displayed. The individual values, which were used to calculate the mean signal shown in Figure 3.5, were re-plotted in Supplementary Figure 6. Similarly, Supplementary Figures 3.7 and 3.8 were made by plotting the individual sample replicates. In all figures, the ThT fluorescence signal was corrected for the background by subtracting the mean fluorescence signal acquired during the first one to two hours of the assay. The fluorescence was also normalized by dividing each measurement by the mean signal collected for two to four hours during which the fluorescent reading had the largest value. If one of the sample replicates had fluorescence reading close to the background, its fluorescence was normalized by dividing it by the mean fluorescence reading

of the remaining replicates. The data was smoothed by substituting each data point with the average value calculated using four data points collected prior and after the point and the point itself. This is equivalent to central moving average with sliding nine data points wide window. The normalized fluorescence signal of the peptides (^{14}VQGIINFE^{21}, ^{33}GSIKGL^{38}, ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153}) incubated without the protein was calculated using the averaged fluorescence signal of maximum fluorescence acquired over two to three hours from the assays done with the largest peptide to protein ratios.

Transmission Electron Microscopy

Negatively stained specimens for transmission electron microscopy (TEM) were prepared by applying 5 μL of sample on hydrophilic 400 mesh carbon-coated formvar support films mounted on copper grids (Ted Pella, Inc.). The samples were allowed to adhere for 3 min, rinsed twice with distilled water and stained for 1 min with 1% uranyl acetate. Grids were examined on either a JEM1200-EX (JEOL) or T12 (FEI) microscope

ACKNOWLEDGEMENTS

We thank Dr. L. Goldschmidt, Dr. M. Chattopadhyay, Dr. R. Nelson, Dr. B. Chan and Prof. J. S. Valentine for discussions, Dr. I. Kourinov, Dr. J. Schuermann, Dr. K. Rajashankar, Dr. N. Sukumar, and Dr. S. Banerjee at APS beam line 24-ID-E and ESRF beamline ID13 for help with x-ray data collection; HHMI, P01 NS049134 , NIH AG029430, NIH P01 and DOE for support to D.E's laboratory and NIH R01 NS39112, VA 1I01BX000506, and the Judith and Jean Pape Adams Charitable Foundation for support to P.J.H.'s laboratory.

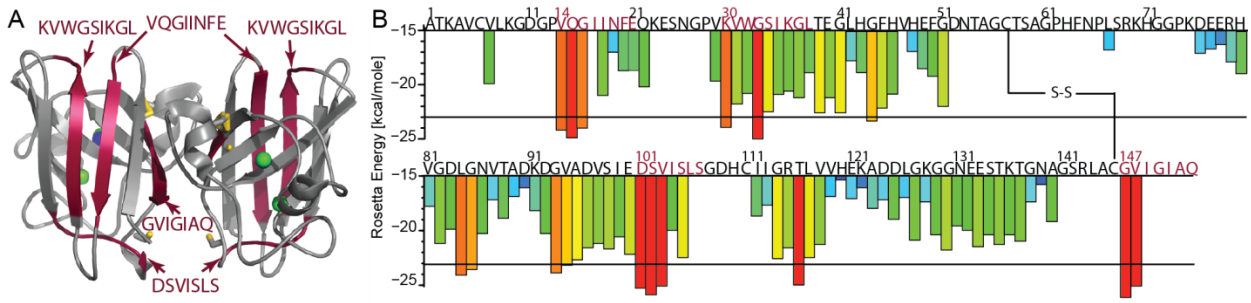


Figure 3.1: The four SOD1 segments predicted to form fibrils form aggregates with fibril morphology. (**A)** Ribbon diagram of SOD1 dimer (PDB code: 2C9U) which shows zinc (green) and copper (blue) atoms coordinated in the metal binding loop. The intra-subunit disulfide bond between Cys57 and Cys146 is shown in gold. The four SOD1 segments predicted to form fibrils are shown in red. **(B)** Diagram of the 3D-profile Rosetta energies (y-axis) calculated for each six-residue segment from the SOD1 sequence (x-axis). Segments predicted to form fibrils are highlighted in red.

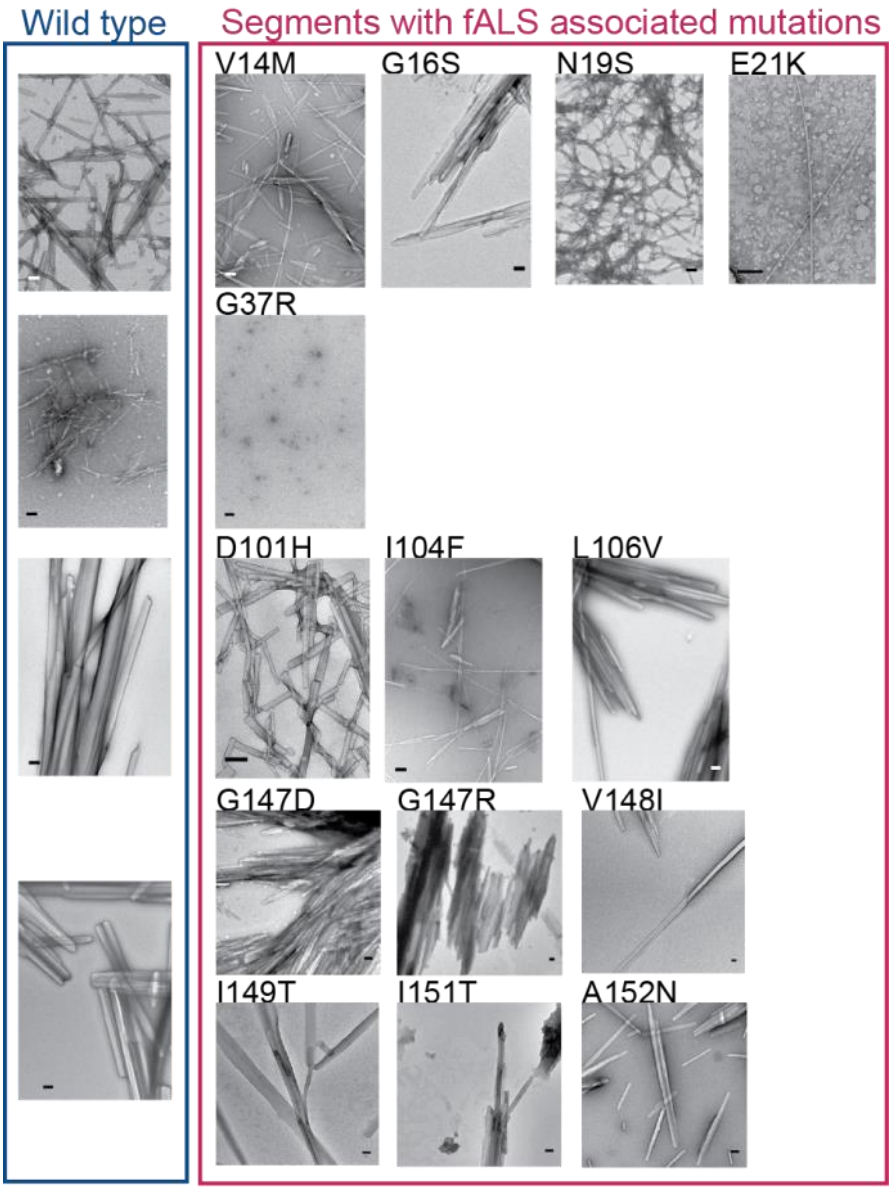


Figure 3.2: Mutations associated with fALS preserve the fibril-forming propensities of ^{13}VQGIINFE^{21}, ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153}. All predicted segments formed fibrils. In addition, thirteen of the fourteen mutated segments that we tested formed aggregates with fibril morphology, confirming the robustness of our predictions. Scale bar is 100nm.

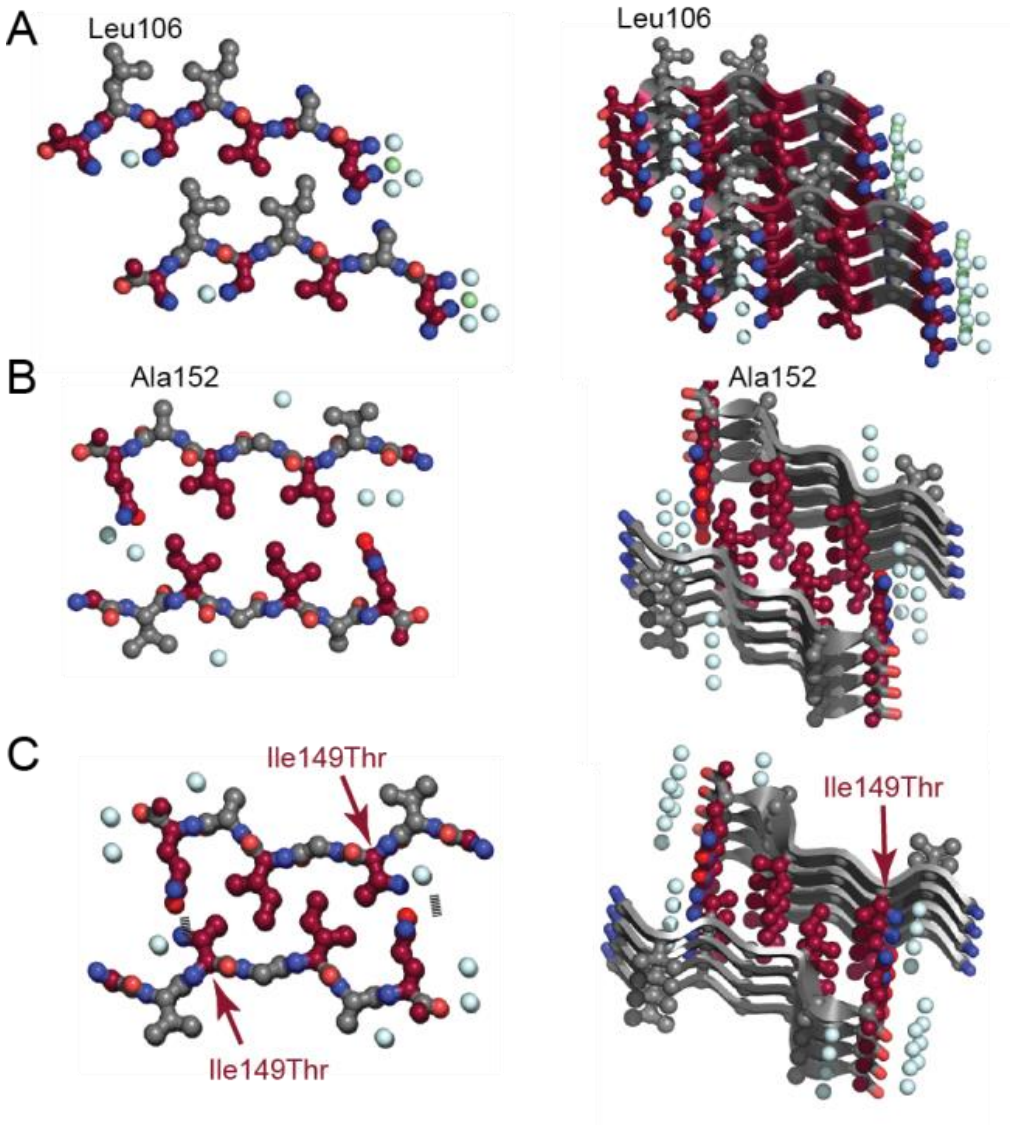


Figure 3.3: Steric zipper structures of (A) ^{101}DSVISLS^{107}, (B) ^{147}GVIGIAQ^{153} and (C) ^{147}GVTGIAQ^{153} (I149T) segments. Off-axis (right) and down the fibril axis (left) views of the dry steric zipper interface of ^{101}DSVISLS^{107}, ^{147}GVIGIAQ^{153} and ^{147}GVIGIAQ^{153}. The dry steric zipper of the ^{147}GVTGIAQ^{153} (I149T) **(C)** compared to the wild-type (B) is stabilized by two hydrogen bonds (dashed lines) formed between the mutated Thr149 and Gln153, showing that this mutation is compatible with the steric zipper. The side chains carbon atoms and Cα atoms of odd numbered residues are colored burgundy to distinguish the two faces of the sheets. This was done to illustrate the face-to-back steric zipper packing of ^{101}DSVISLS^{107} **(A)** and the face-to-face packing of ^{147}GVIGIAQ^{153} **(B)** and ^{147}GVTGIAQ^{153} (I149T) **(C)**.

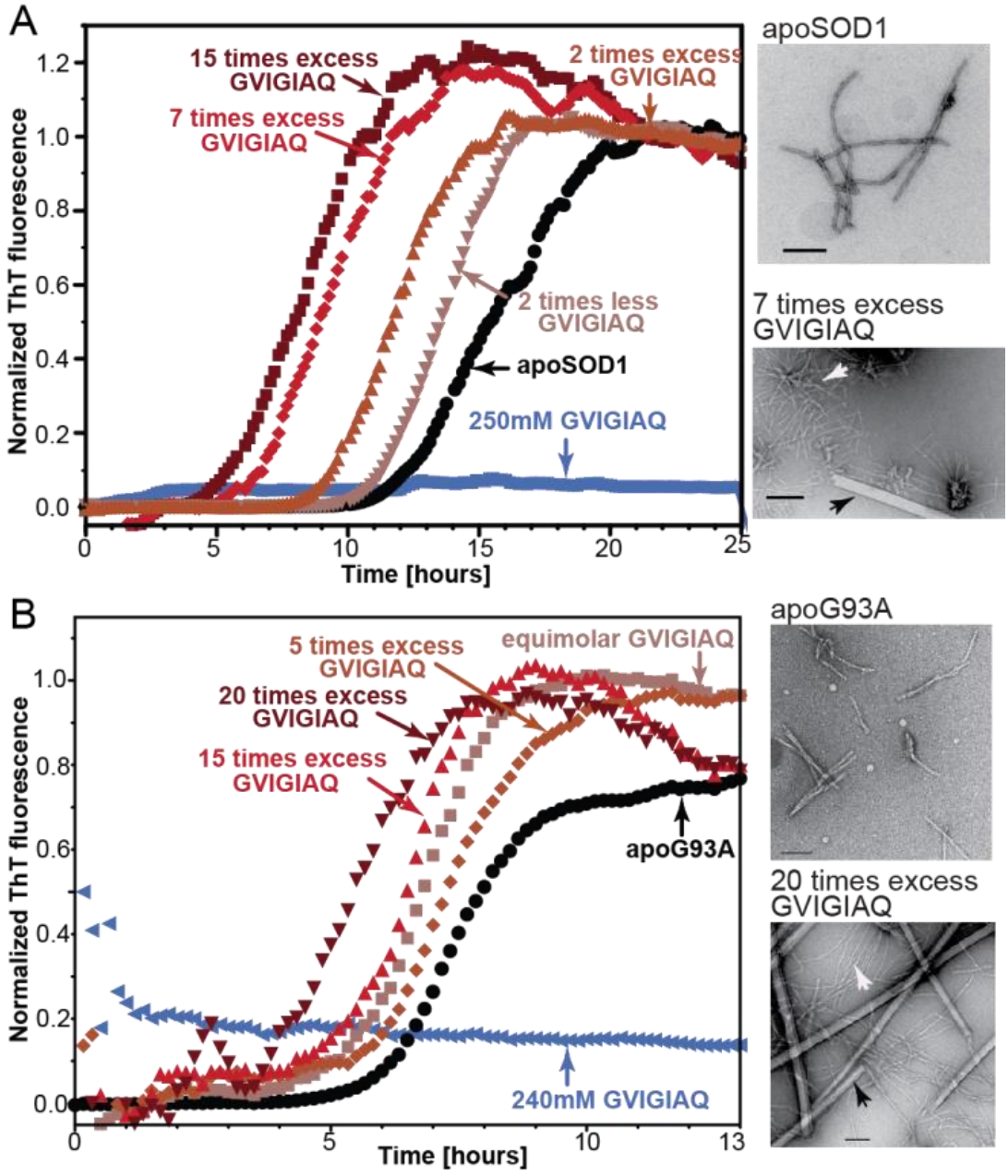


Fig. 3.4: Two C-terminal segment ^{147}GVIGIAQ^{153} nucleates the fibril formation of full-length SOD1. (A) Using a ThT fluorescence assay, we observed that full-length SOD1 co-incubated with ^{147}GVIGIAQ^{153} forms fibrils more rapidly than the full-length protein alone. Adding 15 molar excess of ^{147}GVIGIAQ^{153} shortened the lag time of aggregate formation nearly 50%. To the right are TEM images of the samples taken after the ThT fluorescence signal reached a plateau. Above is a micrograph of apoSOD1^{WT} fibrils. The lower micrograph of

apoSOD1^{WT} co-incubated with the peptide same shows that wild-type apoSOD1 fibrils (white arrows) co-localize with ^{147}GVIGIAQ^{153} needle-like aggregates (black arrows). (**B)** ^{147}GVIGIAQ^{153} seeded the aggregation of mutant apoSOD1^{G93A}. Twenty times molar excess of ^{147}GVIGIAQ^{153} shortened the lag phase of apoSOD1^{WT} aggregation by about one third. ApoSOD1^{G93A} co-incubated with a 20 molar excess of ^{147}GVIGIAQ^{153} (bottom right) formed a mixture of full-length protein fibrils (white arrow) and ^{147}GVIGIAQ^{153} needle-like aggregates (black arrows). Each point in the ThT fluorescence assays is an average of four replicates.

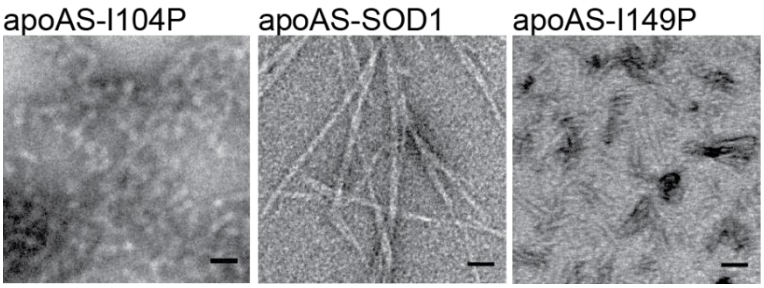
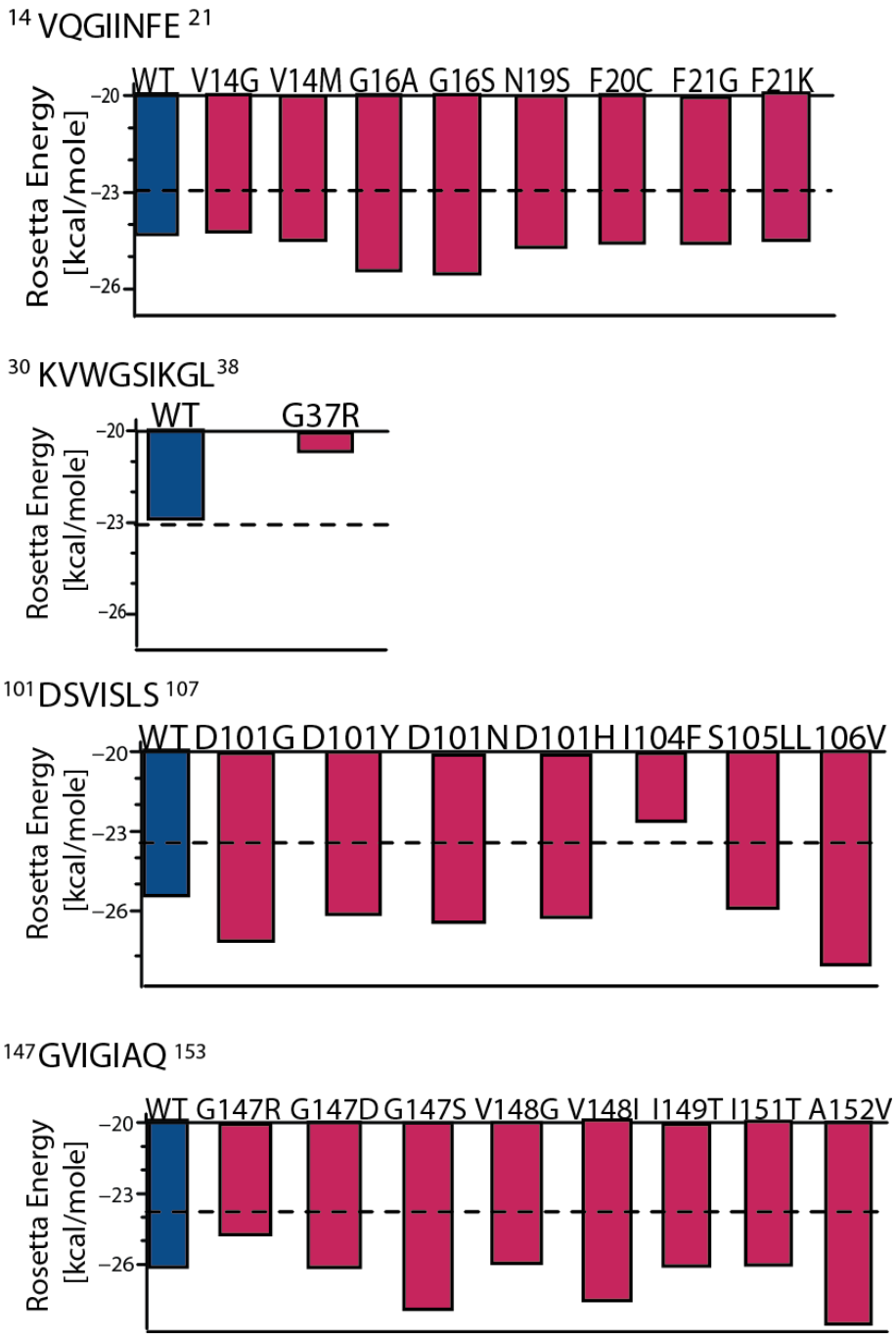
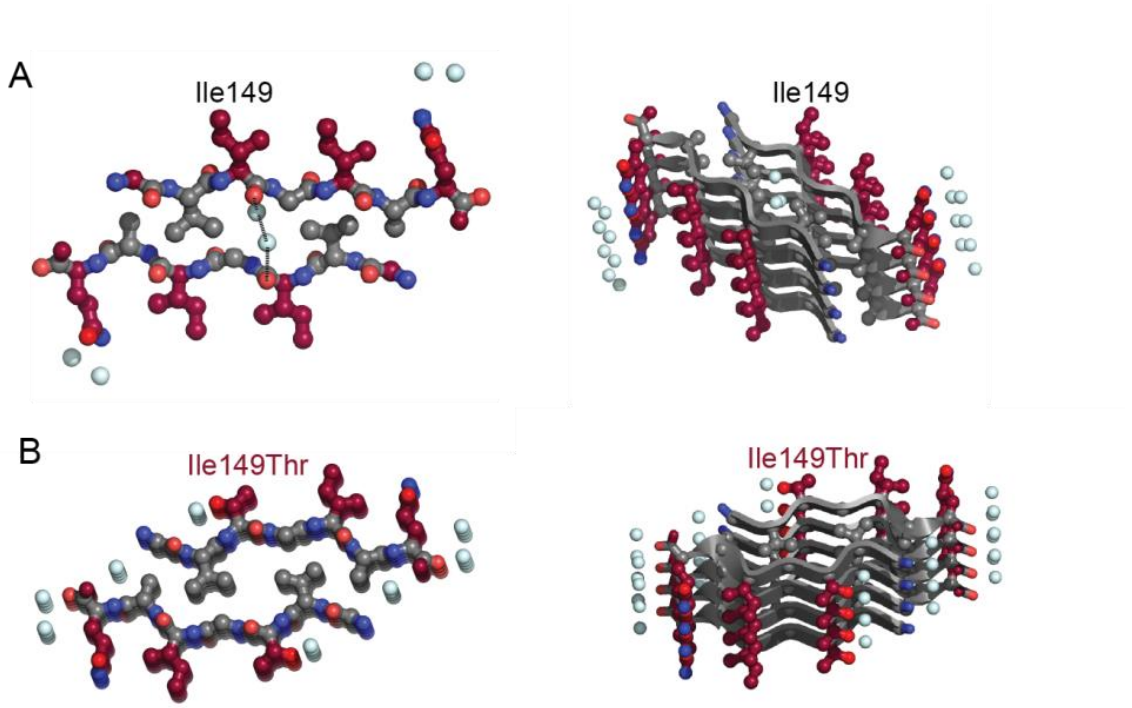


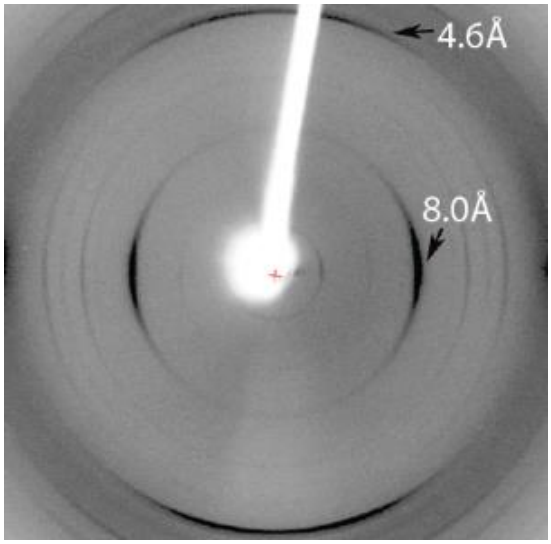
Figure 3.5: Pro substitutions in ^{101}DSVISLS^{107} and ^{147}GVIGIAQ^{153} inhibited aggregation of full-length protein. ApoAS-SOD1^{I104P} formed amorphous aggregates (left), with occasional fibers like those of the wild-type protein (middle). Short rods were observed in apoAS-SOD1^{I149P} sample (right) together with some wild-type looking fibrils.



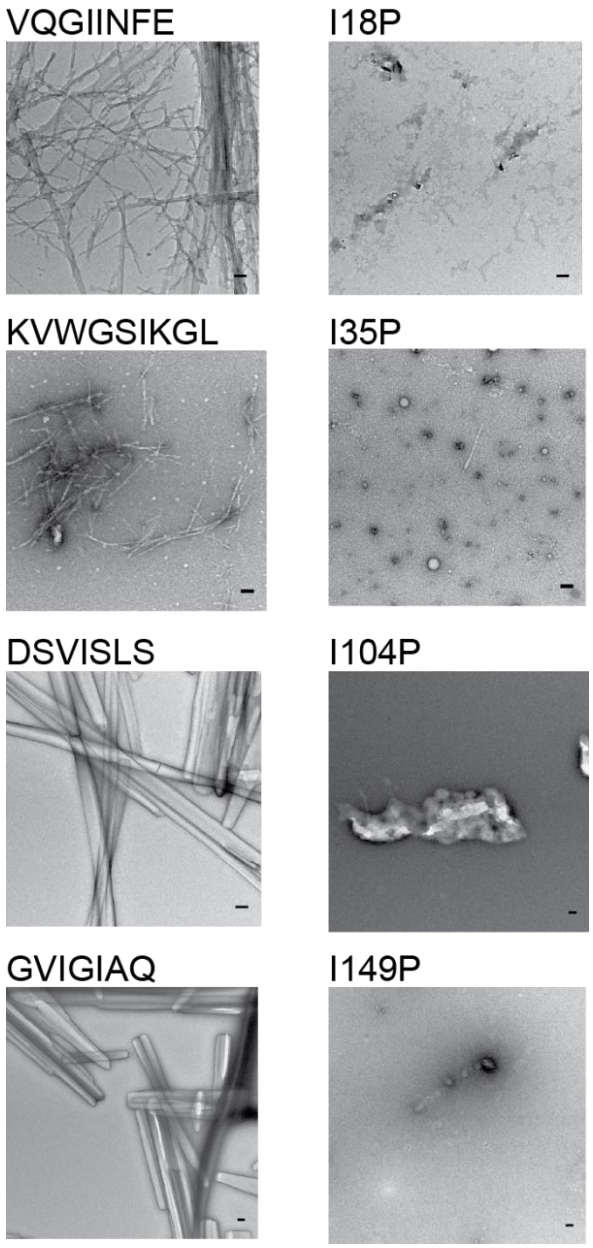
Supplementary Figure 3.1: Mutations associated with fALS preserve the fibril-forming propensities with the exception of G37R in ^{30}KVWGSIKGL^{38} and I104F in ^{101}DSVISLS^{107}. The Rosetta energies of wild-type segments are shown in blue and the energies of segments with fALS mutations are in red. Nearly all mutant segments are predicted to form fibrils with energies below the threshold value of -23 kcal/mol (dashed line).



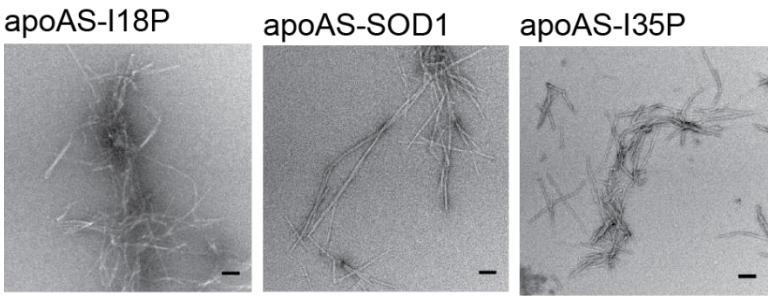
Supplementary Figure 3.2: Wet interfaces of the steric zippers formed of ^{147}GVIGIAQ^{153} and ^{147}GVTGIAQ^{153} (I149T). (A) Val148 of the wild-type segment is intercalated between residues Gly150 and Val151 of the mating β-sheet. The interaction between the two β-sheets is mediated by hydrogen bonds between Gly150 from the mating sheet interfaces and two water molecules. (**B)** In contrast, Val148 of I149T is shifted by about 2Å to face Gly150 of the mating β-sheet and there are no water molecules between the β-sheets.



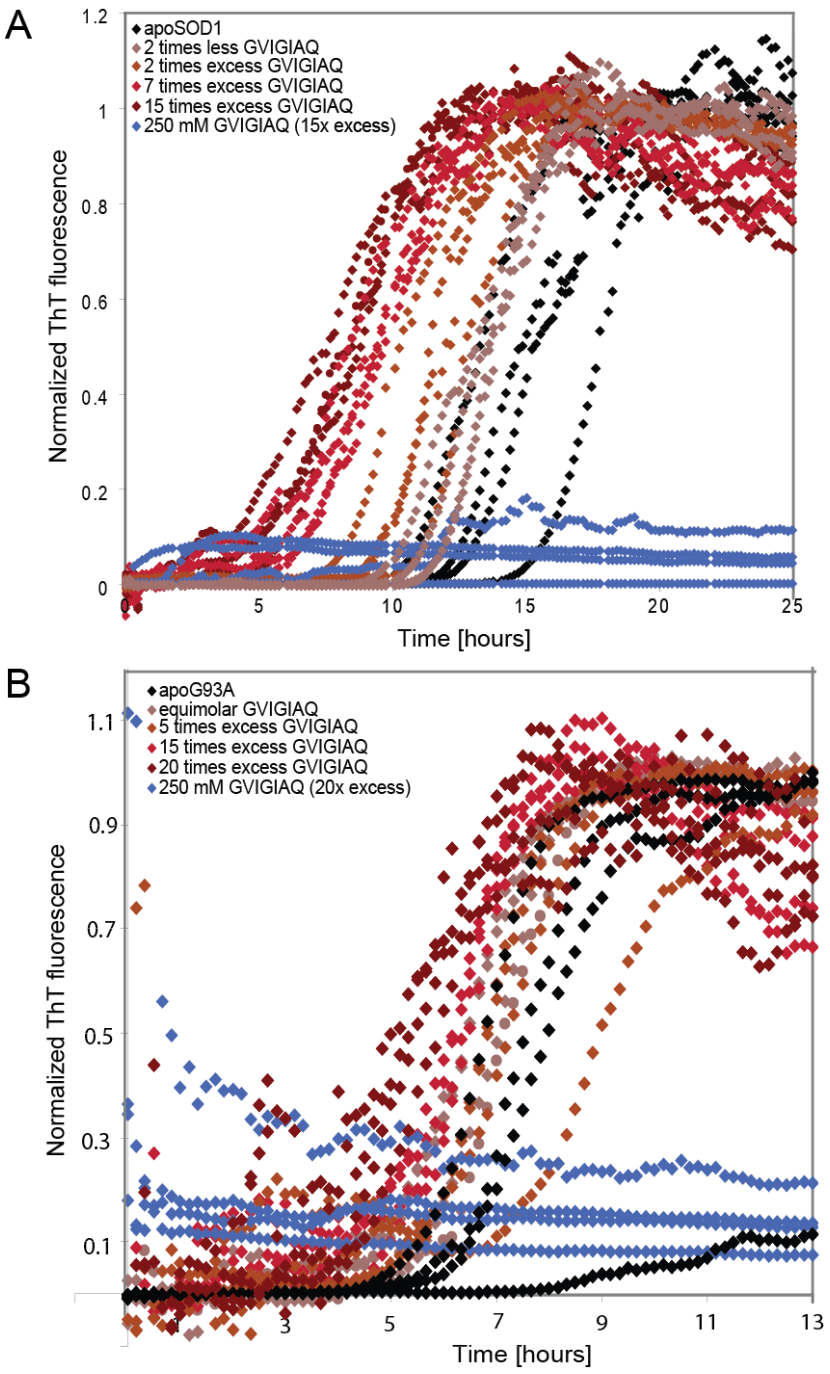
Supplementary Figure 3.3: The V148I mutation in the ^{147}GVIGIAQ^{153} segment does not perturb the propensity of the mutant peptide to form fibril-like aggregates. The fiber x-ray diffraction pattern of ^{147}GIIGIAQ^{153} (V148I) rod-like aggregates grown at pH 2.5 displays a cross-β diffraction pattern typical of amyloid fibrils, with a meridional reflection at 4.7Å and an equatorial reflection at 8.0 Å.



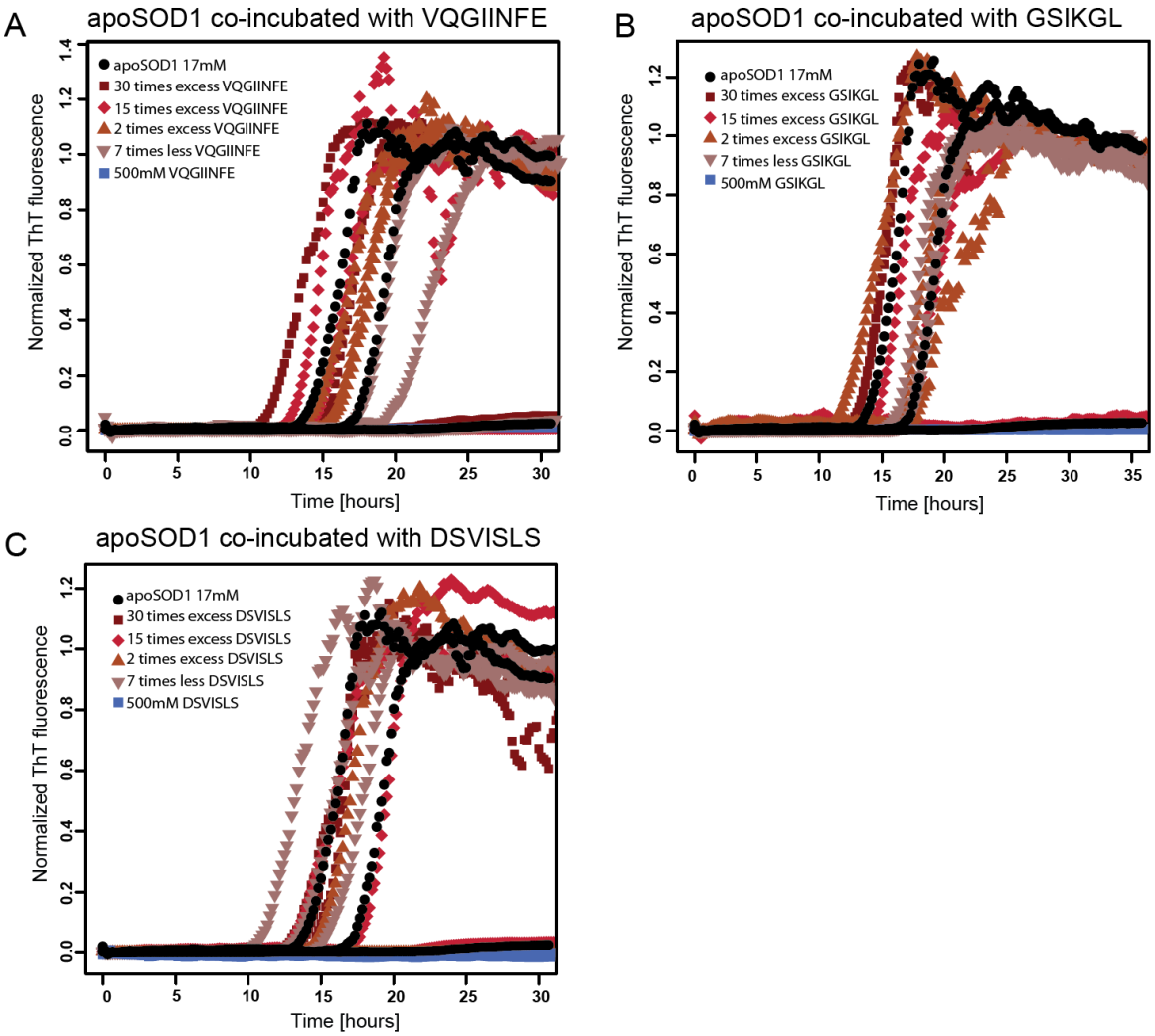
Supplementary Figure 3.4: Single Pro mutation is sufficient to inhibit the aggregation of SOD1 fibril forming peptides. All SOD1 segments with high propensity to form fibrils formed fibril-like aggregates (left panels). Peptides with Pro substitutions in the middle of the same segments formed amorphous looking aggregates which lacked the morphology typical for the amyloid-like fibrils (right panels). Based on this data, we inferred that Pro mutations can block the fibril formation of each SOD1 fibril forming segment in the context of the full-length protein.



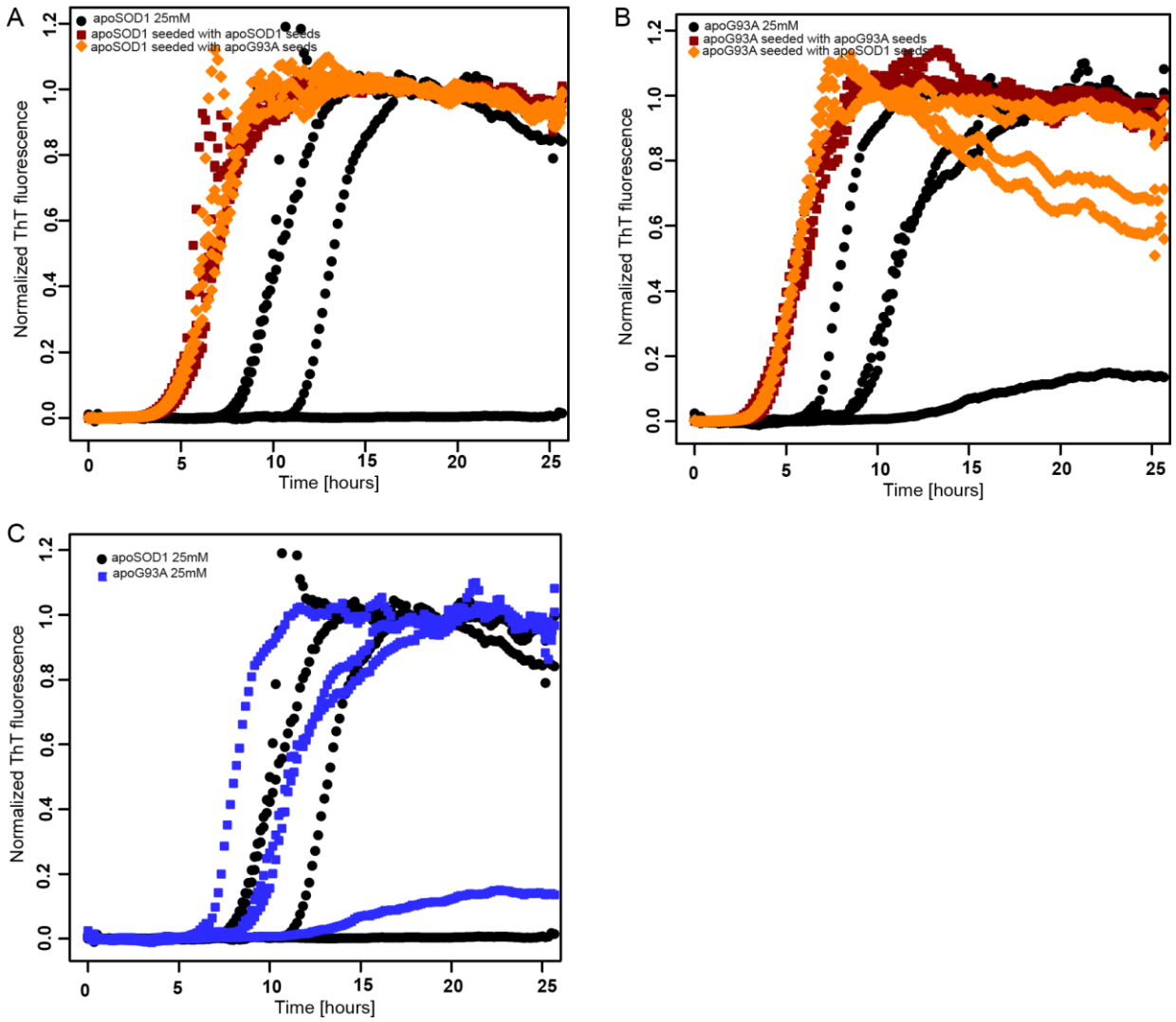
Supplementary Figure 3.5: apoSOD1^{I18P} and apoSOD1 ^{I35P} formed aggregates with morphology similar to apoSOD1^{WT}. TEM was used to assess whether segments ^{14}VQGIINFE^{21} and ^{30}KVWGSIKGL^{38} are important for fibril formation of the full-length protein. Proteins with I18P and I35P substitution that block the aggregation of segments ^{14}VQGIINFE^{21} and ^{30}KVWGSIKGL^{38}, formed fibrils similar in morphology with wild-type protein. Based on this, we can conclude that in the conditions tested, ^{14}VQGIINFE^{21} and ^{30}KVWGSIKGL^{38} are auxiliary rather than the building segments of the fibril core.



Supplementary Figure 3.6: The C-terminal segment ^{147}GVIGIAQ^{153} accelerates the fibril formation of full-length apoSOD1. The individual fluorescent curves representing each of the four sample replicates, which were used to calculate the average fluorescence shown in Fig. 5, are plotted. The C-terminus peptide ^{147}GVIGIAQ^{153} accelerates the fibril formation of both apoSOD1^{WT} **(A)** and apoSOD1^{G93A} **(B).**



Supplementary Figure 3.7: ^{14}VQGIINFE^{21}, ^{33}GSIKGL^{38} and ^{101}DSVISLS^{107} do not nucleate apoSOD1^{WT} fibril formation. **(A)** ^{14}VQGIINFE^{21} does not alter apoSOD1 fibril formation when added 30 times, 15 times, 2 times more or 2 times less than the full-length protein. Similarly, segments ^{33}GSIKGL^{38} **(B)** and ^{101}DSVISLS^{107} (**C**) do not change the fibril formation of the full-length protein. The other segments show no effect on protein aggregation (Fig. S7) but it is possible that under different experimental conditions these other segments can become accessible for self-association. Each sample was done in triplicates and is represented by curves that are same in color.



Supplementary Figure 3.8: apoSOD1^{WT} and apoSOD1 ^{G93A} can cross-seed their fiber formation. (A) The fibril formation of apoSOD1^{WT} is seeded as efficiently with apoSOD1 ^{G93A} nuclei as with tseeds formed with the wild-type protein. **(B)** The reverse it also true. The fibril formation apoSOD1^{G93A} can be cross seeded with apoSOD1^{WT} nuclei. **(C)** apoSOD1^{WT} and apoSOD1 ^{G93A} showed similar rates of aggregation.

Peptide name	Mutation	Rosetta Energy [kcal/mole]
Cluster: VQGIINFE VQGIIN QGIINF GIINFE	**Wild-type**	**-24.4: average** -24.2 -24.9 -24.0
Cluster: GQGIINFE **G**QGIIN QGIINF GIINFE	**V14G**	**-24.4: average** -24.4 -24.9 -24.0
Cluster: MQGIINFE **M**QGIIN QGIINF GIINFE	**V14M**	**-24.3: average** -24.0 -24.9 -24.0
Cluster: VQSIINFE VQ**S**IIN Q**S**IINF **S**IINFE	**G16S**	**-25.6: average** -27.0 -25.5 -24.2
Cluster: VQAIINFE VQ**A**IIN Q**A**IINF **A**IINFE	**G16A**	**-25.4: average** **-25.9** **-25.4** **-24.8**
Cluster: VQGIISFE VQGII**S** Q**G**II**S**F **G**II**S**FE	**N19S**	**-24.8: average** **-25.3** **-24.4** **-24.8**
Cluster: VQGIINCE VQ**G**IIN Q**G**IIN**C** **G**IIN**C**E	**F20C**	**-24.6: average** -24.2 -24.6 -24.4
Cluster: VQGIINFG VQ**G**IIN QGIINF GIINF**G**	**E21G**	**-24.5: average** -24.2 -24.8 -24.6
Cluster: VQGIINFK VQ**G**IIN QGIINF GIINF**K**	**E21K**	**-24.5:average** -24.2 -24.8 -24.5
Cluster: KVWGSIKGL KVWGSI VWGSIK WGSIKG GSIKGL	**Wild-type**	***-22.9: average*** -24.0 *-21.7* *-20.8* -25.0
Segment: KVWGSIK**R**L KVWGSI VWGSIK WGSIK**R**	**G37R**	***-20.9: average*** -24.0 *-21.7* *-17.4*

GSIK**R**L		*-20.4*
Cluster: DSVISLS DSVISL SVISLS	**Wild type**	**-25.5 average** -25.3 -25.9
Cluster: GSVISLS **G**SVISL SVISLS	**D101G**	**-26.5 average** -27.1 -25.9
Cluster: YSVISLS **Y**SVISL SVISLS	**D101Y**	**-25.4 average** -24.8 -25.9
Cluster: HSVISLS **H**SVISL SVISLS	**D101H**	**-25.6 average** -25.4 -25.9
Cluster: NSVISLS **N**SVISL SVISLS	**D101N**	**-25.7: average** -25.5 -25.9
Cluster: DSVFSLS DSV**F**SL SV**F**SLS	**I104F**	***-22.2: average*** -20.5 -23.9
Cluster: DSVILLS DSVI**L**L SVI**L**LS	**S105L**	**-25.2: average** -23.9 -26.4
Cluster: DSVISVS DSVISV SVISVS	**L106V**	**-27.1: average** -26.4 -27.8
Cluster: GVIGIAQ GVIGIA VIGIAQ	**Wild type**	**-26.1: average** -27.1 -25.1
Cluster: RVIGIAQ **R**VIGIA VIGIAQ	**G147R**	**-23.4: average** *-21.6* -25.1
Cluster: DVIGIAQ **D**VIGIA VIGIAQ	**G147D**	**-24.7: average** **-24.2** **-25.1**
Cluster: SVIGIAQ **S**VIGIA VIGIAQ	**G147S**	**-26.0: average** **-26.9** **-25.1**
Cluster: GGIGIAQ G**G**IGIA **GIGIAQ**	**V148G**	**-24.4: average** -25.1 -23.7
Cluster: GIIGIAQ G**I**IGIA **I**IGIAQ	**V148I**	**-25.8: average** -27.1 -24.4
Cluster: GVTGIAQ GV**T**GIA VTGIAQ	**I149T**	**-24.5: average** -24.7 -24.3

Cluster: GVIG<u>T</u>AQ GVIG**<u>T</u>**A VIG**<u>T</u>**AQ	**I151T**	**-24.5: average** -25.5 -23.5
Cluster: GVIGIVQ GVIGI**<u>V</u>** VIGI**<u>V</u>**Q	**A152V**	**-26.7: average** **-27.8** **-25.5**

Supplementary Table 3.1: The effect of fALS-associated mutations on the fibril formation propensity of the SOD1 segments predicted to form fibrils: To compare the fibril formation propensity between segments longer than six residues, we calculated the average Rosetta energy for each cluster (15). The mutated residues are shown in bold and underlined. The segments with Rosetta energies larger than the threshold (-23 kcal/mol) are shown in italic. ^{30}KVWGSIKGL^{38}, had averaged Rosetta energy slightly larger than the threshold energy. This Rosetta energy can be explained with the presence of two six residue segments with unfavorable fibril formation energies. We studied this segment because two six residue segments that are part of ^{30}KVWGSIKGL^{38} are predicted to form fibrils.

Supplementary Table 3.2: Comparison of the steric zippers formed by ^{101}DSVISLS^{107}, ^{147}GVIGIAQ^{153}, and fALS mutant ^{147}GVTGIAQ^{153} (I149T).

Segment Sequence	Shape complementarity of dry interface	Area buried Å^2 dry interface	Sheet-to-sheet Distance dry interface Å	Shape complementarity of wet interface	Area buried Å^2 wet interface	Sheet-to-sheet Distance Wet interface Å
GVIGIAQ	0.81	419	11.4	0.74	462	6.7
GV<u>T</u>GIAQ	0.94	549	9.8	0.85	540	6.7
DSVISLS	0.77	430	9.5	0.77	430	9.5

Supplementary Table 3.3: Statistics for data collection, processing and refinement of ^{147}GVIGIAQ^{153}, ^{147}GVTGIAQ^{153} (I149T) and ^{101}DSVISLS^{107}.

Data Collection	**GVIGIAQ**	**GVTGIAQ**	**DSVISLS**
Space Group	C2	C2	$P2_1$
Resolution (Å)	1.9	1.3	1.4
Unit Cell Dimensions			
a (Å)	50.13	47.79	4.795
b (Å)	4.84	4.79	45.51
c (Å)	18.14	17.69	11.15
β (°)	99.5	109.4	100.4
Measured Reflections	1001	4551	2644
Unique Reflections	415	1053	717
Overall Redundancy	2.4	4.3	3.7
Last Shell Redundancy	2.4	2.4	1.4
Overall Completeness (%)	95.2	95.7	73.6
Last Shell Completeness (%)	89.2	81.6	28.4
Overall R_{sym}	14.4	12.8	18.6
Last Shell R_{sym}	20.6	26.8	16.6
Overall I/σ(I)	6.3	7.3	6.1
Last shell I/σ(I)	4.8	3.6	3.7
Refinement			
R_{work}	23.2	20.18	18.7
R_{free}	24.8	23.32	22.8
r.m.s.d. bond length (Å)	0.010	0.004	0.006
r.m.s.d. bond angle (°)	1.32	1.064	0.92
Number of protein atoms	46	45	50
Number of solvent atoms	3	4	4
Average B factor of protein atoms	16.32	6.01	19.07

REFERENCES

1. Pasinelli P & Brown RH (2006) Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* 7(9):710-723.

2. Tainer JA, Getzoff ED, Beem KM, Richardson JS, & Richardson DC (1982) Determination and analysis of the 2 A-structure of copper, zinc superoxide dismutase. *J Mol Biol* 160(2):181-217.

3. Reaume AG*, et al.* (1996) Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nat Genet* 13(1):43-47.

4. Rowland LP & Shneider NA (2001) Amyotrophic lateral sclerosis. *N Engl J Med* 344(22):1688-1700.

5. Ohi T, Nabeshima K, Kato S, Yazawa S, & Takechi S (2004) Familial amyotrophic lateral sclerosis with His46Arg mutation in Cu/Zn superoxide dismutase presenting characteristic clinical features and Lewy body-like hyaline inclusions. *J Neurol Sci* 225(1-2):19-25.

6. Forsberg K*, et al.* (2010) Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. *PLoS One* 5(7):e11552.

7. Prudencio M, Durazo A, Whitelegge JP, & Borchelt DR (2010) An examination of wild-type SOD1 in modulating the toxicity and aggregation of ALS-associated mutant SOD1. *Hum Mol Genet* 19(24):4774-4789.

8. Wang J*, et al.* (2003) Copper-binding-site-null SOD1 causes ALS in transgenic mice: aggregates of non-native SOD1 delineate a common feature. *Hum Mol Genet* 12(21):2753-2764.

9. Graffmo K. S.*, et al.* (2012) Expression of wild-type human superoxide dismutase-1 in mice causes amyotrophic lateral sclerosis. *Human Molecular genetics* accepted for publication.

10. Wang J*, et al.* (2002) Fibrillar inclusions and motor neuron degeneration in transgenic mice expressing superoxide dismutase 1 with a disrupted copper-binding site. *Neurobiol Dis* 10(2):128-138.

11. Chattopadhyay M*, et al.* (2008) Initiation and elongation in fibrillation of ALS-linked superoxide dismutase. *Proc Natl Acad Sci U S A* 105(48):18663-18668.

12. Fiala M*, et al.* (2010) IL-17A is increased in the serum and in spinal cord CD8 and mast cells of ALS patients. *J Neuroinflammation* 7:76.

13. Roberts K*, et al.* (2013) Extracellular aggregated Cu/Zn superoxide dismutase activates microglia to give a cytotoxic phenotype. *Glia* 61(3):409-419.

14. Chia R, *et al.* (2010) Superoxide dismutase 1 and tgSOD1 mouse spinal cord seed fibrils, suggesting a propagative cell death mechanism in amyotrophic lateral sclerosis. *PLoS One* 5(5):e10627.

15. Goldschmidt L, Teng PK, Riek R, & Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A* 107(8):3487-3492.

16. Sawaya MR, *et al.* (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447(7143):453-457.

17. Ivanova MI, Sievers SA, Sawaya MR, Wall JS, & Eisenberg D (2009) Molecular basis for insulin fibril assembly. *Proc Natl Acad Sci U S A* 106(45):18990-18995.

18. Kayatekin C, Zitzewitz JA, & Matthews CR (2010) Disulfide-reduced ALS variants of Cu, Zn superoxide dismutase exhibit increased populations of unfolded species. *J Mol Biol* 398(2):320-331.

19. Furukawa Y, Kaneko K, Yamanaka K, & Nukina N (2010) Mutation-dependent polymorphism of Cu,Zn-superoxide dismutase aggregates in the familial form of amyotrophic lateral sclerosis. *J Biol Chem* 285(29):22221-22231.

20. Chan PK, *et al.* (2013) Structural similarity of wild-type and ALS-mutant superoxide dismutase-1 fibrils using limited proteolysis and atomic force microscopy. *Proc Natl Acad Sci U S A* 110(27):10934-10939.

21. Orrell RW, *et al.* (1997) Clinical and functional investigation of 10 missense mutations and a novel frameshift insertion mutation of the gene for copper-zinc superoxide dismutase in UK families with amyotrophic lateral sclerosis. *Neurology* 48(3):746-751.

22. Doucette PA, *et al.* (2004) Dissociation of human copper-zinc superoxide dismutase dimers using chaotrope and reductant. Insights into the molecular basis for dimer stability. *J Biol Chem* 279(52):54558-54566.

23. Seetharaman SV, *et al.* (2009) Immature copper-zinc superoxide dismutase and familial amyotrophic lateral sclerosis. *Exp Biol Med (Maywood)* 234(10):1140-1154.

24. Seetharaman SV, *et al.* (2010) Disrupted zinc-binding sites in structures of pathogenic SOD1 variants D124V and H80R. *Biochemistry* 49(27):5714-5725.

25. Lelie HL, *et al.* (2011) Copper and zinc metallation status of copper-zinc superoxide dismutase from amyotrophic lateral sclerosis transgenic mice. *J Biol Chem* 286(4):2795-2806.

26. Sheng Y, Chattopadhyay M, Whitelegge J, & Valentine JS (2012) SOD1 aggregation and ALS: role of metallation states and disulfide status. *Curr Top Med Chem* 12(22):2560-2572.

27. Kawamata H & Manfredi G (2008) Different regulation of wild-type and mutant Cu,Zn superoxide dismutase localization in mammalian mitochondria. *Hum Mol Genet* 17(21):3303-3317.

28. Johnson LM, *et al.* (2012) Enhancement of alpha-helix mimicry by an alpha/beta-peptide foldamer via incorporation of a dense ionic side-chain array. *J Am Chem Soc* 134(17):7317-7320.

29. Otwinowski ZM, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307-326.

30. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26:795– 800.

31. Winn MD, *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.

32. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.

33. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.

34. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.

35. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.

36. DeLano WL (2002) The PyMOL Molecular Graphics System. *http://www.pymol.org*.

37. Anonymous (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760-763.

38. Lawrence MC & Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234(4):946-950.

39. Doucette PA, *et al.* (2004) Dissociation of human copper-zinc superoxide dismutase dimers using chaotrope and reductant. Insights into the molecular basis for dimer stability. *J Biol Chem* 279(52):54558-54566.

40. DiDonato M, *et al.* (2003) ALS mutants of human superoxide dismutase form fibrous aggregates via framework destabilization. *J Mol Biol* 332(3):601-615.

41. Banci L, *et al.* (2007) Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: a possible general mechanism for familial ALS. *Proc Natl Acad Sci U S A* 104(27):11263-11267.

CHAPTER 4

Atomic structure of a toxic, oligomeric segment of SOD1 linked to Amyotrophic Lateral Sclerosis (ALS)

INTRODUCTION

Since Alzheimer's pioneering report in 1906, fibrillar protein deposits have been linked to neurodegenerative diseases. More recently, this link has been challenged by findings that transient, soluble oligomers formed by these proteins are cytotoxic (1-3). Whereas atomic-resolution structures of the spines of amyloid fibrils have shown tightly packed β-sheets with interdigitated side chains(4-9), atomic-level details of toxic oligomers remain elusive. Various reports suggest that toxic intermediates formed by amyloid-forming proteins consist of antiparallel β-sheet rich structures (10-15). These reports utilized chemical cross-linking, analytical size exclusion, electron microscopy and FTIR, but no atomic structure of toxic amyloid oligomers has been reported.

ALS is a debilitating disease, destroying spinal motor neurons and often leading to death within a few years of symptom onset. More than 170 different mutations in SOD1, a metal-binding, homodimeric protein of 153 residues, are found in familial cases of ALS (16). SOD1 aggregation has also been implicated in sporadic ALS pathogenesis (17,18). Most of these SOD1 mutants show little change in enzymatic function suggesting that toxicity derives not from a loss of native function, but from a gain of toxic function (19-21). Transgenic mouse models of the familial mutants show motor neuron degeneration and stain positive for SOD1-containing inclusions, suggesting that protein aggregation is a mode of toxicity (22-24). Enrichment of oligomers has also been observed in cell culture (25, 26) and in transgenic mice (27, 28). However, a causal relationship between the appearance of aggregates and neuronal death has not been conclusively supported, and no atomic structure has been described for toxic oligomers

of SOD1 or any other neurodegenerative-disease-related protein. Here we propose a structure for toxic oligomers formed by SOD1.

RESULTS

Crystal structure of SOD1 residues 28-38 reveals an antiparallel β-sheet oligomer

Based on mutational studies of others (29-32) and our own (see below), we focused on residues 28-38 of SOD1 (with sequence PVKVWGSIKGL) as having the potential to form a toxic amyloid oligomer (Supplementary Figures 4.1 and 4.2). To determine the structure of this segment we needed to increase its solubility. We achieved this by engineering a single residue replacement: proline to lysine at the N-terminal position 28. Rod-like crystals, 5 μm in the shortest dimension appeared overnight and upon further optimization yielded diffraction to 2.0 Å resolution. We determined the phases by single isomorphous replacement with anomalous scattering (SIRAS) from crystals soaked in potassium iodide (Supplementary Table 4.1).

The crystal structure shows a twisted β-sheet built of antiparallel, out-of-register β-strands. Describing its shape, we term it the *corkscrew* (Figure 4.1A). Each b-strand in the sheet contains eight residues, from Lys28 to Ile35. The three C-terminal residues, Lys36, Gly37, and Leu38, adopt a β-hairpin conformation, positioning the C-terminal carboxylate to hydrogen-bond with the N-terminal residue of an adjacent strand. The twist of the sheet is left-handed, as is commonly observed for β-sheets. The sheet undergoes a full turn every 16 strands, with a helical pitch of 71 Å corresponding to the unit cell "c" dimension. Unit cell repeats extend the corkscrew throughout the length of the crystal. Examination of Lys28 reveals that the P28K substitution made to facilitate crystallization affects only crystal lattice contacts (Supplementary Figure 4.4), and suggests that the native sequence would adopt the corkscrew structure seen here. Indeed, SOD1 residues 28-38 assembled preferentially into a corkscrew in our MD simulations (Figure 4.1F and Supplementary Figure 4.5B).

The corkscrew architecture differs markedly from amyloid fibrils. Sheets from adjacent corkscrews do not mate together tightly as sheets do in amyloid fibrils, but instead touch lightly through sparse contacts involving polar and charged side chains lining the exterior of the corkscrew (Lys28, Lys30, Ser34, and Lys36), Trp32 and water-mediated contacts. Hence, unlike amyloid fibrils, the corkscrew has no dry interface between sheets to stabilize its assembly. Instead, the corkscrew assembly is stabilized by weaker hydrophobic forces arising from the concave interior filled with aliphatic side chains of Val29, Val31, Ile35, and Leu38 (Figure 4.1A and Supplementary Figure 4.3).

The absence of a stable amyloid-like dry sheet-sheet interface suggests that fragmentation of the corkscrew could be relatively facile and its subsets of various sizes could fit the definition of soluble oligomers. Indeed, ion mobility mass spectrometry experiments confirm that this SOD1 segment 28-38 forms low molecular weight oligomers in solution similar in cross-section to the crystal structure of the corkscrew (Supplementary Figure 4.6), supporting our hypothesis that the corkscrew represents the structure of a soluble oligomer. Furthermore, the β-sheet rich nature of the corkscrew is a property shared in common with other amyloid-related oligomers such as α-synuclein, amyloid-β and HET-s (11-13, 33). In fact, several amyloid oligomers have been reported to share antiparallel, out-of-register β-strand architecture (13, 34, 35).

The corkscrew shares several structural features with another soluble oligomer, cylindrin, from the non-pathogenic amyloid-forming protein, aB-crystallin(36). Both oligomers are composed of antiparallel β-strands, shifted out-of-register by two residues. Both sheets are highly curved and have an antifacial geometry, meaning opposite faces of the sheet display different sets of side chains, not related by symmetry (37). In both oligomers, individual strands hydrogen-bond to neighboring strands through alternating weak and strong interfaces (Supplementary Figure 4.5A). The strong interface of the corkscrew is composed of 9 inter-chain hydrogen bonds and the weak interface is composed of 7 hydrogen bonds - 1 intra-main chain

and 6 inter-main chain hydrogen bonds. The β-strands in both oligomers have similar values of buried surface area (984 Å^2 vs 943 Å^2) and shape complementarity (0.79 vs 0.74) (Supplementary Table 4.2). The primary difference between the two architectures is that cylindrin is a closed barrel of defined stoichiometry (6 strands) whereas the corkscrew, though highly curved, is not entirely closed and is likely to exist in a range of oligomeric sizes.

Before proposing a role for corkscrew in ALS, it is important to show that the corkscrew can accommodate full-length SOD1 without steric conflict. We modeled the remainder of SOD1 around the corkscrew scaffold keeping the tertiary structure of SOD1 intact everywhere except near the corkscrew (Figure 4.1B). In our model, strands 2, and 3 detach from the native fold, exposing the corkscrew-forming residues 28-38 residing in strand 3. This local unfolding may be triggered by dissociation of the SOD1 dimer and metal depletion. Biochemical studies have noted that metal-depleted monomer is prevalent in patient tissues (38-40) as well as in mouse models (28, 41) and is an intermediate in the conversion of native holo-SOD1 to pathological aggregates. While this model is speculative, it agrees with H/D exchange, molecular dynamics (MD) and mass spectrometry studies by others showing that most familial SOD1 mutants have minimal change in their secondary structure and contain a partially unfolded β-barrel at physiological temperature with local unfolding in β-strand 3 (30, 42).

To probe the role of the corkscrew in SOD1-associated cytotoxicity we took advantage of the sensitivity of the corkscrew structure to mutations at Gly33. The absence of a bulky side chain at position 33 appears essential to form the concave inner surface of the corkscrew (Figure 4.1C). Mutation of Gly33 to a larger residue, such as tryptophan (Figure 4.1D) or valine would introduce severe steric clashes with Val29 and Val31, destabilizing the corkscrew structure. Consistent with this observation, our all-atom MD simulations revealed that the corkscrew was less stable for W33 than for G33, regardless of whether the N-terminus was native proline or lysine (Figure 4.1E and Supplementary Figure 4.4C). Thus, we set out to infer the role of the

corkscrew in cytotoxicity by ascertaining whether G33 mutants (presumably deficient in forming corkscrews) alleviated toxicity of SOD1 familial mutants.

Segment (28-38) is necessary and sufficient for toxicity

We asked if the segment is cytotoxic and if the corkscrew structure underlies its toxicity. To examine cytotoxicity, we used embryonic stem-cell-derived motor neurons expressing green fluorescent protein to facilitate visualization of neuron morphology(43). The corkscrew-forming segment was aggregated and applied to motor neurons and cellular viability was measured upon overnight treatment. We found that viability as measured by MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) reduction assay was reduced by 60% compared to buffer-treated cells at physiological concentrations of 8-100 μM in a dose-dependent manner (Figure 4.2A). In contrast, the segment harboring a single point replacement of G33W, which disrupts corkscrew formation did not induce any toxicity in cells at any concentration tested (Figure 4.2B). We also tested replacement of Gly33 with a less bulky residue, valine and found that it was also sufficient to alleviate toxicity of the segment (Figure 4.2B). The segment harboring the native proline residue is also cytotoxic, and mutation at Gly33 to tryptophan or valine renders the peptide segment non-toxic (Supplementary Figure 4.7).

To further validate our hypothesis that the corkscrew structure is essential for SOD1-mediated toxicity in ALS, we introduced corkscrew-disrupting mutations at G33 in two familial mutants of full-length SOD1 – A4V and G93A (44) and tested for alleviation of toxicity. Proteins were recombinantly expressed, purified and aggregated by demetallation and agitation at 37°C for 12 hours, which produced a mixture of fibrils and oligomers. Motor neurons were treated with aggregated proteins overnight and cellular viability was measured thereafter. A4V and G93A mutants were toxic to the cells at 8 μM, which looked drastically degenerated compared to buffer-treated cells (Figure 4.2C). In contrast, A4V/G33V, A4V/G33W, G93A/G33V and G93A/G33W proteins were non-toxic at the same concentration, and cellular morphologies were indistinguishable from those of the buffer-treated cells. We also quantified the cellular

viability, confirming that A4V and G93A were toxic to cells (Figure 4.2D) and that the introduction of the corkscrew disrupting substitutions at Gly33 alleviates their toxic effects. Together, these data suggest that the segment 28-38 is both necessary and sufficient for motor neuron toxicity.

Toxicity of full-length SOD1 derives from soluble oligomers

We asked if toxicity of full-length SOD1 derives from soluble oligomers or insoluble fibrils. We tested the cytotoxic effects of the non-fibril-forming mutant (I104P) (26). Even though it did not form any mature fibrils (Figure 4.3A), it was toxic to motor neurons, and addition of the corkscrew-disrupting substitution, G33W, alleviated the cytotoxicity (Figures 4.3B and 4.3C). These results suggest that fibril formation is not essential for cytotoxicity.

To identify which species of SOD1 aggregate is toxic, we monitored the toxicity of various SOD1 mutants as their aggregates evolved over time. We found that both familial mutants (A4V and G93A) were toxic to cultured neurons when aggregated for 12-16 hours but extended aggregation for 72 hours rendered the protein non-toxic (Figures 4.3D and 4.3E). A4V/G33W and G93A/G33W were non-toxic irrespective of the duration of aggregation. Negative stain electron microscopy showed abundant fibrils in the samples aggregated for 72 hours (Figure 4.3F) and immunoblotting with the conformational oligomer-specific antibody (A11) suggested samples aggregated for 72 hours contained no oligomers (Figure 4.3G). These results suggest that toxicity of aggregated SOD1 mutants derives from oligomers.

Several reports suggest that wt-SOD1 (WT) aggregates in sporadic ALS (45-47) analogous to mutant SOD1 aggregates in familial ALS; however, it is not clear if WT and mutant SOD1 aggregate by the same mechanism or have the same toxicity. We developed assays to answer these questions in the absence of established cellular and animal models that recapitulate WT toxicity. To monitor the kinetics of fibril formation in these experiments, we used Thioflavin T, a dye that fluoresces upon binding to fibrils. The fibril-forming behavior of WT, G93A, and G33W constructs were similar, displaying a lag time of 2-4 hours, presumably a

period in which small-oligomers form, followed by a short log phase and then a plateau indicative of mature-fibrils (Supplementary Figure 4.8). We then performed the motor neuron cytotoxicity assay with protein aggregates from several time points. We found that WT or G93A protein samples that were allowed to aggregate for up to 12 hours were toxic to the motor neurons, whereas the same protein preparations that were aggregated for 16 hours or more exhibited no toxicity (Supplementary Figures 4.8A and 4.8B). Electron microscopy showed no mature fibrils in samples aggregated up to 12 hours (Supplementary Figure 4.8D). In contrast, samples shaken for 16 hours or longer exhibited large fibril loads and were non-toxic to motor neurons. G33W, which cannot form the corkscrew, was non-toxic regardless of the extent of fibril formation (Supplementary Figure 4.8C). Thus, WT and mutant SOD1 share similar mechanisms of cytotoxicity that depend on exposure and self-assembly of residues 28-38; this exposure may be enhanced by structural perturbations induced by familial mutations.

Corkscrew disruption alleviates defects in a *Danio rerio* (zebrafish) ALS model:

To determine whether corkscrew-disrupting mutations alleviate axonopathies caused by familial SOD1 mutants in vivo, we conducted experiments using a zebrafish model of ALS (48-51) . We expressed A4V SOD1 with and without the corkscrew-disrupting substitution at Gly33 in a zebrafish TDL6 line in which the primary motor neurons are labeled with GFP and the mitochondria are labeled with dsRed (52). We analyzed the axons at 2 days post fertilization (dpf). A4V caused 8% reduction in axon length (Figures 4.4A and 4.4B) as has been previously shown [50], but A4V/G33V-injected fish were significantly better (Figures 4.4A and 4.4B). Additionally, we observed 30% of A4V-injected fish were severely deformed and could not be imaged suggesting that an acute phenotype is lethal (Supplementary Figure 4.9C). In contrast the WT and double mutant expressing fish did not display large mortality. We observed a similar phenotype upon expression of G93A – G93A-expressing zebrafish have 5% reduction in axon lengths while G93A/G33V expressing zebrafish were significantly better (Figures 4.4C and 4.4D).

Defects in mitochondrial assembly and trafficking along with vacuolation (53, 54) and abnormal clustering in neuronal processes (55-58) are established pathological phenotypes observed in transgenic mice, patient-derived cells and other models. However, they have not been reported in any zebrafish model of ALS thus far. Therefore, we analyzed the mitochondrial morphology upon expression of SOD1 familial mutants. Expression of A4V mutant protein caused remarkable mitochondrial pathology characterized by abnormal diffused clustering at the branch points indicative of defective mitochondria (Figure 4.4E) whereas A4V/G33V-expressing fish had a mitochondrial network similar to WT fish. These defects were quantified by measuring the size and fluorescence intensity of the mitochondria confirming that A4V-expressing fish displayed enlarged mitochondria (Figure 4.4F and Supplementary Figure 4.9), which were fewer in number (Figure 4.4G and Supplementary Figure 4.9). We also observed similar defects in the G93A-expressing zebrafish (Figures 4.4H and 4.4I and Supplementary Figure 4.9). Thus, disrupting the corkscrew segment alleviates ALS-linked axonopathies and mitochondrial defects in this in vivo model.

DISCUSSION

Our experiments suggest that segment 28-38 of SOD1 is important for SOD1-mediated toxicity. The crystal structure of this segment revealed an oligomer composed of antiparallel, out-of-register β-strands, which assemble into a corkscrew-like structure. G33V and G33W point mutants that were designed to disrupt the observed oligomer alleviated toxicity of both the isolated peptide and full-length SOD1, which contained toxic familial mutants. In a zebrafish model of ALS, G33V prevented axonopathies and mitochondrial defects, two characteristic features of ALS-linked pathology. Taken together, these results suggest that the corkscrew structure is critical for SOD1-mediated cytotoxicity.

The corkscrew structure explains its oligomeric state and suggests the identity of its potential interacting partners in the cell. The corkscrew is composed of a single twisted sheet

rather than pairs of tightly mated sheets, as observed in 88 steric zipper structures published thus far. The lack of a tight sheet-sheet interface suggests the corkscrew is labile to fragmentation into small, soluble oligomers of varying size, rather than fibers. A clue about the identity of the corkscrew's interacting partners in the cell is offered by examining the functions of its structural homologs. A search for corkscrew homologs in the Protein Data Bank using the DALI server (59) yielded matches with other highly twisted β-sheet proteins such as membrane receptor proteins, enzymes and bactericidal-permeability increasing (BPI) protein (Fig. S10). The twisted sheet seen in the crystal structure of BPI has been shown to bind lipids and destabilize membranes(60). It is conceivable that the cleft seen in the corkscrew structure is important for cytotoxicity potentially as a binding site for lipids.

Our results demonstrate that toxicity derives from the corkscrew oligomers rather than fibrils (Supplementary Figure 4.11). Previously, we have shown that out-of-register oligomers are likely off-pathway from in-register fibril formation due to the large energetic cost of rearrangement of out-of-register oligomers into in-register fibrils(61). Although we cannot ascertain whether the fibrils observed in our experiments are in-register, our results are consistent with this hypothesis. The corkscrew-disrupting mutations of G33V/G33W attenuate cytotoxicity but do not attenuate fibril formation. Cytotoxicity assays of the non-fibril forming mutant (I104P) and the time course assays with the familial mutants, A4V and G93A suggest that toxicity is a property of soluble oligomers and not of large insoluble fibrils. These findings for SOD1 align with the hypotheses proposed by others for amyloid-β and huntingtin that large insoluble aggregates are relatively inert deposits (62-65).

From a molecular perspective, it would be unlikely to find ALS-linked mutations in the 28-38 segment of SOD1 given its structural importance for mediating toxicity. Indeed compared to other regions of SOD1, this segment contains few familial mutations, and no mutations are found in the core of this segment spanning residues 32-36(29). Notably familial mutants including G37R and the rare mutants V29A and V31A are found near the ends of this segment.

From our crystal structure, we infer that all these mutations are compatible with the corkscrew structure, although it is unclear if they actively promote oligomer assembly.

In summary, we have identified an 11-residue segment in ALS-associated SOD1 that is necessary for its cytotoxicity. Our data support the hypothesis that SOD1 forms toxic oligomers composed of antiparallel, out-of-register β-sheet structures involving residues 28-38. This cytotoxic segment may be a target for developing structure-based ALS therapeutics.

MATERIALS AND METHODS

SOD1(28-38) structure determination

Crystals of SOD1(28-38) with P28K substitution were grown by hanging drop vapor diffusion using VDX plates (Hampton Research, Aliso, Viejo, CA). Lyophilized peptide at 98% purity (Genscript Inc.) was dissolved to 50 mg/ml in 50 mM Tris-base buffer. The reservoir solution contained 0.2 M sodium citrate pH 5 and 13% PEG 6000. Crystallization drops were prepared by mixing peptide with reservoir in a 2:1 ratio, in a total volume of 2 µL. Thick needle-like crystals appeared overnight and harvested after 2-3 days growth at ambient temperature. For isomorphous replacement, the crystals were soaked for 30 seconds in a solution composed of approximately 60% KI stock solution (prepared by dissolving solid KI to 0.5 M in reservoir), 30% reservoir solution, and 10% glycerol stock for cryoprotection. The crystals were subsequently flash frozen in liquid nitrogen. Single crystals were mounted with CrystalCap HT Cryoloops (Hampton Research, Aliso Viejo, CA). All data were collected at the Advanced Photon Source (Chicago, IL) on beamline 24-ID-E, which has a 5 µm microfocus beam suitable for the crystals. All data were processed using DENZO and SCALEPACK or XDS (66). SIRAS phasing produced an interpretable map and the structure was built using COOT(67). Model refinement was performed using REFMAC (68) and BUSTER(69). Figures were generated using PyMol(70). Unless stated otherwise, SOD1(28-38) refers to KVKVWGSIKGL.

Molecular Dynamics simulations

We performed all-atom MD simulations of SOD1 and αB crystallin on Anton(71, 72), a special-purpose computer designed to accelerate standard molecular dynamics simulations by orders of magnitude. Protein, water, and ions were represented explicitly, using the AMBER 99SB*-ILDN (73-75) force field and the TIP3P water model(76). Simulations of the SOD1 corkscrew were initiated from the crystal structure comprised of an eight-stranded corkscrew with either a wild-type (KVKVWGSIKGL and PVKVWGSIKGL) or G33W mutant (KVKVWWSIKGL and PVKVWWSIKGL) monomer sequence. Initial structures were placed in a cubic box 76 Å on each side and solvated with approximately 13,000 water molecules and 150 mM NaCl. The systems were minimized and then equilibrated in the NPT ensemble at 1 bar and 300 K for 50 ns with 5 kcal mol^{-1} $Å^{-2}$ harmonic position restraints applied to all non-hydrogen atoms of the protein; these restraints were tapered off linearly over 50 ns. Production runs were initiated from the final snapshot of the equilibration run and performed in the NVT ensemble using a 2.5 fs time step. Non-bonded interactions were truncated at 12 Å. The wild-type (KVKVWGSIKGL and PVKVWGSIKGL) and G33W mutant (KVKVWWSIKGL and PVKVWWSIKGL) corkscrew simulations were run for 16.4 μs, 13.8 μs, 16.6 μs and 15.2 μs, respectively. RMSDs in Figure 4.1E and Supplementary Figure 4.4C were calculated on the hydrophobic core of the corkscrew (Cα atoms of residues 3 to 9, excluding the chains at the end).

SOD1 oligomer assembly simulations were initiated from three monomers of a tandem repeat structure (KVKVWGSIKGL--GG--KVKVWGSIKGL) built in homology to a monomer of the tandem repeat structure of αB crystallin cylindrin (PDB ID 3SGR (36)). These monomers were separated in space by at least 30 Å, placed in a box 62 Å on each side, and solvated with approximately 7,600 water molecules and 150 mM NaCl. System preparation and simulations then proceeded as described above for the SOD1 corkscrew simulations. Each monomer was weakly restrained such that its Cα root-mean-square deviation (RMSD) was within approximately 2 Å of the starting structure. In particular, a flat-bottom harmonic restraint, *U*,

was applied to each monomer such that $U(R) = \frac{1}{2}k(R - R_0)^2$ for $R > R_0$, and $U(R) = 0$ otherwise. Here, R is the RMSD from the initial structure, $R_0 = 2$ Å, and $k = 50$ kcal mol^{-1} Å^{-2}. The flat-bottom RMSD restraints limit the monomers' sampling of very collapsed structures and structures that have significantly reduced β-sheet content, and allow simulations to focus on discovering possible amyloid-like oligomers. We ran 20 assembly simulations beginning from the same starting conformation with different initial velocities drawn from a Maxwell-Boltzmann distribution. Of the twenty simulations, four were run for 4 μs and the rest were run for 1 μs.

The αB crystallin assembly simulations were initiated from three monomers of the tandem repeat structure of αB crystallin (GKLKVLGDVIEV--GG--KLKVLGDVIEV, PDB ID 3SGR). System preparation and simulation protocol were as above, except that we ran 10 assembly simulations in this case. Of these 10 simulations, two were run for 5 μs, two were run for 2 μs, and the rest were run for 1 μs.

Expression and purification of SOD1 constructs

All SOD1 constructs were expressed recombinantly in *E. coli*. All recombinant proteins are based on a SOD1 background strain that carries a C6A, C111S double mutation to simplify the purification. The SOD1 gene was inserted into the pET22b vector from Novagen with NcoI and SalI restriction enzyme digestion sites (gift from Professor Joan Valentine's lab at UCLA). Mutations in the SOD1 gene were made using the QuickChange Site-Directed mutagenesis Kit (Stratagene). The plasmid was transformed into the BL21(DE3)Gold (Agilent Technologies) expression strain. For expression, 10 ml LB + Amp (100 μg/mL) was inoculated from frozen stock and grown overnight. 10 ml of starting culture was added to a 2 L flask of 1 L LB + Amp (100 μg/mL) and grown for 3 hours at 37 °C to OD_{600} = 0.6. IPTG was then added to 1 mM and Zn^{2+} was added to 0.05 mM to induce protein expression, which continued for an additional 3 hours. The bacterial pellet was collected by centrifugation at 4000 rpm for 10 mins.

To purify SOD1 from the bacteria, osmotic shock was used to release proteins including SOD1 from the periplasm. First, the cell pellet was resuspended in 30 mL of 30 mM Tris-HCl pH 8 with 20% sucrose and stirred slowly for 20 mins. The cells were collected by centrifugation at 10,000 × g at 4 °C for 10 mins. The pellet was then resuspended in 30 mL of ice-cold water and stirred slowly for 20 mins on ice to release periplasmic proteins. Next, cell debris was removed by centrifugation at 4 °C for 10 mins at 10,000 × g. The contaminant proteins were then removed by precipitation with ammonium sulfate. 0.326 g/ml $(NH_4)_2SO_4$ was added to the supernatant, and the solution was stirred at 4 °C for 45 mins. The protein was purified out of ammonium sulfate by separation over a phenyl sepharose column. The column was equilibrated with buffer A (2 M $(NH_4)_2SO_4$, 0.15 M NaCl, 0.05 M NaH_2PO_4, pH 7.0) and buffer B (0.15 M NaCl, 0.05 M NaH_2PO_4, pH 7.0). The protein eluted at 6-30% buffer B. The protein was then concentrated and further purified by size exclusion chromatography on a silica G3000 column (Tosoh Bioscience). The column buffer comprised 0.1 M sodium sulfate, 25 mM sodium phosphate, and 1 mM sodium azide, pH 6.5.

SOD1 demetallation (required for aggregation)

The protein constructs purified by size exclusion chromatography were dialyzed in four buffers at 4°C to remove the bound metals. Protein preparations were put in a dialysis cassette (10 kDa molecular weight cutoff) and exchanged with buffer 1 twice (100 mM Na Acetate pH 3.8, 10 mM EDTA pH 8.0), followed by exchange with buffer 2 (100 mM Na Acetate pH 3.8, 100 mM NaCl, 1 mM EDTA pH 8.0), buffer 3 (100 mM Na Acetate pH 5.5, 100 mM NaCl, 1 mM EDTA pH 8.0), and finally buffer 4 (20 mM K Phosphate pH 7.0, 1 mM EDTA pH 8.0). Each exchange lasted 12 hours. To avoid metal contamination, all stock solutions (except for EDTA) were treated with Chelex 100 resin (Bio-Rad) to remove trace metals and all beakers were rinsed with metal-free water.

Protein and peptide aggregation

Full-length SOD1: Following metal removal, proteins were concentrated using a centrifugal concentrator (10 kDa molecular weight cutoff) to 800 μM (monomer) stock concentration and stored at -80 °C in 50 μL aliquots until further use. Aggregation assays were carried out in 1.5 mL microcentrifuge tubes (Fisher Scientific) under acidic conditions with 80 μM SOD1, 50 mM sodium acetate, 1 mM EDTA and 1 mM TCEP, pH 3.5. Preparations were agitated at 900 rpm (in Torrey Pine Shakers) at 37 °C for indicated times. These samples were then dialysed with 10 mM potassium phosphate and 1 mM EDTA, pH 7.0 for 4 hours at 4 °C.

Peptide: Peptide segments were aggregated by incubation at 37 °C for 12 hours with agitation in 50 mM tris-base buffer and added to cultured neurons.

Differentiation of mESCs to motor neurons

Hb9:eGFP mESCs were maintained and differentiated into motor neurons as previously described(43). Briefly, mESCs were first plated on gelatin to remove mouse embryonic fibroblasts (MEFs) prior to differentiation, and then plated in 60 mm bacterial petri dishes in core motor neuron medium (DMEM/F12, Neurobasal, 10% Knockout Serum Replacement, Pen-Strep, Glutamax) to induce embryoid body (EB) formation. Two days later, N2 supplement (1x), retinoic acid (1 μM; Sigma), and SAG (1 μM; Calbiochem) were added to the EBs. Media was changed every two days. After five days post-retinoic acid and SAG addition, Hb9:GFP EBs were dissociated using ice-cold 0.25% Trypsin-EDTA for 6 minutes at room temperature, followed by trituration in L-15 medium (Hyclone). Dissociated motor neurons were plated in core motor neuron medium with GDNF, BDNF, CNTF (all 10 ng/ml; Peprotech) on 96-well plates that were previously coated with poly-L-ornithine (0.01%; Sigma) and laminin (5 μg/ml; BD Biosciences) and incubated for 12-16 hrs before treatment with the SOD1 peptides or proteins. Approximately 80,000 cells were plated per well. All reagents are from Invitrogen, unless otherwise noted.

Cytotoxicity Assays

Aggregated protein preparations were added to cultured neurons at the given final concentration. Unless stated otherwise, all full-length proteins were aggregated for 12 hours before adding to motor neurons at 8 μM final concentration. Motor neurons were incubated for 12-16 hrs and imaged (Axio Observer.D1 microscope, Zeiss), followed by cell viability assay. Cell viability was measured using the commercially available MTT assay (Promega Cat#G4100) according to the manufacturer's protocol with 4 hour incubation with the MTT reagent. All data were normalized using buffer as 100% viability and 0.2% SDS as 0% viability. Experiments were performed in triplicates and repeated a minimum of three times.

Electron Microscopy

Each 3-5 μL sample was spotted directly on freshly glow-discharged carbon-coated electron microscopy grids (Ted Pella, Redding, CA). After 4 min incubation, grids were rinsed twice with 5 μL distilled water and stained with 5 μl of 2% uranyl acetate for 1 min. Excess uranyl acetate was removed by blotting and specimens were examined on a T-12 electron microscope at an accelerating voltage of 80 kV. Images were recorded digitally by wide angle (top mount) BioScan 600W 1 × 1K digital camera (Gatan, Pleasanton, CA).

Dot Blot Assay

3-5 μL sample was spotted on a nitrocellulose membrane (Trans-Blot, Bio-Rad) followed by blocking with 10% fat-free milk in 1X TBST buffer (50 mM Tris, 150 mM NaCl, 0.05% Tween20) for 30 mins. Membranes were incubated with primary antibody SOD100 (Enzo Life Sciences) at 1:1000 dilution, A11 (EMD Millipore) at 1:500 diluted in 5% fat free milk in TBST buffer at room temperature for 1 hour. The membranes were washed three times in TBST buffer before incubating with HRP-linked secondary antibody - anti-mouse HRP secondary antibody (Life Technologies) at 1:10000 dilution, or anti-rabbit HRP secondary antibody (Life Technologies) at 1:10000 diluted in 5% fat-free milk in TBST buffer at room temperature for 1 hour. Pierce ECL Western Blotting Substrate, (Thermo Scientific, #32209), a chemiluminescent

substrate was used for the detection of horseradish peroxidase (HRP) on the immunoblots following the manufacturer's instructions. A detailed protocol can be found at the following URL,

https://tools.thermofisher.com/content/sfs/manuals/MAN0011536_Pierce_ECL_West_Blot_Subs_UG.pdf

Thioflavin-T (ThT) Assay

Fibril formation assays were performed with 80 μM protein concentration in 50 mM sodium acetate buffer, 1 mM EDTA, 1 mM TCEP, pH 3.5 and 10 μM ThT – conditions identical to those used for aggregating proteins for the cytotoxicity assays, but with the addition of ThT. All assays were carried out in black Nunc 96-well optical bottom plates (Thermo Scientific). Plates were agitated at 300 rpm in 3-mm rotation diameter in a Varioskan microplate reader (Thermo) at 37 °C. Fluorescence measurements were recorded every 30 mins using λ_{ex} = 444 nm, λ_{em} = 482 nm, with an integration time of 200 μs.

Mass Spectrometry

Peptide segments were dissolved in 20 mM ammonium acetate buffer (pH = 7) to a final concentration of 200 μM and incubated at room temperature without agitation. Experiments were carried out at time t_o and every day for a week and ion mobility was used to calculate the cross sections of oligomers. The peptide samples were nano-electrosprayed on a home-built, high-resolution ion mobility mass spectrometer (77) consisting of a source ion-funnel, a two-meter long drift cell, an exit funnel, a quadrupole mass analyzer and detectors. The ions were stored and focused in the source funnel and subsequently pulsed into the drift cell filled with helium gas at high pressure. A drift voltage was applied across the cell to create a weak electrical field. The ions moved forward under the influence of the electrical field, and at the same time experienced the drag force due to multiple collisions with helium buffer gas. An ion traveled with a constant velocity that reflected its size and shape. At the end of the drift cell, the ions were collected by an exit funnel where they were steered and focused into a quadrupole mass

analyzer and detected as a function of arrival time in the arrival time distributions (ATDs). By measuring ATDs at different pressure to voltage (P/V) ratios, the absolute mobility of the ion of interest could be measured:

$$t_A = \frac{l^2}{K_0}\frac{273.16\,K}{760\,torr \cdot T}\cdot\frac{P}{V} + t_0\ (Eq.1)$$

In Eq. 1, l is the length of the drift cell, P is the gas pressure in torr, V is the voltage across the cell, and t_o is the time the ions spend outside the drift cell before reaching the detector. The collision cross section σ can then be approximated from K_o (78)

$$\sigma \approx \frac{3q}{16N}\left(\frac{2\pi}{\mu k_B T}\right)^{\frac{1}{2}}\frac{1}{K_0}\ (Eq.2)$$

Here, N is the buffer gas density, μ is the reduced mass of the collision system (He+ion) and k_B is Boltzmann's constant. The cross section contains information about the three dimensional structure of the ion. In the ATDs shown in Fig. S6, underneath each experimental peak is a dotted line representing the peak shape expected for a single conformer (79).

Zebrafish

All zebrafish (*Danio rerio*) were maintained in accordance with standard laboratory conditions (80). The University of California, Los Angeles (UCLA) Chancellor's Animal Research Committee approved all experiments performed on zebrafish. The TDL6 (Tübingen driver line) zebrafish used in all experiments were identified in a screen for developmentally regulated enhancers that drive tissue-specific expression. Gal4-driven GFP expression marks primary motor neurons. Mitochondria were marked by a TOL2-mediated insertion of a Gal4-UAS-MLSDsRed construct (52).

RNA production and injections

SOD1 constructs used for zebrafish expressions were the full-length sequence with Cys6 and Cys111 intact. The constructs (WT, G93A, G93A/G33V, A4V, A4V/G33V were subcloned into pcGlobin2 vector and the constructs were linearized using XbaI to generate RNA. The

Translation efficiency of the RNA of all constructs was tested by TnT® Quick Coupled Transcription/Translation System (Promega) using the manufacturer's instructions. Construct G93A/G33V has an additional mutation of I18G designed to prevent steric clashes due to the presence of G33V in the native SOD1 barrel. As a control, we also tested the cytotoxicity of this construct in cell culture and found that it was non-toxic similar to G33V construct (Fig. S9B). RNA was transcribed using the mMessage Machine T7 kit (Ambion) and the mRNA was diluted to 500 ng/μL and 1000 ng/μL aliquots and stored at -80°C. 1 μL mRNA at 500 ng/μL (WT, G93A, G93A/G33V) and at 1000 ng/ μL (WT, A4V, A4V/G33V) was injected into TDL6 embryos from timed mating at the 1-4 cell stage. Embryos were incubated in 1x E3 buffer at 28.5 °C.

Analysis of axonopathy

At 2 dpf (days post fertilization) zebrafish larvae microinjected with SOD1 constructs were anesthetized in 0.01% tricaine and imaged using a Leica MZ16F fluorescent stereoscope at 11.5X magnification. Observers were blinded and the length from the exit point of the spinal cord to the tip of the motor axon was measured for 5 axons per embryo corresponding to axons 12-16 - using Fiji Simple Neurite Tracer. Experiments were repeated multiple times and a minimum of 72 embryos were assessed for each construct.

Analysis of mitochondrial defects

The mitochondrial network from the base of the spinal cord to the tip of the motor axon was imaged on a Leica TCS SP8 confocal microscope at 10X magnification with a 1.5X optical zoom. Images were processed with the Leica AF software and mitochondria size was measured for 5 axons per embryo using Fiji Image J Particle Analysis with the described algorithm (81). Before quantification, fluorescence from the spinal cord and regions outside of the larvae body were excluded. The program allowed automatic detection of isolated fluorescent spots by threshold and watershedding of binary formatted images without introducing user-bias. Spots with a circularity of 0.5-1 were selected to have their intensity and area quantified, where 1 is a complete circle. Intensity was quantified for each spot via overlapping gray levels.

Protein expression in zebrafish

Protein expression was confirmed by western blotting (Supplementary Figure 4.9). Zebrafish embryos were pooled from each group at 2 dpf and washed twice in Ringer's buffer (NaCl 13 5 mM, KCl 5 mM, $MgSO_4$ 1mM, K_2HPO_4 0.4 mM, Glucose 5.5 mM, HEPES 20 mM pH 7.4). They were subsequently deyolked and dechorionated and lysed in RIPA buffer (Sigma) (volume equals twice the number of embryos in µL) with Halt protease inhibitor cocktail EDTA-free (Thermo Scientific) added. Lysates were homogenized by sonication and protein concentration was determined using the BCA protein assay kit (Thermo Fisher Cat #23252). Equal amounts of protein were loaded on 12% Tris-glycine PAGE gels and transferred onto a PVDF membrane using an iBlot dry blotting system (Thermo Fisher) at 20 V for 6 mins. Membranes were washed with methanol followed by water and then blocked with 10% fat free milk in TBST for 1 hr at room temperature and washed three times with TBST buffer. Membranes were incubated with primary antibodies – SOD100 (Enzo Life sciences) at 1:500 dilution in 5% milk in TBST and incubated for 1hr at room temperature or anti β-actin (Abcam) at 1:1000 diluted in 5% milk in TBST and incubated overnight at 4°C. Membranes were first probed with SOD100 and then stripped using Restore PLUS Western Blot stripping buffer (Thermo Scientific) and subsequently probed with anti β-actin. Blots were washed thrice with 1X TBST and subsequently incubated with the secondary anti-rabbit HRP secondary antibody (Sigma) at a 1:10000 dilution in 5% milk in TBST for 1hr at room temperature. Pierce ECL Western Blotting Substrate, (Thermo Scientific, #32209), a chemiluminescent substrate was used for the detection of horseradish peroxidase (HRP) on the immunoblots following the manufacturer's instructions.

Statistical Significance

All statistical analyses were done using GraphPad Prism 7.0.

Antibodies used

The following antibodies were used in this study. SOD100 (EnzoLife sciences) at 1:1000 dilution, A11 (EMD Millipore) at 1:500 dilution, anti β-actin (Abcam) at 1:1000, anti-mouse HRP secondary antibody (Life technologies) 1:10000 dilution, anti-rabbit HRP secondary antibody (Life technologies) 1:10000 dilution.

ACKNOWLEDGEMENTS

We thank Lisa Johnson, David Borchelt, and Joan Valentine for discussions; Hamilton Trinh, Michael Collazo, Duilio Cascio, and staff at Argonne Photon Source (APS), Northeastern Collaborative Access Team beamline 24-ID-E. The beamline is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector is funded by a NIH-ORIP HEI grant (S10 RR029205). APS is 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. We thank HHMI, DOE, and NIH (AG029430) for support to DSE; UCLA Broad Center for Regenerative Medicine and Stem Cell Research and NINDS (NS072804), Muscular Dystrophy Association (92901) and California Institute for Regenerative Medicine (RB1-01367; RB5-07480 grants to BGN; RT307678 grant to CMK); NIGMS (GM61721) grants to CMK; NIH (AG047116) and NSF (CHE-1301032) grants to MTB. SS was supported by the Whitcome Pre-Doctoral fellowship; KLA by the UCLA CMB Training program (Ruth L. Kirschstein NIH GM00785), the UCLA-California Institute for Regenerative Medicine Training Grant, and a UCLA Graduate Division Dissertation Year Fellowship; RN by a Larry L. Hillblom Foundation fellowship and CJ was a Beckman Research Scholar. Atomic coordinates and structure factors for the corkscrew have been deposited in the Protein Data Bank as 5DLI and 5IIW.

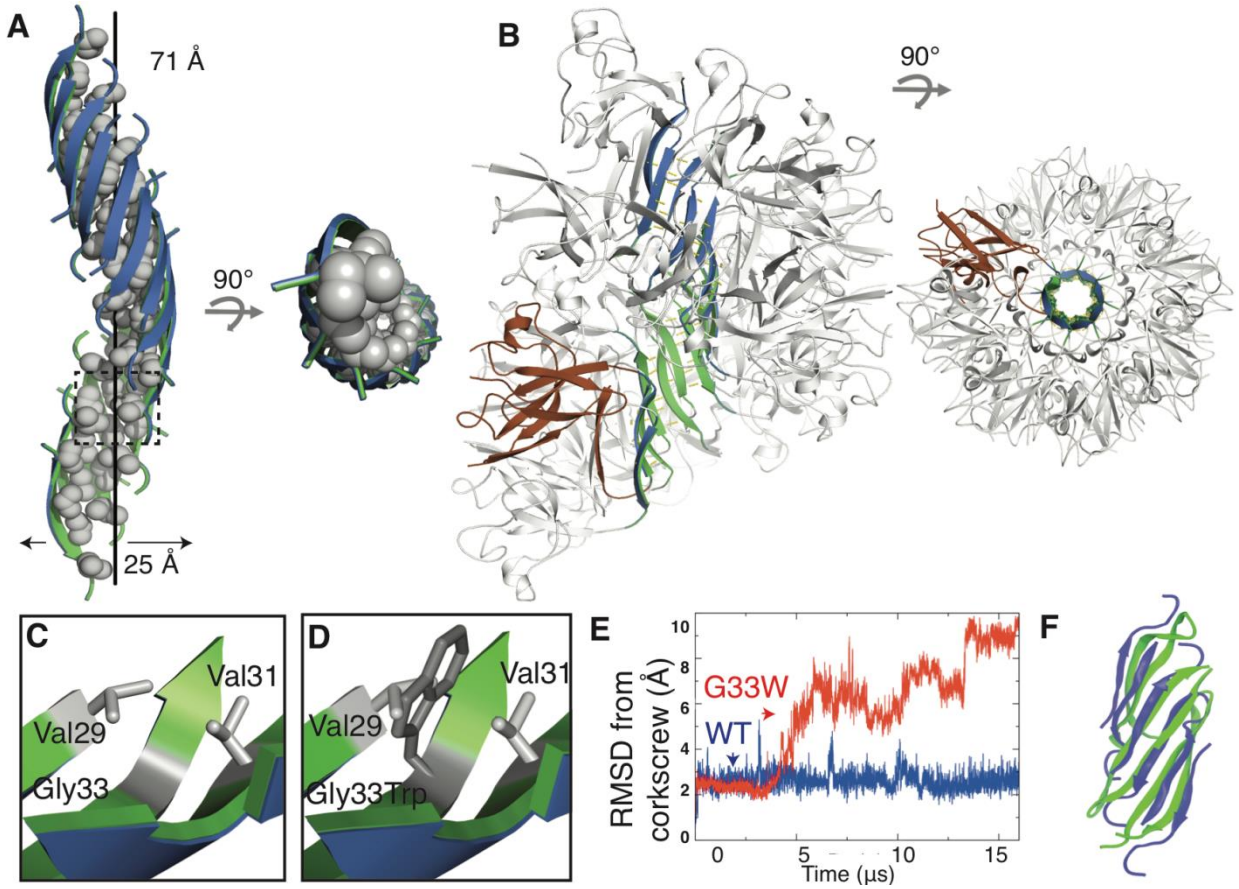


Figure 4.1. Structure of an 11-residue segment derived from SOD1 in its oligomeric state. (A) 2.0 Å resolution crystal structure of the SOD1 segment with sequence KVKVWGSIKGL shows antiparallel, out-of-register β-strands forming a continuous left-handed helix. Sixteen strands form one complete turn of the helix, with 25 Å outer diameter and 71 Å pitch. The hydrophobic interior lined with valine, isoleucine, and leucine side chains (shown in spheres) excludes water molecules, as shown in side and top views. (B) Model of a full-length SOD1 toxic oligomer. This model contains 16 protomers. Strands 2 and 3 of each protomer detach from the native fold to form the corkscrew spine (green interior, blue exterior) as observed in the crystal structure of residues 28-38. The remaining protein decorates the exterior of corkscrew retaining much of the SOD1 native structure. The model is illustrated in the same two orientations as (A). One monomer is colored brown for clarity. Note that the segment 28-38 is buried in the core but it is exposed at the ends. (C) Gly33 is essential for creating the concave

inner surface of the corkscrew. The lack of side chain on Gly33 permits the close approach of hydrophobic side chains of Val29 and Val31 located on β-strands bordering opposite sides of Gly33. (D) Unavoidable steric clash results from mutating Gly33 to tryptophan. (E) All-atom MD simulations of the corkscrew-forming segment suggest that introduction of the G33W substitution destabilizes the structure. Blue and red curves correspond to Cα root-mean-square deviations (RMSD) from the corkscrew crystal structure in MD simulations of 8 chains of the corkscrew segment (KVKVWGSIKGL) and G33W mutant segment (KVKVWWSIKGL), respectively. The structure of the corkscrew segment remained stable throughout the length of the simulation, whereas the G33W mutant deviated from the corkscrew structure. (F) The SOD1 segment 28-38 preferentially assembled into a corkscrew-like structure in MD simulation. MD simulations of weakly restrained monomers of SOD1 spontaneously assembled into a corkscrew-like structure. A snapshot of an assembled corkscrew-like structure from the MD simulations (green) is overlaid onto the crystal structure (blue). As a control, we found that monomers of the cylindrin-forming segment of αB crystallin spontaneously assembled into a cylindrin structure using the same simulation protocol.

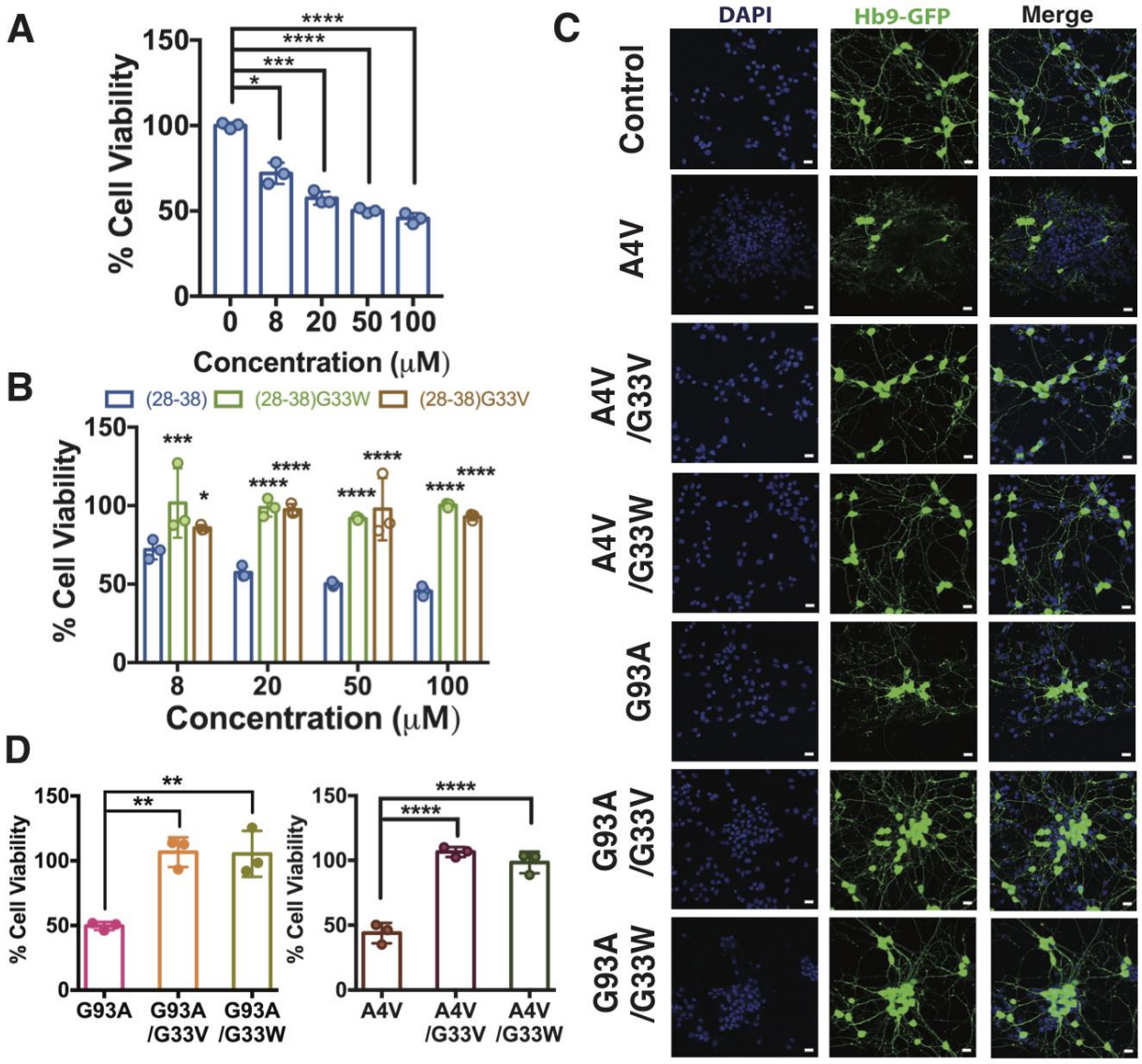


Figure 4.2. Corkscrew forming segment 28-38 is necessary and sufficient for cytotoxicity. (A) Cell viability of motor neurons measured by MTT reduction assay shows that the corkscrew segment (KVKVWGSIKGL) is toxic to primary motor neurons in a dose-dependent manner. Results shown as Mean±SD (n=3). Symbols represent individual values of triplicates and bars represent average values. Statistical significance was analyzed using two tailed T-tests with Welch's correction. (B) Corkscrew-forming segment (28-38) harboring single point substitutions at Gly33 (G33V and G33W) is non-toxic to motor neurons. All peptide

segments were prepared identically and motor neurons were treated with different final concentrations. Statistical significance of G33V and G33W mutants was compared to the 28-38 segment by two-way ANOVA. (C) Hb9-GFP-labeled motor neurons treated with 8 μM aggregated full-length familial mutants (A4V and G93A) lose neurites but the corresponding corkscrew-disrupting mutants (G93A/G33V, G93A/G33W, A4V/G33V and A4V/G33W) are non-toxic and neurons look healthy. (Scale bar = 20 μm) (D) Cell viability measured by MTT reduction assay confirming that the familial mutants A4V and G93A are toxic and substitution of Gly33 with valine or tryptophan renders the protein non-toxic. Results shown as Mean$\pm$SD (n=3). Symbols represent individual values of triplicates and bars represent average values. Statistical significance was analyzed by one-way ANOVA. (*$p < 0.05$,**$p < 0.01$,***$p < 0.001$,****$p < 0.0001$).

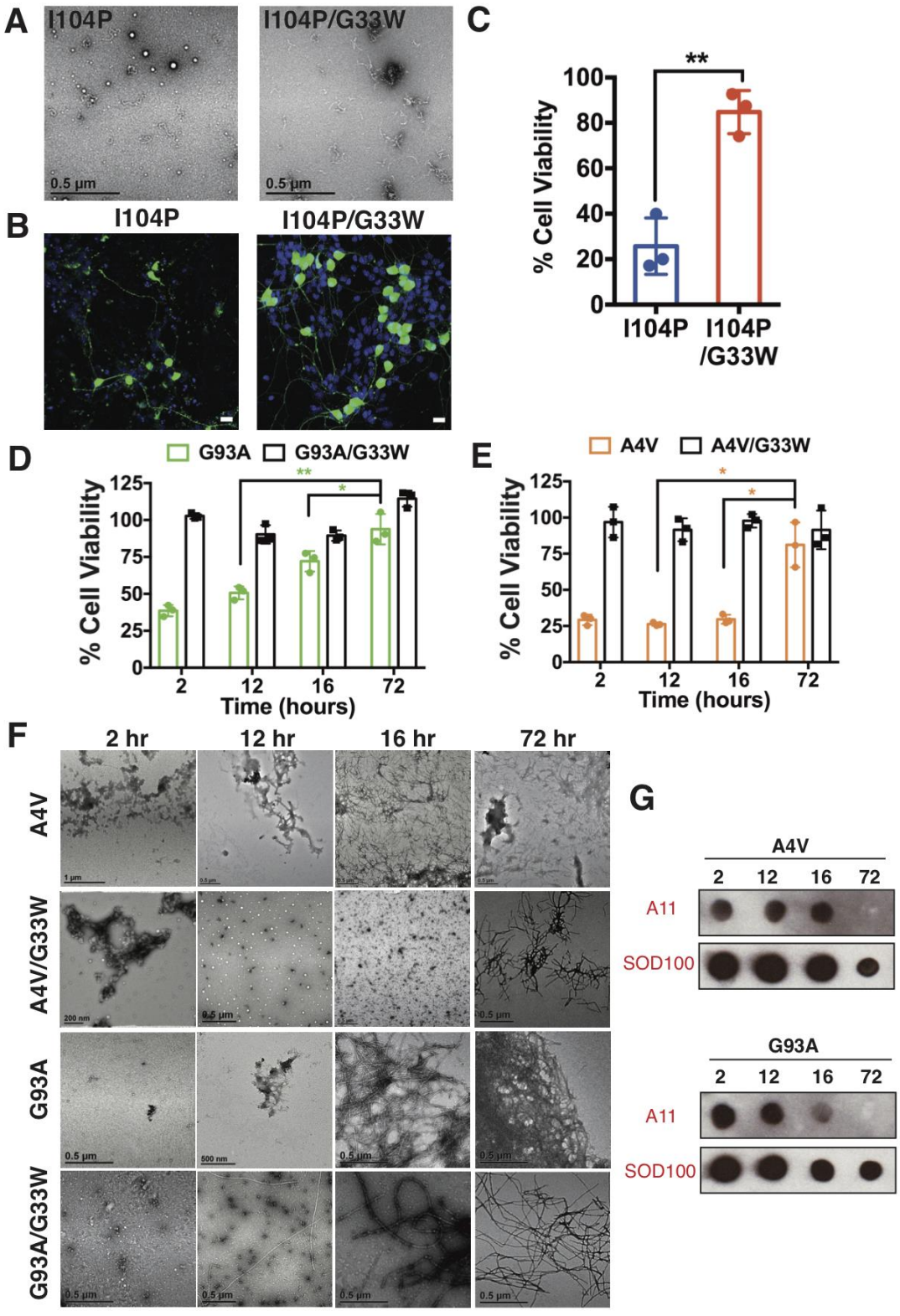


Figure 4.3. Toxicity of full-length SOD1 derives from soluble oligomers. (A) Electron micrographs of a non-fibril forming SOD1 mutant (I104P) and the corresponding double mutant

(I104P/G33W) show some aggregates but no large fibrils. (B) Motor neurons treated with I104P lose neurites and have shrunken cell bodies (left) but I104P/G33W treated cells look healthy. (C) Cell viability measured by MTT reduction assay confirmed that I104P is toxic and I104P/G33W is non-toxic. Statistical significance was analyzed using a two-tailed T-test with Welch's correction. (D,E) The toxic properties of SOD1 mutants depend on the duration of aggregation. A4V and G93A mutants aggregated for 12-16 hours are toxic to motor neurons while extended agitation for 72 hours renders the proteins non-toxic. The corkscrew-disrupted proteins (A4V/G33W and G93A/G33W) are non-toxic irrespective of the duration of aggregation. Results shown as Mean$\pm$SD (n=3). Symbols represent individual values of triplicates and bars represent average values. Statistical significance was analyzed by two-tailed T-tests with Welch's correction. (*$p < 0.05$,**$p < 0.01$,***$p < 0.001$,****$p < 0.0001$). (F) Representative electron micrographs of various preparations of the familial mutants, A4V and G93A and the double mutants, A4V/G33W and G93A/G33W. Some large aggregates can be seen at 2-16 hour time points but no fibrils, whereas at 72 hours all constructs show large fibril loads. (G) Immunoblots of the familial mutants aggregated for different time points. Samples aggregated for 12-16 hours are A11 positive and both proteins when aggregated for 72 hours lose A11 reactivity. SOD100 was used as loading control.

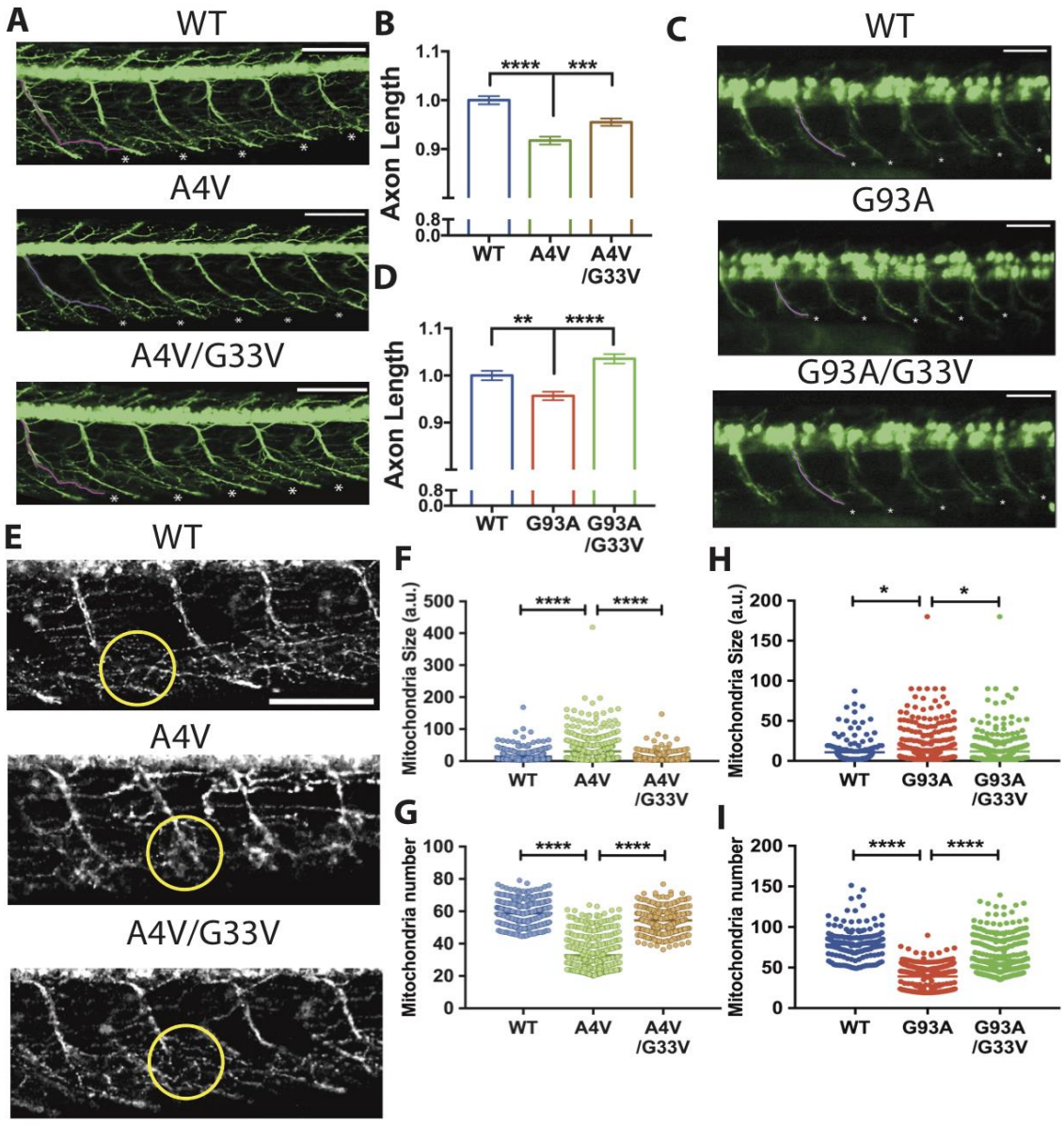
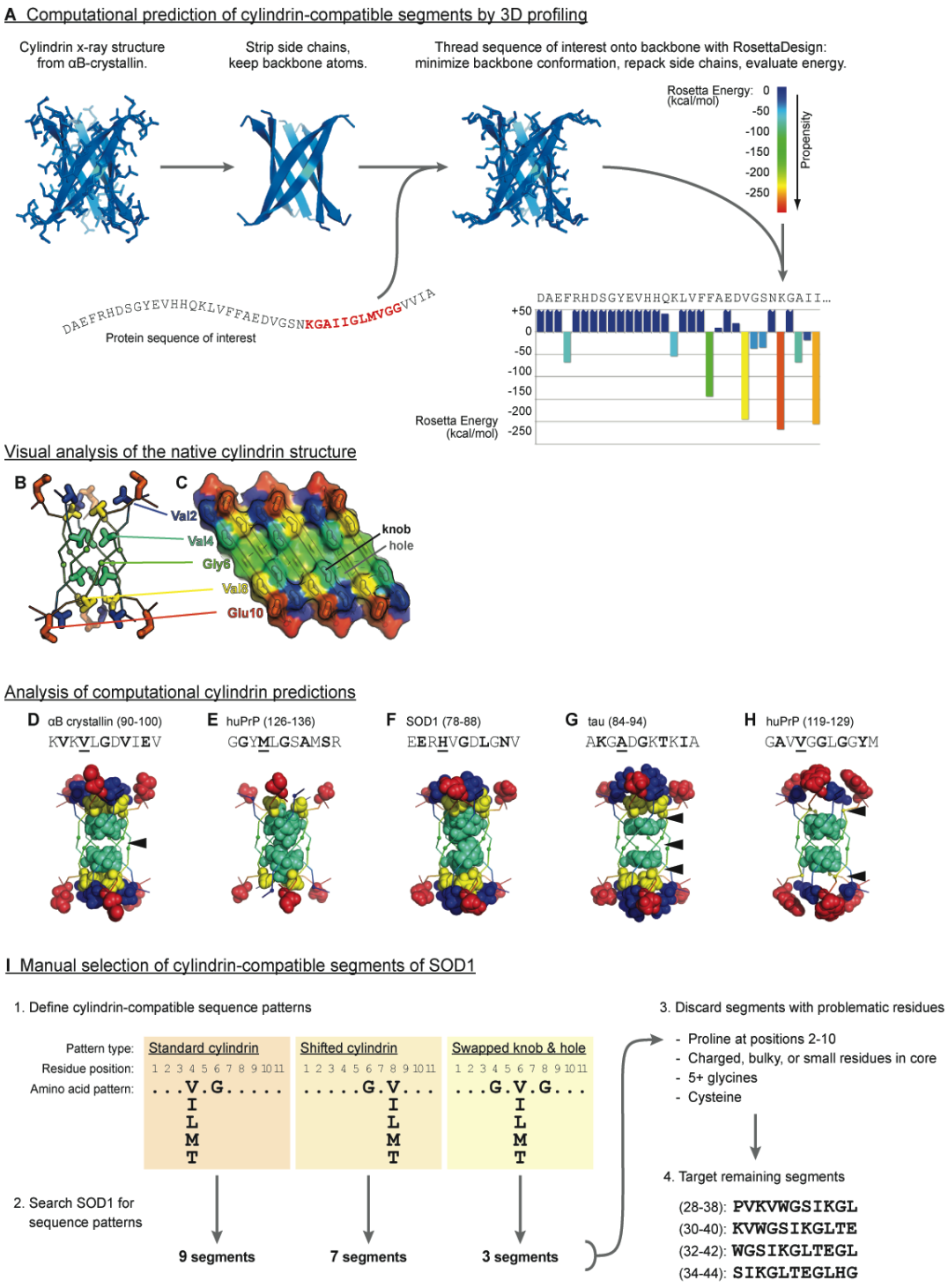


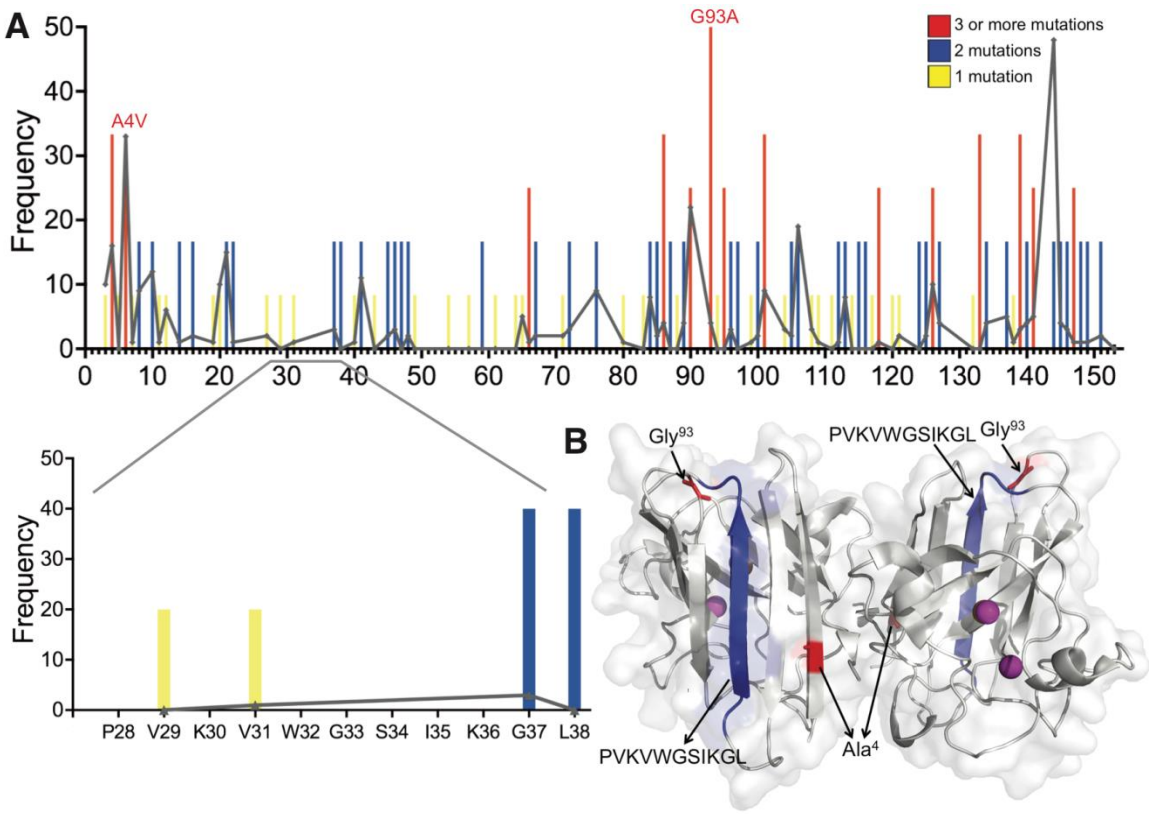
Figure 4.4. Corkscrew-disrupting substitution of G33V alleviates axonopathies in a *Danio rerio* (zebrafish) ALS model. (A) Micrographs of zebrafish embryos at 2 dpf show reduction in axon lengths of A4V–expressing embryos, whereas corkscrew-disrupted A4V/G33V–expressing embryos have significantly better axon lengths. The first axon is highlighted in pink for clarity and axons measured for quantification are marked by an asterisk. Scale bar 100 μm. (B) Quantification of axon lengths show that A4V-expressing embryos have shorter axons than WT-expressing embryos. The corkscrew-disrupting substitution G33V

alleviates the defect. Results shown as Mean±SEM relative to WT for at least 72 embryos. Statistical significance was analyzed by one-way ANOVA. (C) Micrographs of zebrafish embryos at 2 dpf show reduction in axon lengths of G93A–expressing embryos, whereas corkscrew-disrupted G93A/G33V–expressing embryos have significantly better axons. The first axon is highlighted in pink for clarity, and axons measured for length are marked by an asterisk. Scale bar 100 μm. (D) Quantifications of axon lengths show that G93A-expressing fish have shorter axons than WT-expressing fish. The corkscrew-disrupting substitution, G33V alleviates the defect. Results shown as Mean±SEM relative to WT for axons 12-16 of at least 73 embryos. Statistical significance was analyzed by one-way ANOVA. (E) Micrographs show A4V-expressing zebrafish have impaired mitochondria which are clustered at the branch points (encircled) as compared to WT. The mitochondrial network of A4V/G33V-expressing fish is similar to WT. (F,G) Quantitative analysis of the mitochondrial network shows A4V-expressing fish have larger mitochondrial size (30.17 a.u.) and diffused clustering (fluorescence intensity 32 a.u.) in the axons indicative of defective fission, whereas A4V/G33V-expressing fish have healthy mitochondria (10.57 a.u. size and 54 a.u. fluorescence intensity) similar to WT (14.08 a.u. size and 58 a.u. fluorescence intensity). Symbols represent individual measurements made for each group. Statistical significance was analyzed by one-way ANOVA. (H,I) Quantitative analysis of the mitochondria network show G93A-expressing fish have larger, more diffusely clustered mitochondria (14.68 a.u. size and 39 a.u. fluorescence intensity) indicative of defective fission. Mitochondrial network of G93A/G33V (11.68 a.u. size and 59 a.u. fluorescence intensity) expressing fish is similar to WT (10.78 a.u. size and 78 a.u. fluorescence intensity). Statistical significance was analyzed by one-way ANOVA. Symbols represent individual measurements for each group. (*$p < 0.05$,**$p < 0.01$,***$p < 0.001$,****$p < 0.0001$).

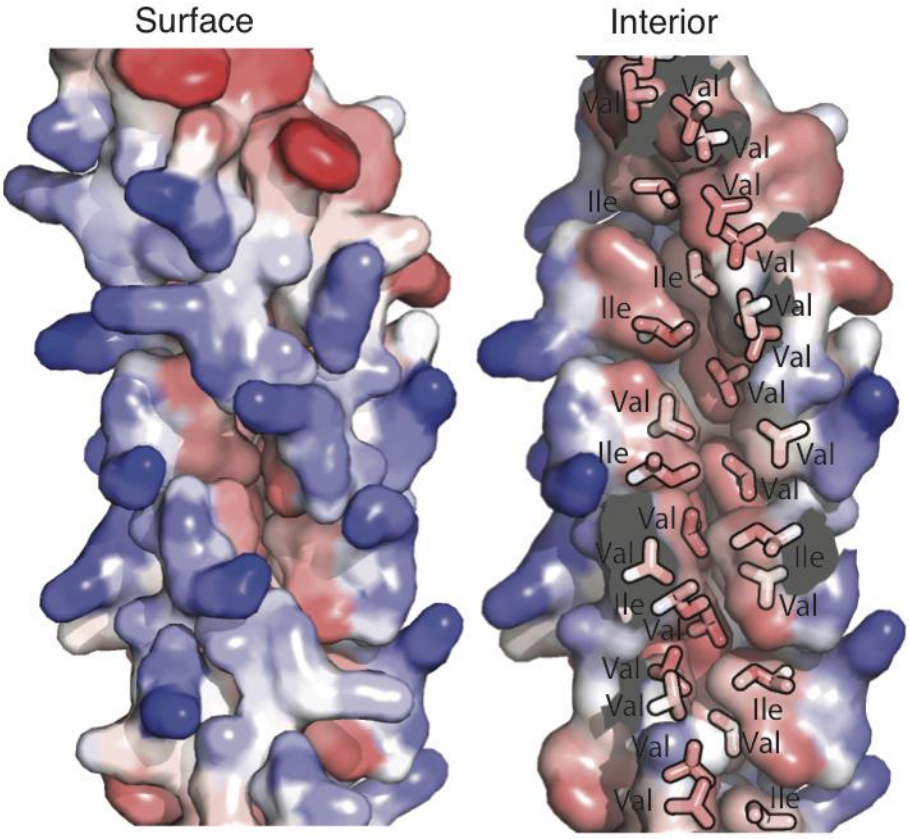


Supplementary Figure 4.1. Prediction of cylindrin-compatible segments. (A) 3D profile method for predicting cylindrin-compatible segments in amyloid proteins. An example of threading with the amyloid-β peptide is shown. (B,C) Visual analysis of the cylindrin structure revealed a layering of side chains in the core (b; backbone trace with even-numbered, 'inward-

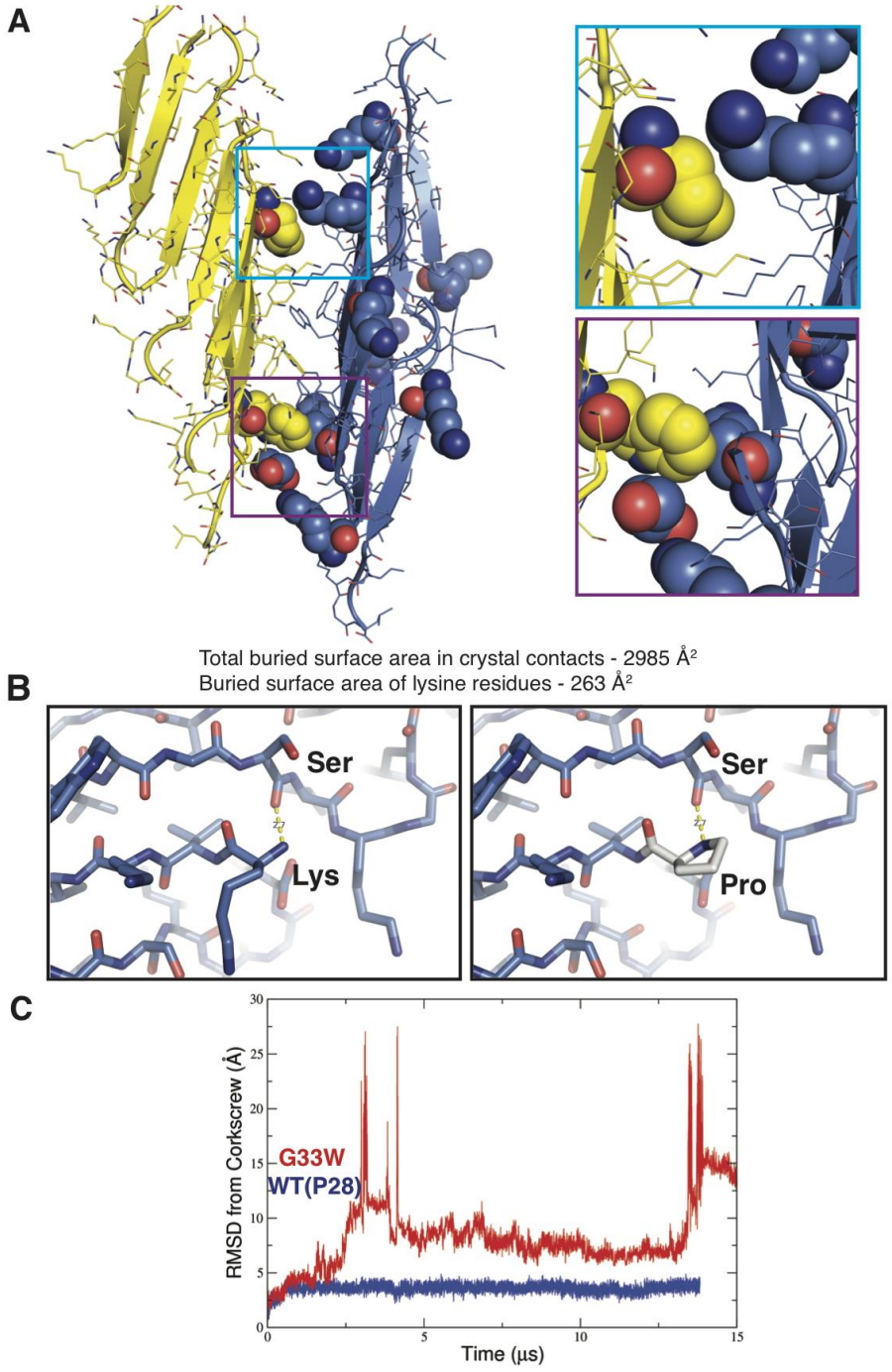
facing' side chains shown as sticks, and glycine Cα shown as a small sphere), and a 'knob in hole' packing of valine and glycine residues (c; surface rendering) that allows for the small circumference of the barrel. (D-H) Structures resulting from the 3D profile predictions were analyzed for packing in the barrel core. Representation is similar to panel B, but side-chain atoms are shown as spheres with van der Waals radii. The protein of origin and segment sequence are given, with even-numbered, 'inward-facing' residues in bold, and the critical position 4 residue underlined. (D) The original cylindrin structure, after Rosetta-based side-chain repacking and energy minimization. Notice the packing of valine 4, cyan, near the center. Although there is a gap between the two layers (arrowhead), it is small. (E) In this model of a segment from the human prion protein, methionine (cyan) occupies the central position 4, filling the space more completely than valine of the cylindrin sequence. (F) A histidine side chain at position 4 can also fill the central space, but is not able to satisfy hydrogen bonding groups. (G) Alanine at position 4 does not fill the core as well, leaving gaps (arrowheads). (H) Adjacent alanine and glycine, at positions 2 and 8 respectively, leave large gaps in outer layers of the cylindrin model core (arrowheads), in spite of a native valine packing in the middle at position 4. (I) Manual selection of cylindrin-compatible segments. First, three sequence-based criteria, which define the glycine position and the position of an adjacent, appropriate side chain (with cylindrin-compatible size and charge) in the core, were used in the initial selection of segments. The resulting sequences were further culled based on the presence of undesirable amino acids in specific positions. Prolines were avoided because of their tendency to disrupt β-structure. Charged and bulky inward-facing residues were avoided because of likely disruption to the packing. Small inward-facing residues were avoided because of gaps in packing, *e.g*. as in panel H. Large numbers of glycines were avoided because of the tendency for these sequences to be flexible rather than adopt a fixed conformation. Cysteines were avoided for experimental purposes. Four overlapping segments of SOD1 resulted from this manual selection.



Supplementary Figure 4.2. Frequency distribution of ALS-related mutations in SOD1. (A) Histogram representation of the number of familial-disease mutations found at each residue position. Gray connected line refers to the total number of patients known to carry mutations at that residue. The segment 28-38 is expanded. Val29 and Val31 each have one known familial mutation and Gly37 and Leu38 have two known familial mutations. Notice that the segment 28-38 has low frequency of individuals with ALS linked mutations with none in residues 32-36. A total of 4 individuals are known to carry mutations in this region. Data were obtained from http://alsod.iop.kcl.ac.uk (B) Ribbon diagram of dimeric SOD1 (PDB ID: 2C9S) with the segment 28-38 highlighted in blue. Zinc and copper atoms are shown as magenta spheres, and the familial mutation residues, A4V and G93A are colored red.

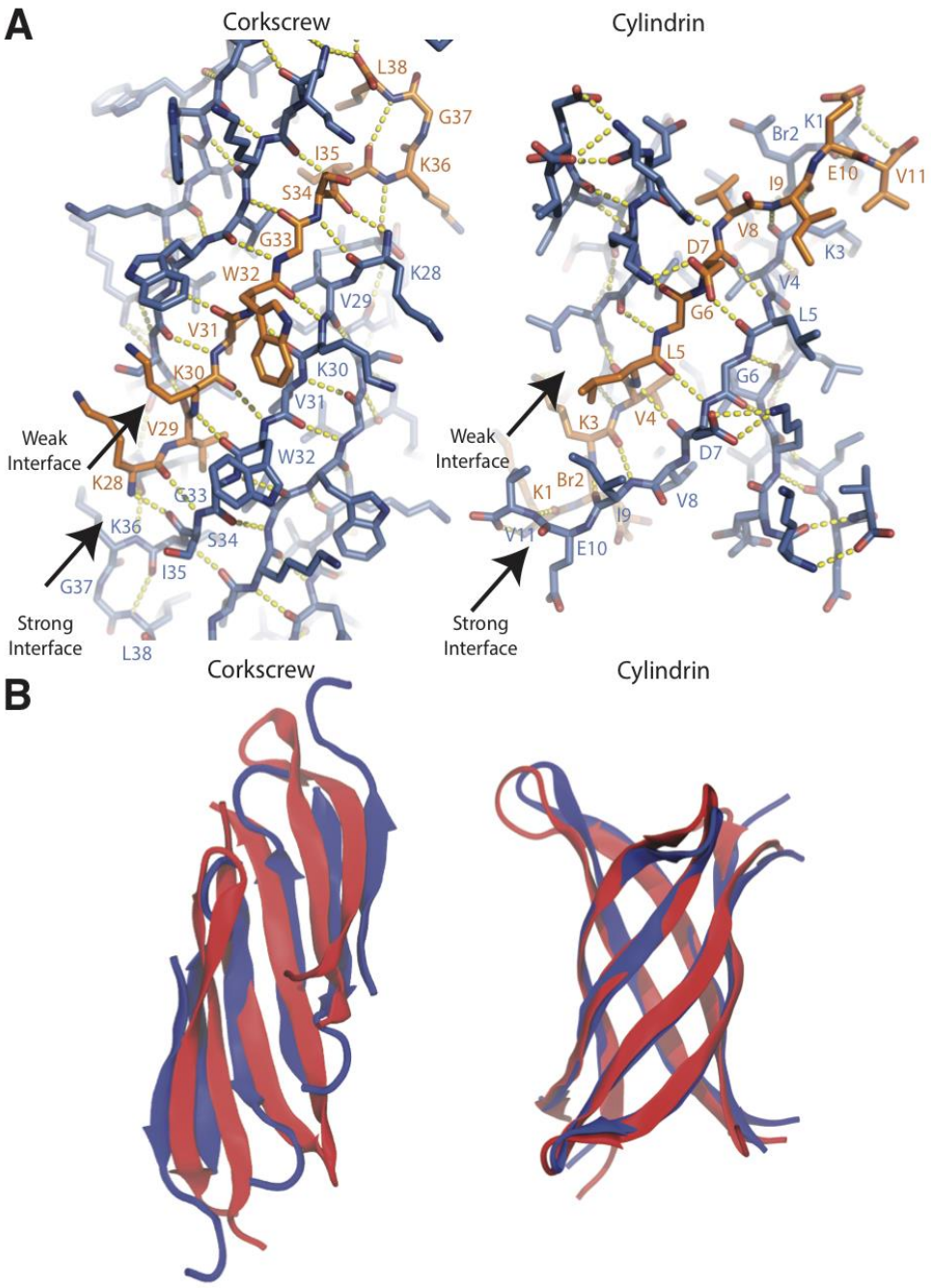


Supplementary Figure 4.3. Hydrophobicity mapped on the surface of the corkscrew crystal structure (left), and a lengthwise cross-section (right). The cleft interior is strongly hydrophobic (red) made up of valine and isoleucine side chains of Val29, Val31, and Ile35, whereas the exterior is strongly hydrophilic (blue) arising from Lys28, Lys30, Ser34, and Lys36.



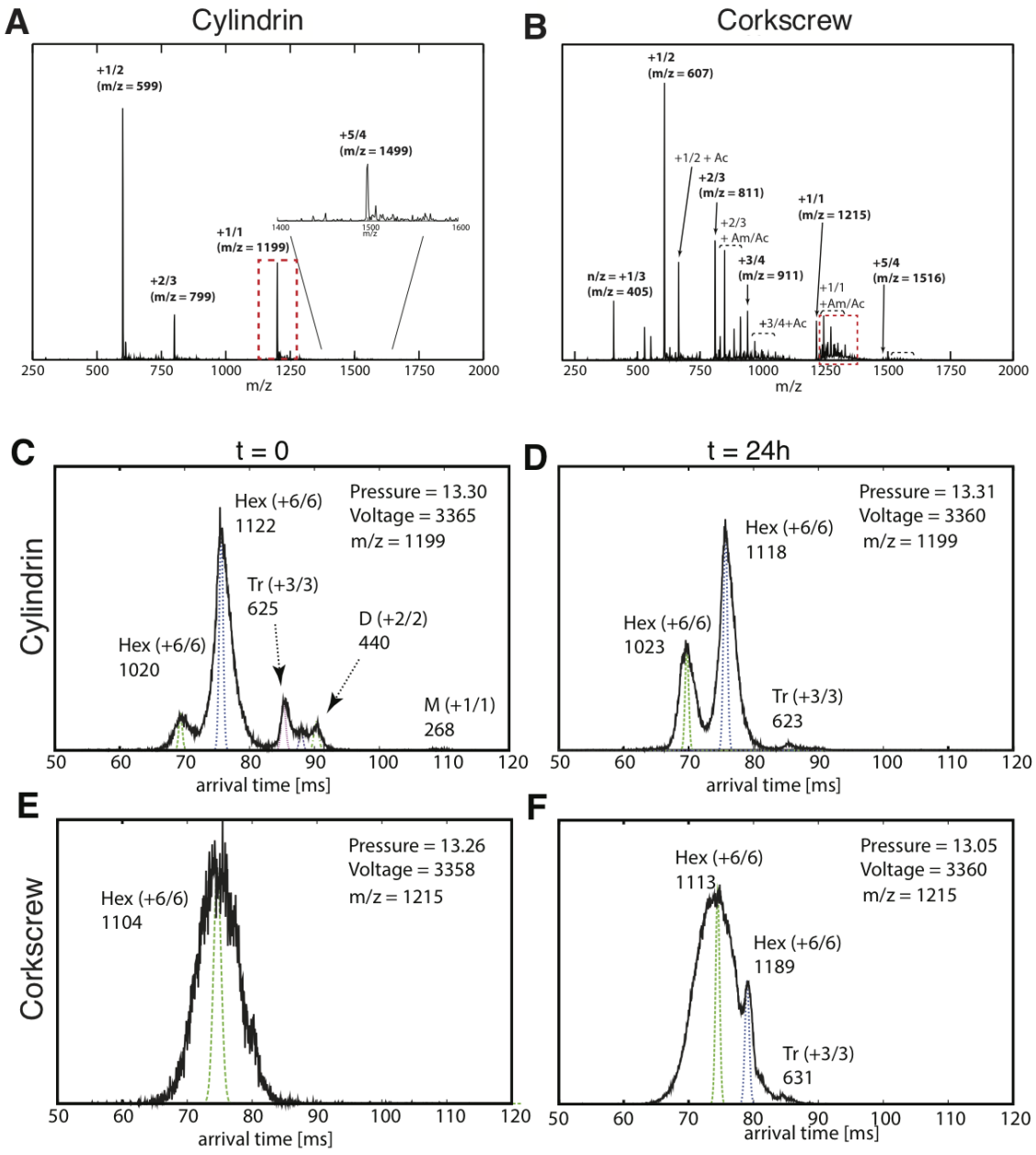
Supplementary Figure 4.4. Lysine substitution at position 28 contributes weakly to the crystal packing and not to the stability of the corkscrew structure. (A) Crystal packing of the corkscrew structure. Two asymmetric units forming crystal contacts are shown in blue and yellow from neighboring corkscrews and N-terminal lysines are shown in sphere representation. Among the eight N-terminal lysines in each asymmetric unit, 5 form weak van

der Waals interactions with a lysine residue and a tryptophan residue of the neighboring corkscrew. These interactions contribute weakly (~9%) of the total buried surface area in the crystal contacts. Buried surface area in the crystal contacts was calculated by AREAIMOL. (B) Lysine at position 28 is not essential for corkscrew formation and replacing it with the native proline residue does not disrupt the potential for hydrogen bonding. The N-terminal lysine forms a hydrogen bond with the carbonyl oxygen of Ser34 in an adjacent β-strand (left). Substituting the N-terminal lysine with the native proline residue (gray) maintains the hydrogen bond (right). (C) The blue and red curves correspond to Cα root-mean-square deviations (RMSD) from the corkscrew crystal structure in MD simulations of the 8-chain wild-type (PVKVWGSIKGL) and G33W mutant (PVKVWWSIKGL) corkscrews, respectively. The G33W mutant deviates from the corkscrew structure, whereas the wild-type structure remains stable throughout the length of the simulation. The simulations suggest that the dynamics of the backbone at residue 33 may contribute to the instability of the G33W mutant. The G33 residue in the corkscrew structure occupies a region of Ramachandran space different from the allowable region for tryptophan and other non-glycine amino acids. In the wild-type simulations, the backbone dihedral angles of G33 tend to remain in the lower right quadrant of a Ramachandran plot, whereas those of W33 in the mutant quickly relax to the β-sheet/α-helix regions, leading to deviations from the corkscrew structure. The extra hydrophobicity of G33W may also have an effect, making the mutant corkscrew more likely to collapse.



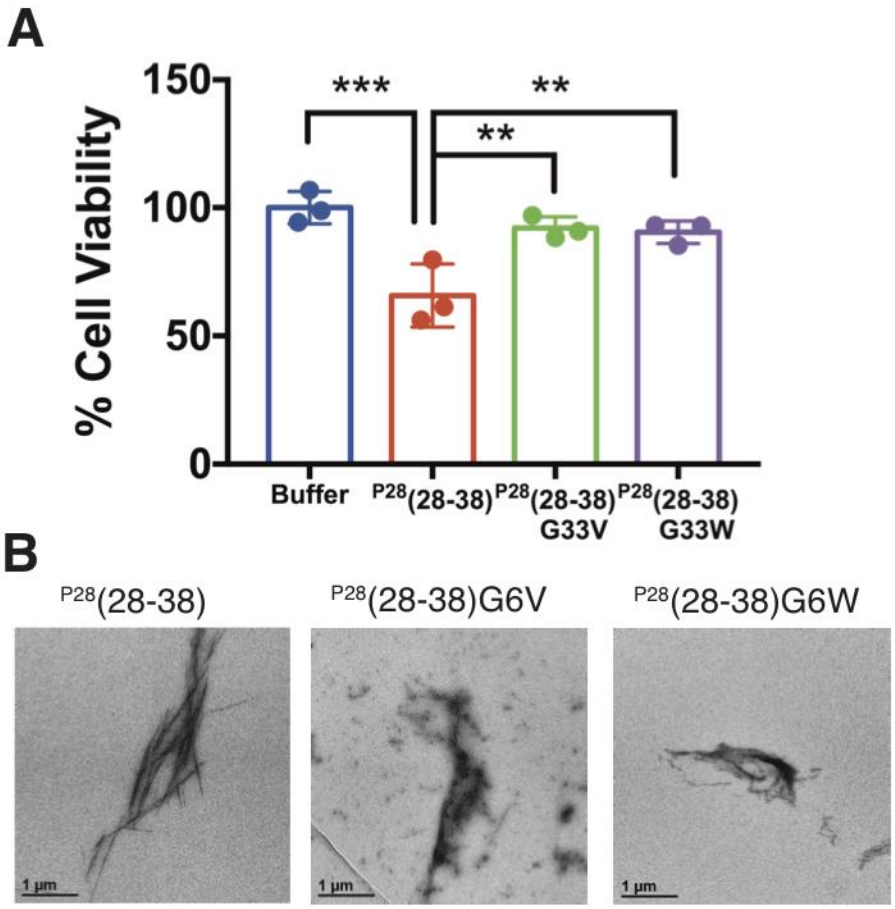
Supplementary Figure 4.5. Structural comparison of corkscrew and cylindrin. (A) Hydrogen bonding network of the corkscrew structure (left). Shown here the orange β-strand has a strong interface with the β-strand below composed of 9 hydrogen bonds and a weak interface with the β-strand above composed of 7 hydrogen bonds. Hydrogen bonding network in the cylindrin structure (Right, PDB ID: 3SGO) shares a similar architecture where the orange

strand has a stronger interface with the β-strand below composed of 12 hydrogen bonds and a weak interface with the β-strand above composed of 8 hydrogen bonds. Br2 refers to the non-natural amino acid 2-bromoallyl-glycine that was used for phasing. (B) MD simulations of weakly restrained monomers of SOD1 spontaneously assembled into a corkscrew-like structure. A snapshot of an assembled corkscrew-like structure from the MD simulations (red) is overlaid onto the crystal structure (blue). Out of 20 simulations, three successfully assembled to a corkscrew-like structure. Interestingly, in trajectories where a corkscrew did not form, individual monomers often formed interfaces composed of out-of-register β-sheets. As a control, we found that monomers of αB crystallin spontaneously assembled into a cylindrin structure using the same simulation protocol. A snapshot of an assembled cylindrin from the MD simulation (red) is overlaid onto the crystal structure (blue, PDB ID 3SGR). Out of 10 simulations, one successful assembly to cylindrin was obtained. As in the SOD1 assembly runs, in αB crystallin assembly trajectories where a cylindrin did not form, we often observed dimers composed of out-of-register β-sheet interfaces.

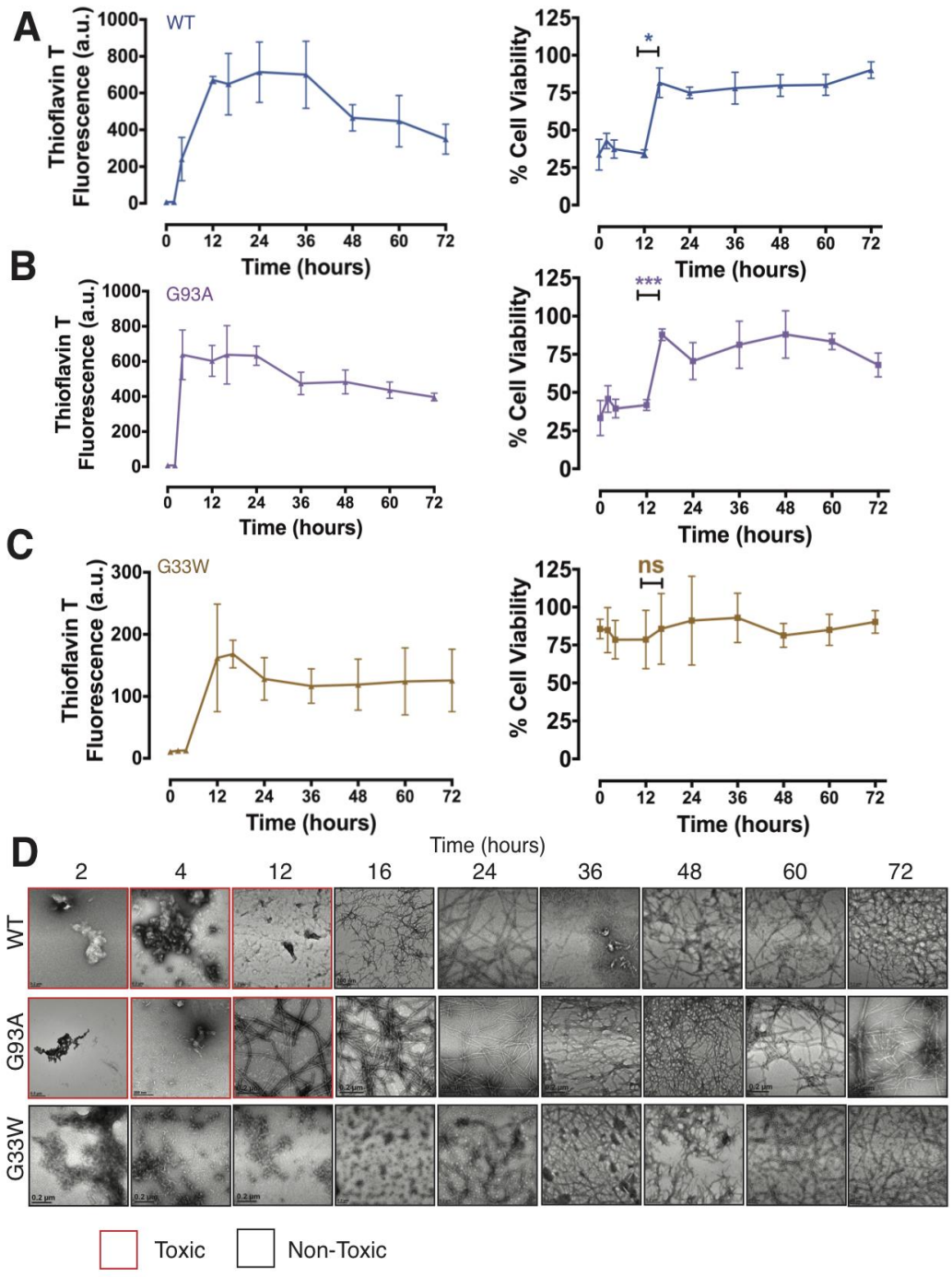


Supplementary Figure 4.6. The corkscrew-forming segment, SOD1(28-38) forms oligomers in solution with cross section similar to the crystal structure. (A, B) n-ESI-q mass spectrum of 200 μM cylindrin and corkscrew peptides in 20 mM ammonium acetate buffer (pH = 7.0), respectively. Each mass spectral peak is annotated with *n/z* where *n* is the oligomer number and *z* is the charge. (C-F) Representative ATDs of *n/z* = 1/1 peaks of the cylindrin and corkscrew peptides incubated for t = 0 and 24 hours. Each ATD feature is

annotated with *n/z* and an experimental collision cross section (M = monomer, D = dimer, Tr = trimer, Hex = hexamer). The narrow dashed lines are the peak shapes predicted for a single conformer of the cross sections given.

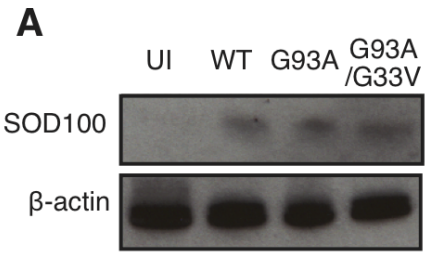


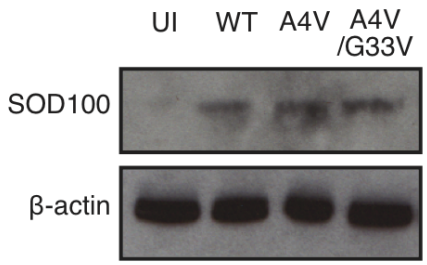
Supplementary Figure 4.7. Segment (28-38) with native proline is toxic and substitution at Gly33 renders it non-toxic. (A) Peptides were prepared identically by solubilizing in 50 mM tris-base buffer and overnight incubation at 37 °C with agitation and ES-derived motor neurons were treated with 100 µM final concentration. ^{P28}(28-38) is toxic while the peptides with substitutions G33V and G33W are non-toxic. Symbols represent individual values and bars represent average values. Statistical significance was measured by ANOVA. (**p < 0.01, ***p < 0.001). (B) Electron micrographs of all three peptides showed similar fibrils suggesting that substitution at Gly33 does not change the fibrillation propensity.

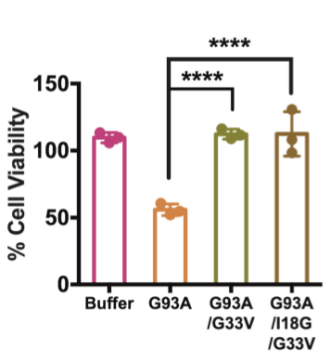


Supplementary Figure 4.8. WT and mutant SOD1 show similar kinetics of aggregation and cytotoxicity. (A) Thioflavin T assay of 80 μM WT protein shows a lag time of 2-4 hours before fibrils first appear (left). Cell viability (right) of ES-derived primary motor neurons incubated with protein aggregated for indicated times shows WT is only toxic when shaken up to 12 hours, when abundant fibrils first appear. (B) Similar to WT, the mutant G93A

shows a lag time of 2-4 hours in a thioflavin T assay (left). Cell viability (right) of ES-derived primary motor neurons incubated with protein aggregated for indicated times shows G93A is also toxic when shaken up to 12 hours, when abundant fibrils first appear. (C) Thioflavin T assay (left) shows G33W, the corkscrew-disrupted mutant protein behaves similar to WT and G93A with a lag phase of 2-4 hours. However, it is non-toxic at all time points (right). Results shown as Mean$\pm$SD (n=3). Statistical significance was analyzed using a two-tailed T-test with Welch's correction (*$p < 0.05$, ***$p < 0.001$, ns not significant). (D) Electron micrographs of SOD1 constructs aggregated for different lengths of time. Samples aggregated up to 12 hours tend to show some aggregates and sparse fibrils while samples aggregated for 16 hours or more show a large fibril load. Scale bar 200 nm.







C

	% Deformed
WT	2.33
G93A	10.34
G93A/G33V	2.88

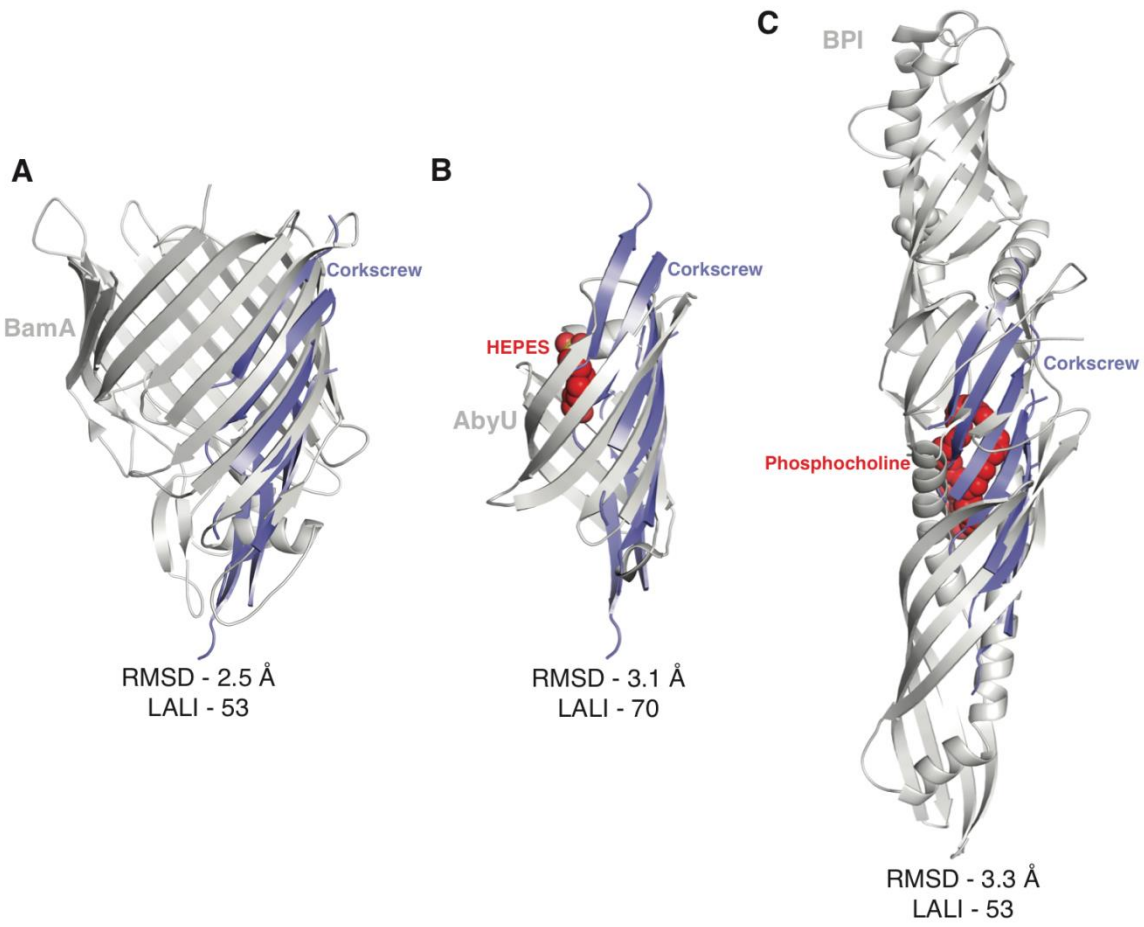
	% Deformed
WT	0
A4V	30
A4V/G33V	3.125

D

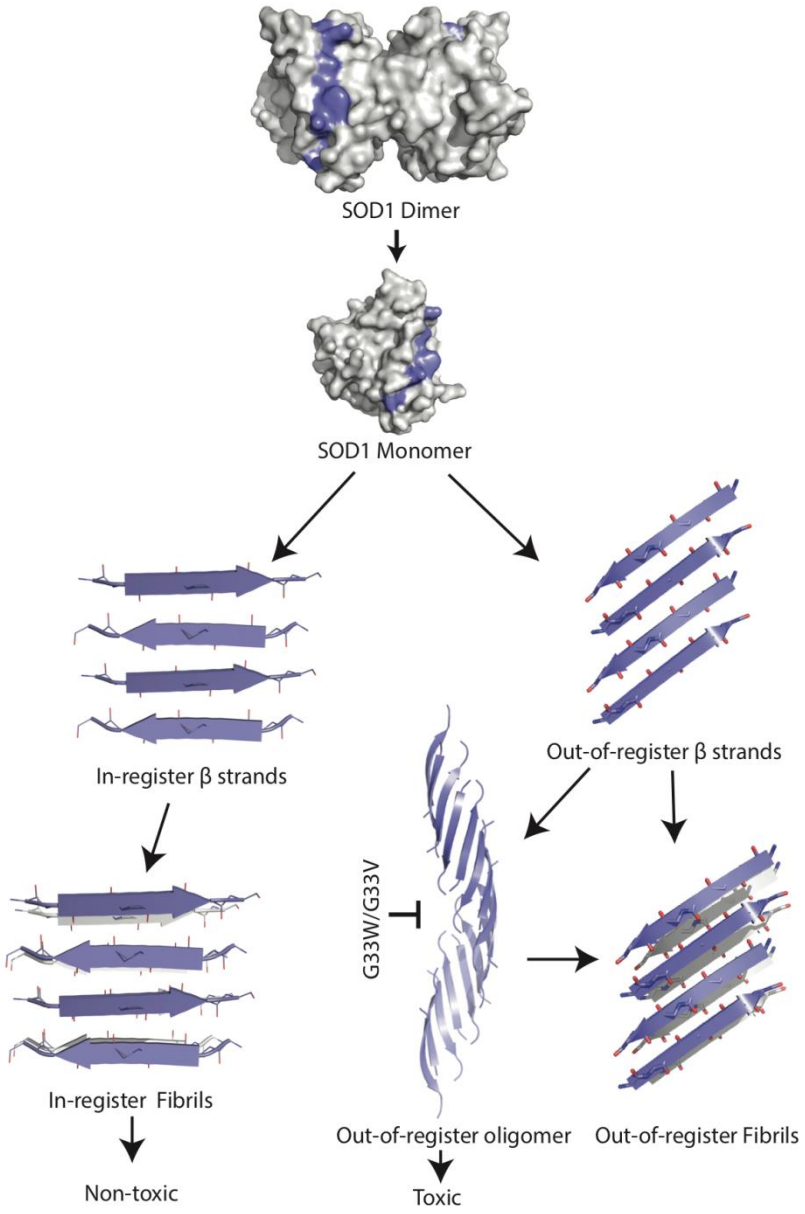
	Axon Length (n)	Mitochondria (n)	
		Size (a.u.)	Number
WT	1± .01 (365)	14.08	58 (365)
A4V	0.92 ± .01 (515)	30.17	32 (468)
A4V/G33V	0.95 ± .01 (575)	10.57	54 (400)

	Axon Length (n)	Mitochondria (n)	
		Size (a.u.)	Number
WT	1 ± .01 (360)	10.78	78 (171)
G93A	0.95 ± .01 (390)	14.68	39 (374)
G93A/G33V	1.03 ± .01 (375)	11.68	59 (361)

Supplementary Figure 4.9. Expression of A4V and G93A mutant proteins causes axon shortening and abnormal mitochondrial clustering in zebrafish model. (A) Western blotting of zebrafish embryos to confirm expression of the mRNA injected. SOD100 was used to detect SOD1 and β-actin was used as loading control. UI refers to uninjected zebrafish. (B) Construct G93A/G33V/I18G is non-toxic in cell culture model similar to the G93A/G33V mutant. (C) Zebrafish at 2 dpf were imaged and higher percentage of A4V (30%) and G93A (10%) injected zebrafish displayed significant deformation compared to WT and the G33V expressing fish. (D) Average axon length (normalized to WT construct) and mitochondrial size of the different constructs. n refers to the total number of measurements made for each construct.



Supplementary Figure 4.10. Structural comparison of corkscrew with other β-sheet proteins. The DALI server was used to search the Protein Data Bank for proteins with a similar structural fold. Three main categories of proteins were found – big barrel proteins, small β-sheet proteins that bind ligands and open twisted β-sheet proteins that bind ligands. (A) Overlay of corkscrew (purple) with the structure of BamA (PDB: 4K3C), a β-barrel membrane protein (gray). (B) Overlay of corkscrew (purple) with the structure of AbyU (PDB: 5DYV), a diels alderase (gray) with HEPES (red) bound in the barrel. (C) Overlay of corkscrew (purple) with BPI (PDB: 1BP1), bactericidal protein with phosphocholine bound. Notice that the twisted sheet creates a groove for binding phosphocholine (red). RMSD – root mean square deviation of Cα, LALI – length of aligned residues.



Supplementary Figure 4.11. Model of SOD1 aggregation pathway. SOD1 forms a stable dimer that is resistant to aggregation. Its destabilization into monomers renders it prone to aggregation into β-sheet rich fibrils. The corkscrew structure of segment 28-38 (purple) determined here, suggests that there is an off-pathway for oligomer formation independent of the in-register fibrils. The mutations, G33W/G33V specifically prevent corkscrew formation, an out-of-register oligomer. Fibril formation assays suggest that G33W/G33V do not affect the in-register fibril formation pathway and large fibrils are non-toxic.

Supplementary Table 4.1. X-Ray Data Collection and Refinement Statistics

	KVKVWGSIKGL	KVKVWGSIKGL (iodide)
Beamline	APS 24-ID-E	APS 24-ID-E
Resolution Å	1.9	2.1
Total unique reflections	8378	11489
Total reflections observed	47323	43500
Unit cell dimensions **a, b, c (Å)** **α, β, γ (°)**	 33.2, 44.4, 71.2 90.0, 90.0, 90.0	 33.1, 44.4, 71.4 90.0, 90.0, 90.0
Space group	$P2_12_12_1$	$P2_12_12_1$
R_{merge}	16.5% (48.4%)	9.5% (41.8%)
I/σ	8.4 (4.3)	11.1 (3.6)
Completeness	95.5 %	96.7%
Wavelength (Å)	0.9791	0.9791
Refinement		
Resolution (Å)	35.5-2.0	19.3-2.0
Reflections for refinement (after merging Friedel Pairs)	6690	6249
R_{free}/R_{work} (%)	25.08/21.09	26.2/21.2
Molecules per asymmetric unit	8	8
Solvent content (%)	54.6	54.6
Matthews coefficient (Å³/Da)	2.71	2.71
Total water molecules	34	52
Total iodide atoms	0	2
Glycerol molecule	0	1
Rmsd bond length (Å)	0.010	0.010
Rmsd angles (°)	1.022	0.88
Ramachandran plot **Allowed** **Generous** **Disallowed**	 100 0 0	 100 0 0

Supplementary Table 4.2. Comparison of shape complementarity (Sc) and buried surface area (A_b) of corkscrew with cylindrin and steric zippers.

	Corkscrew	Cylindrin*	Steric Zipper**
Sc	0.79	0.74	0.68
A_b (Å²)	984	943	1034
A_b/Residue	89	86	94

* PDB ID: 3SGO

** PDB ID: 4RIL

For the Sc calculation, we examined the interface between one chain and the remaining chains of the assembly.

A_b values were calculated using AREAIMOL. The area buried was calculated by subtracting the solvent accessible surface area of one chain of the assembly from the total solvent accessible surface area of an isolated chain.

REFERENCES

1. Shankar GM, et al. (2008) Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat Med* 14(8):837–842.

2. Baglioni S, et al. (2006) Prefibrillar Amyloid Aggregates Could Be Generic Toxins in Higher Organisms. *J Neurosci* 26(31):8160–8167.

3. Eisele YS, et al. (2015) Targeting protein aggregation for the treatment of degenerative diseases. *Nat Rev Drug Discov* 14(11):759–780.

4. Nelson R, et al. (2005) Structure of the cross-β spine of amyloid-like fibrils. *Nature* 435(7043):773–778.

5. Sawaya MR, et al. (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* 447(7143):453–457.

6. Xiao Y, et al. (2015) Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol* 22(6):499–505.

7. Colvin MT, et al. (2016) Atomic Resolution Structure of Monomorphic $A\beta_{42}$ Amyloid Fibrils. *J Am Chem Soc* 138(30):9663–9674.

8. Tuttle MD, et al. (2016) Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat Struct Mol Biol* 23(5):409–415.

9. Wälti MA, et al. (2016) Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *Proc Natl Acad Sci* 113(34):E4976–E4984.

10. Zou Y, Li Y, Hao W, Hu X, Ma G (2013) Parallel β-Sheet Fibril and Antiparallel β-Sheet Oligomer: New Insights into Amyloid Formation of Hen Egg White Lysozyme under Heat and Acidic Condition from FTIR Spectroscopy. *J Phys Chem B* 117(15):4003–4013.

11. Berthelot K, Ta HP, Géan J, Lecomte S, Cullin C (2011) In Vivo and In Vitro Analyses of Toxic Mutants of HET-s: FTIR Antiparallel Signature Correlates with Amyloid Toxicity. *J Mol Biol* 412(1):137–152.

12. Celej MS, et al. (2012) Toxic prefibrillar α-synuclein amyloid oligomers adopt a distinctive antiparallel β-sheet structure. *Biochem J* 443(3):719–726.

13. Cerf E, et al. (2009) Antiparallel β-sheet: a signature structure of the oligomeric amyloid β-peptide. *Biochem J* 421(3):415–423.

14. Sandberg A, et al. (2010) Stabilization of neurotoxic Alzheimer amyloid- oligomers by protein engineering. *Proc Natl Acad Sci* 107(35):15595–15600.

15. Wille H, Drewes G, Biernat J, Mandelkow E-M, Mandelkow E (1992) Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau in vitro. *J Cell Biol* (118(3): 573–584.).

16. Rosen DR, et al. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362(6415):59–62.

17. Zou Z-Y, Liu M-S, Li X-G, Cui L-Y (2015) Mutations in SOD1 and FUS caused juvenile-onset sporadic amyotrophic lateral sclerosis with aggressive progression. *Ann Transl Med* (3(15):221).

18. Pasinelli P, Brown RH (2006) Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* 7(9):710–723.

19. Gurney M, et al. (1994) Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science* 264(5166):1772–1775.

20. Borchelt DR, et al. (1994) Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc Natl Acad Sci* 91(17):8292–8296.

21. Taylor JP, Brown RH, Cleveland DW (2016) Decoding ALS: from genes to mechanism. *Nature* 539(7628):197–206.

22. Bruijn LI (1998) Aggregation and Motor Neuron Toxicity of an ALS-Linked SOD1 Mutant Independent from Wild-Type SOD1. *Science* 281(5384):1851–1854.

23. Bruijn LI, et al. (1997) ALS-Linked SOD1 Mutant G85R Mediates Damage to Astrocytes and Promotes Rapidly Progressive Disease with SOD1-Containing Inclusions. *Neuron* 18(2):327–338.

24. Oztug Durer ZA, et al. (2009) Loss of Metal Ions, Disulfide Reduction and Mutations Related to Familial ALS Promote Formation of Amyloid-Like Aggregates from Superoxide Dismutase. *PLoS ONE* 4(3):e5004.

25. Park Y-N, et al. (2012) Huntingtin Fragments and SOD1 Mutants Form Soluble Oligomers in the Cell. *PLoS ONE* 7(6):e40329.

26. Kim J, et al. (2014) Dimerization, Oligomerization, and Aggregation of Human Amyotrophic Lateral Sclerosis Copper/Zinc Superoxide Dismutase 1 Protein Mutant Forms in Live Cells. *J Biol Chem* 289(21):15094–15103.

27. Zetterström P, et al. (2007) Soluble misfolded subfractions of mutant superoxide dismutase-1s are enriched in spinal cords throughout life in murine ALS models. *Proc Natl Acad Sci* 104(35):14157–14162.

28. Wang J, et al. (2009) Progressive aggregation despite chaperone associations of a mutant SOD1-YFP in transgenic mice that develop ALS. *Proc Natl Acad Sci* 106(5):1392–1397.

29. Wright GSA, Antonyuk SV, Kershaw NM, Strange RW, Samar Hasnain S (2013) Ligand binding and aggregation of pathogenic SOD1. *Nat Commun* 4:1758.

30. Durazo A, et al. (2009) Metal-free Superoxide Dismutase-1 and Three Different Amyotrophic Lateral Sclerosis Variants Share a Similar Partially Unfolded -Barrel at Physiological Temperature. *J Biol Chem* 284(49):34382–34389.

31. Shaw BF, et al. (2006) Local Unfolding in a Destabilized, Pathogenic Variant of Superoxide Dismutase 1 Observed with H/D Exchange and Mass Spectrometry. *J Biol Chem* 281(26):18167–18176.

32. Taylor DM, et al. (2007) Tryptophan 32 Potentiates Aggregation and Cytotoxicity of a Copper/Zinc Superoxide Dismutase Mutant Associated with Familial Amyotrophic Lateral Sclerosis. *J Biol Chem* 282(22):16329–16335.

33. Do TD, et al. (2016) Amyloid β-Protein C-Terminal Fragments: Formation of Cylindrins and β-Barrels. *J Am Chem Soc* 138(2):549–557.

34. Liu P, et al. (2015) Quaternary Structure Defines a Large Class of Amyloid-β Oligomers Neutralized by Sequestration. *Cell Rep* 11(11):1760–1771.

35. Hoyer W, Gronwall C, Jonsson A, Stahl S, Hard T (2008) Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid- peptide inhibits amyloid formation. *Proc Natl Acad Sci* 105(13):5099–5104.

36. Laganowsky A, et al. (2012) Atomic View of a Toxic Amyloid Small Oligomer. *Science* 335(6073):1228–1231.

37. Eisenberg DS, Sawaya MR (2017) Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem* 86(1). doi:10.1146/annurev-biochem-061516-045104.

38. Rakhit R, et al. (2007) An immunological epitope selective for pathological monomer-misfolded SOD1 in ALS. *Nat Med* 13(6):754–759.

39. Liu H-N, et al. (2012) Targeting of Monomer/Misfolded SOD1 as a Therapeutic Strategy for Amyotrophic Lateral Sclerosis. *J Neurosci* 32(26):8791–8799.

40. Rakhit R, et al. (2004) Monomeric Cu,Zn-superoxide Dismutase Is a Common Misfolding Intermediate in the Oxidation Models of Sporadic and Familial Amyotrophic Lateral Sclerosis. *J Biol Chem* 279(15):15499–15504.

41. Lelie HL, et al. (2011) Copper and Zinc Metallation Status of Copper-Zinc Superoxide Dismutase from Amyotrophic Lateral Sclerosis Transgenic Mice. *J Biol Chem* 286(4):2795–2806.

42. Valentine JS, Hart PJ (2003) Misfolded CuZnSOD and amyotrophic lateral sclerosis. *Proc Natl Acad Sci* 100(7):3617–3622.

43. Wichterle H, Lieberam I, Porter JA, Jessell TM (2002) Directed Differentiation of Embryonic Stem Cells into Motor Neurons. *Cell* 110(3):385–397.

44. Wang Q, Johnson JL, Agar NYR, Agar JN (2008) Protein Aggregation and Protein Instability Govern Familial Amyotrophic Lateral Sclerosis Patient Survival. *PLoS Biol* 6(7):e170.

45. Gruzman A, et al. (2007) Common molecular signature in SOD1 for both sporadic and familial amyotrophic lateral sclerosis. *Proc Natl Acad Sci* 104(30):12524–12529.

46. Forsberg K, et al. (2010) Novel Antibodies Reveal Inclusions Containing Non-Native SOD1 in Sporadic ALS Patients. *PLoS ONE* 5(7):e11552.

47. Guareschi S, et al. (2012) An over-oxidized form of superoxide dismutase found in sporadic amyotrophic lateral sclerosis with bulbar onset shares a toxic mechanism with mutant SOD1. *Proc Natl Acad Sci* 109(13):5074–5079.

48. Ramesh T, et al. (2010) A genetic model of amyotrophic lateral sclerosis in zebrafish displays phenotypic hallmarks of motoneuron disease. *Dis Model Mech* 3(9–10):652–662.

49. Sakowski SA, et al. (2012) Neuromuscular effects of G93A-SOD1 expression in zebrafish. *Mol Neurodegener* 7(1):44.

50. Van Hoecke A, et al. (2012) EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. *Nat Med* 18(9):1418–1422.

51. Lemmens R, et al. (2007) Overexpression of mutant superoxide dismutase 1 causes a motor axonopathy in the zebrafish. *Hum Mol Genet* 16(19):2359–2365.

52. Levesque MP, Krauss J, Koehler C, Boden C, Harris MP (2013) New Tools for the Identification of Developmentally Regulated Enhancer Regions in Embryonic and Adult Zebrafish. *Zebrafish* 10(1):21–29.

53. Wong PC, et al. (1995) An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron* 14(6):1105–1116.

54. Shi P, Gal J, Kwinter DM, Liu X, Zhu H (2010) Mitochondrial dysfunction in amyotrophic lateral sclerosis. *Biochim Biophys Acta BBA - Mol Basis Dis* 1802(1):45–51.

55. Kiskinis E, et al. (2014) Pathways Disrupted in Human ALS Motor Neurons Identified through Genetic Correction of Mutant SOD1. *Cell Stem Cell* 14(6):781–795.

56. Vande Velde C, et al. (2011) Misfolded SOD1 Associated with Motor Neuron Mitochondria Alters Mitochondrial Shape and Distribution Prior to Clinical Onset. *PLoS ONE* 6(7):e22031.

57. Sotelo-Silveira JR, et al. (2009) Axonal Mitochondrial Clusters Containing Mutant SOD1 in Transgenic Models of ALS. *Antioxid Redox Signal* 11(7):1535–1545.

58. Magrane J, Cortez C, Gan W-B, Manfredi G (2014) Abnormal mitochondrial transport and morphology are common pathological denominators in SOD1 and TDP43 ALS mouse models. *Hum Mol Genet* 23(6):1413–1424.

59. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server):W545–W549.

60. Beamer LJ (1997) Crystal Structure of Human BPI and Two Bound Phospholipids at 2.4 Angstrom Resolution. *Science* 276(5320):1861–1864.

61. Liu C, et al. (2012) Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc Natl Acad Sci* 109(51):20913–20918.

62. Treusch S, Cyr DM, Lindquist S (2009) Amyloid deposits: Protection against toxic protein species? *Cell Cycle* 8(11):1668–1674.

63. Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431(7010):805–810.

64. Kuperstein I, et al. (2010) Neurotoxicity of Alzheimer's disease Aβ peptides is induced by small changes in the Aβ42 to Aβ40 ratio. *EMBO J* 29(19):3408–3420.

65. Martins IC, et al. (2008) Lipids revert inert Aβ amyloid fibrils to neurotoxic protofibrils that affect learning in mice. *EMBO J* 27(1):224–233.

66. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* 26(6):795–800.

67. Emsley P, Cowtan K (2004) *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(12):2126–2132.

68. Vagin AA, et al. (2004) *REFMAC* 5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60(12):2184–2195.

69. Bricogne G, et al. BUSTER version 1.10.0. Cambridge, United Kingdom: Global Phasing Ltd.

70. Delano W The PyMOL Molecular Graphics System. (2002).
71. Shaw DE, et al. (2009) Millisecond-scale molecular dynamics simulations on Anton (ACM Press), p 1.
72. Shaw DE, et al. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer (IEEE), pp 41–53.
73. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinforma* 65(3):712–725.
74. Best RB, Hummer G (2009) Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. *J Phys Chem B* 113(26):9004–9015.
75. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma*:NA-NA.
76. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926.
77. Kemper PR, Dupuis NF, Bowers MT (2009) A new, higher resolution, ion mobility mass spectrometer. *Int J Mass Spectrom* 287(1–3):46–57.
78. Transport properties of ions in gases.
79. Gidden J, Ferzoco A, Baker ES, Bowers MT (2004) Duplex Formation and the Onset of Helicity in Poly d(CG)n Oligonucleotides in a Solvent-Free Environment. (126 (46), 15132-15140).
80. Westerfield, M. The Zebrafish Book. The University of Oregon Press, Eugene, Oregon. (2003).
81. Levesque MP, Krauss J, Koehler C, Boden C, Harris MP (2013) New Tools for the Identification of Developmentally Regulated Enhancer Regions in Embryonic and Adult Zebrafish. *Zebrafish* 10(1):21–29.
82. Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging* 13(1):146.

CHAPTER 5

Structure of the toxic core of α-synuclein from invisible crystals

INTRODUCTION

The presynaptic protein α-synuclein (α-syn), found in both soluble and membrane-associated fractions of the brain, aggregates in Parkinson's Disease (PD). These aggregates are the main component of Lewy bodies, the defining histological feature of this neurodegenerative disease, and have been shown to accompany neuronal damage (1). Two other observations point to aggregated α-syn as a molecular cause of Parkinson's disease (2). The first is that families with inherited forms of PD carry mutations in α-syn, such as A53T, and abundant Lewy bodies (3-5). The second is that families with duplicated or triplicated genes encoding α-syn develop early onset PD, presumably because at high local concentrations α-syn is forced into amyloid (6,7).

Our focus is on a central segment of α-syn, residues 68-78, that we call NACore (Figure 5.1), because of its critical role in both the aggregation and cytotoxicity of α-syn. NACore lies within a 35 residue domain of α-syn termed NAC (Non Amyloid-β Component, originally reported to be deposited with amyloid b in the brains of Alzheimer's disease patients), which the studies of others have established as necessary and sufficient for aggregation and toxicity of α-syn (8-12) (Supplementary Figure 5.1). For example, deletion of residues 71-82 prevents aggregation of α-syn *in vitro*, and abolishes both its aggregation and neurotoxicity in a drosophila model of PD (12). Yet this segment in isolation from the rest of α-syn readily forms amyloid fibrils and is highly cytotoxic (13, 14). Also, β-synuclein, the close homolog of α-syn, which does not aggregate and is not found in Lewy bodies, differs in sequence from α-syn principally by the lack of residues 74-84 that are part of NACore (9).

Segments outside NAC also influence the aggregation of α-syn and have been associated with fibril structure (15-17). In brain extracts from patients with multiple system atrophy, the core of α-syn fibrils extends approximately from residue 30 to 100 (18). Also the A53T mutation of α-syn can accelerate its transition into the amyloid state, and hence accelerate PD (19). This mutation was found to induce the onset of PD at an early age (20), and consistent with this, α-syn containing this A53T mutation forms fibrils *in vitro* more rapidly than wild type (3). Thus we carried out screens for crystals of peptide segments within the NAC domain and adjacent regions, seeking structural information on the molecular basis of aggregation and toxicity of α-syn.

RESULTS

Structure determination byMicroED

Extensive crystal screens of two segments, NACore, residues ^{68}GAVVTGVTAVA^{78}, and PreNAC, ^{47}GVVHGVTTVA^{56}, seemingly produced non-crystalline, amorphous aggregates. But on examination by electron microscopy, we found the aggregates to be clusters of elongated nanocrystals only 50-300 nm in cross section and thus invisible by conventional light microcopy (Figure 5.1). We confirmed well-ordered crystallinity of NACore at both the SACLA and LCLS free electron lasers. We also found that a 9-residue fragment within the NACore, which we term SubNACore, ^{69}AVVTGVTAV^{77}, yielded crystals 1,000 – 10,000 times larger in volume than the NACore nanocrystals (Figure 5.1). We were therefore able to apply synchrotron methods(21, 22) to these larger crystals to determine the structure of their amyloid-like fibrils. Although this 9-residue fragment is missing only two residues compared with NACore, it is not as toxic (23), offering insight described below, into the toxicity of α-syn.

To determine the structure of the invisible crystals of NACore and PreNAC, we turned to Micro-Electron Diffraction (MicroED) (24-26). In MicroED, an extremely low dose electron beam is directed on a nanocrystal within a transmission electron microscope under cryogenic

conditions, yielding diffraction patterns such as that in Figure 5.2. As the wavelength used in our experiments at 200keV is very small (0.025Å), the Ewald sphere is essentially flat yielding diffraction patterns that closely resemble a 2D slice through 3D reciprocal space. As the crystal is continuously rotated in the beam, a series of such diffraction patterns is collected (25). Scaling together diffraction data collected from multiple crystals produces a full 3D diffraction dataset. MicroED has been successfully applied to the well-known structures of hen egg-white lysozyme (25, 26), bovine liver catalase (27) and Ca^{2+}-ATPase (28). But NACore and PreNAC are the first previously unknown structures determined by MicroED.

For NACore and PreNAC, we collected microED patterns from nano-crystals that lay preferentially oriented, flat on the surface of a holey carbon Quantifoil grid, in a frozen-hydrated state. Grids were first screened for appropriately sized crystals, and candidate crystals screened for diffraction. We used crystals showing strong diffraction for data collection by continuous unidirectional rotation about a fixed axis, acquiring a series of diffraction frames at fixed time intervals (25). The needle-shaped crystals typically exceeded the length needed for MicroED; those that were unbent and 100 to 300 nm wide produced the best diffraction patterns. Data from multiple crystals were integrated, scaled and merged together (Supplementary Table 5.1).

The multi-crystal NACore and PreNAC datasets were phased by molecular replacement, using the atomic model of SubNACore and an ideal beta strand model, respectively, as probes. As shown in Supplementary Figure 5.2, diffraction phases calculated from the SubNACore probe structure and NACore structure factors yielded a difference density map (Supplementary Figure 5.2), which clearly reveals the positions of the missing residues, after subsequent refinement, two water molecules, and several hydrogen atoms (Figure 5.3e). Full models of NACore and PreNAC were refined against the MicroED data, producing structures at 1.4Å resolution with acceptable R-factors (Supplementary Table 5.1). Electron scattering factors were used in the refinement calculations (29).

NACore Structure

The structure of the NACore peptide chain is a nearly fully extended β-strand (Figure 5.3 and Supplementary Figure 5.3). These NACore strands stack in-register into β-sheets, as had been predicted by site-directed spin labeling (16, 17). The sheets are paired (Figure 5.3b), as is usual in amyloid spines, and the pairs of sheets form typical steric-zipper protofilaments (Figure 5.3c), previously seen as the spines in many amyloid-like fibrils formed from short segments of fibril-forming proteins (21). The unusual features of this steric zipper are that the 11-residue width of the zipper is longer than has been previously observed (22), and each pair of sheets contains two water molecules, each associated with a threonine sidechain, within the interface instead of being completely dry. Also, in our crystals of NACore, each sheet forms two snug interfaces: Interface A with 268 $Å^2$ of buried accessible surface area per chain, is more extensive and presumably stronger than Interface B (167 $Å^2$), because the terminal residues of the chains in opposing sheets bend towards each other (Figure 5.3 and Supplementary Figure 5.4). The structure of PreNAC reveals a peptide chain that forms a β-strand kinked at residue glycine 51. These strands are arranged into pairs of β-sheets that like the NACore structure interdigitate to form steric zipper protofilaments (Figure 5.3). Of special note, a five residue segment of PreNAC (^{51}GVTTV^{55}) differs in only one residue from a five residue segment of NACore (^{73}GVTAV^{77}) and their backbones and identical sidechains superimpose closely with an alpha carbon RMS deviation of 1.5Å (Supplementary Figure 5.4). This means that weaker interface B of NACore mimics a hypothetical interface between NACore and PreNAC (Figure 5.3d).

Relevance of NACore to Parkinson Disease

The relevance of the structure of NACore to fibrils of full length α-syn is established by the resemblance of their diffraction patterns. Specifically, the fiber diffraction pattern of aligned fibrils of full-length and N-terminally acetylated (30) α-syn protein display the same principal peaks as the diffraction of aligned NACore nanocrystals (Figure 5.2). All three fibrils display the strong reflection at 2.4Å in their diffraction patterns. As seen in Figure 5.3 and Supplementary

Figure 5.5 this reflection arises in NACore because one β-sheet of the steric zipper is translated along the fiber axis with respect to the other β-sheet by 2.4Å, one half the 4.8Å spacing between β-strands, permitting the two sheets to interdigitate tightly together. All three share a strong 4.6Å reflection, which in NACore results from both the stacking of β-strands and the staggering between adjacent β-sheets of the steric zipper, while a shared reflection at near 8.2Å likely arises from the distance between the adjacent pairs of β-sheets that make up the α-syn fibril (Figure 5.2 and Supplementry Figure 5.5). This comparison of fiber diffraction patterns (Supplementary Table 5.2) strongly suggests that the structure of NACore is similar to the spine of our toxic fibrils of full α-syn.

The combined structures of NACore and PreNAC allow us to construct a speculative model for much of the ordered segments of the A53T early onset mutant α-syn (Figure 5.3d). Experimental support of this model comes from the agreement of its simulated fiber diffraction with the measured diffraction patterns of α-syn and N-acetyl α-syn fibrils as well as aligned NACore nanocrystals (Supplementary Table 5.2). Above we hypothesized that the weaker Interface B of NACore might mimic an intramolecular interaction of PreNAC with NACore (Figure 5.3). In fact, the interacting sidechains in the weaker NACore Interface B (G73, T75, and V77) are identical to the sidechains (G51, T53, V55) interacting in the hypothetical interface of PreNAC with NACore. Assuming that this interface actually forms in fibrils of the early onset mutant A53T, we have built the model shown in Figure 5.3d. The hypothetical interface of this model offers a possible reason for a greater propensity of the A53T mutant to aggregate than the wild type sequence, conceivably leading to early onset of PD.

The identity and structure of the cytotoxic amyloid formed by α-syn remains a subject of intensive research (19, 31-35). The weight of evidence over the past decade has tilted scientific opinion from the fully developed amyloid fibrils found in Lewy bodies as the toxic entities to smaller, transient amyloid oligomers. Yet recently, quantitative arguments have been put forward in favor of fibrils (36). Our experiments of the cytotoxicity of NACore on

neuroblastoma cells (Figure 5.4) are consistent with the view that fibrils are toxic: we find that NACore shaken and aggregated for 72 hours displays abundant fibrils, is more toxic than freshly dissolved NACore (Figure 5.4) and is comparably toxic to similarly aggregated full α-syn. We also find greater cytotoxicity of NACore than SubNACore, which is shorter by two residues. This is consistent with the more rapid fibril formation of NACore than of SubNACore (Figure 5.4d). These observations do not rule out the formation of a non-fibrillar, oligomeric assembly, present, but undetected, in our aggregated samples of NACore and α-syn. Of course, NACore is merely a fragment of full length α-syn, and lacks most of the membrane-binding motifs of the N-terminus of the protein, which have been implicated in membrane disruption (37, 38). Yet it is clear that NACore is the minimum entity that recapitulates all the features of full length a-syn aggregation and toxicity.

DISCUSSION

The miniscule size of NACore crystals is typical of amyloid and also of various other biological crystals of interest. For amyloid crystals, our speculation is that the tiny size is a consequence of the natural twist of β-sheets that form the protofilaments of the fibrils. The crystal lattice restrains the twist, creating a strain in these crystals, which increases as crystals grow. Eventually this strain prevents further addition of β-strands, limiting the thickness of the needle crystals. In our experience, longer segments (for example, 11 residues compared to 9 residues) limit crystal growth even more; in the case of 11-residue NACore and 10-residue PreNAC, the strain produces nanocrystals, invisible by optical microscopy. These crystals are too small for mounting and conventional synchrotron data collection, but are ideally suited for analysis by MicroED. Our structures of NACore and PreNAC demonstrate that MicroED is capable of determining new and accurate structures of biological material at atomic resolution. This finding paves the path for applications of MicroED to other biological substances of importance, for which only nanocrystals can be grown. In our particular application, we have

been able to learn the atomic arrangement of the core of the crucial NAC domain. This opens opportunities for structure-based design of inhibitors of amyloid formation of α–syn (39).

MATERIALS AND METHODS

Crystallization

Microcrystals of SubNACore, ^{69}AVVTGVTAV^{77}, were grown from synthetic peptide purchased from CS Bio. Crystals were grown at room temperature by hanging drop vaporization. Lyophilized peptide was dissolved in water at 2.9 mg/ml concentration in 48 mM lithium hydroxide. Peptide was mixed in a 2:1 ratio with reservoir containing 0.9 M ammonium phosphate, and 0.1M sodium acetate pH 4.6.

Nanocrystals of NACore, ^{68}GAVVTGVTAV^{78}, were grown from synthetic peptide purchased from CS Bio. Ten batches of synthesized peptide (CSBio) at a concentration of 1 mg/ml in sterile water were shaken at 37°C on a Torrey Pines orbital mixing plate at speed setting 9, overnight. The insoluble material was washed in 30% (w/v) glycerol then stored in water at room temperature before diffraction. The sample contained a mixture of fibrils and crystals.

Nanocrystals of PreNAC, ^{47}GVVHGVTTVA^{56}, were grown from synthetic peptide purchased from InnoPep. Crystallization trials of synthesized peptide were prepared in batch. Peptide was weighed and dissolved in sterile-filtered 50mM phosphate buffer pH 7.0 with 0.1% DMSO at a concentration of 5mg/ml. This solution was shaken at 37°C on a Torrey Pines orbital mixing plate at speed setting 9, overnight.

Data Collection and Processing

X-ray diffraction data from microcrystals of SubNACore were collected using synchrotron radiation at the Advanced Photon Source, Northeast Collaborative Access Team micro focus beam line 24-ID-E. The beam line was equipped with an ADSC Quantum 315 CCD detector. Data from a single crystal were collected in 5° wedges at a wavelength of 0.9791 Å

using a 5 μm beam diameter. We used data from three different sections along the needle axis. The crystals were cryo-cooled (100 K) for data collection. Data were processed and reduced using Denzo/Scalepack from the HKL suite of programs (43).

Electron diffraction data from nanocrystals of NACore and PreNAC were collected using MicroED techniques (25, 26). These nanocrystals typically clump together. To break up the clumps, an approximately 100μL volume of nanocrystals was placed in a sonication bath for 30 minutes. Nanocrystals were deposited onto a Quantifoil holey-carbon EM grid in a 2-3μl drop after appropriate dilution, which optimized for crystal density on the grid. All grids were then blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (FEI), then transferring to liquid nitrogen for storage. Frozen hydrated grids were transferred to a cryo-TEM using a Gatan 626 cryo-holder. Diffraction patterns and crystal images were collected using an FEG-equipped FEI Tecnai F20 TEM operating at 200 kV and recorded using a bottom mount TVIPS F416 CMOS camera with a sensor size of 4000 squared pixels, each 15.6 μm in size per square dimension. Diffraction patterns were recorded by operating the detector in rolling shutter mode with 2 x 2 pixel binning, producing a final image 2000 squared pixels in size. Individual image frames were taken with exposure times of 3-4 seconds per image, using a selected area aperture with an illuminating spot size of approximately one micron. This geometry equates to an electron dose of less than 0.1 e–/Å^2 per second. During each exposure, crystals were continuously rotated within the beam at a rate of 0.3° per second, corresponding to 1.2° wedge per frame. Diffraction data were collected from several crystals each oriented differently with respect to the rotation axis. These data sets each spanned wedges of reciprocal space ranging from 40° to 80°.

X-ray diffraction data from nanocrystals of NACore were collected using XFEL radiation at the CXI instrument (Coherent X-ray Imaging) at the Linear Coherent Light Source (LCLS)-SLAC . The photon energy of the X-ray pulses was 8.52 keV (1.45 Å). Each 40 fs pulse contained up to 6 x 10^{11} photons at the sample position taking into account a beam line transmission of

60%. The diameter of the beam was approximately 1 μm. We used a concentration of approximately 25 μl of pelleted material suspended in 1 mL water. The sample was injected into the XFEL beam using a liquid jet injector and a gas dynamic virtual nozzle (44). The micro jet width was approximately 4 μm and the flow rate was 40 μl/min. The sample caused noticeable sputtering of the liquid jet. XFEL data were processed using cctbx.xfel (45, 46).

Calibration of the sample to detector distance in microED was accomplished using a polycrystalline gold standard and by referencing the prominent reflections in the electron diffraction experiment with the corresponding reflections in the XFEL data. Calibration of the x/y locations of the 64-tile CSPAD detector was performed by *cctbx.xfel* by refining the optically measured tile positions against a thermolysin data set (45).

To gain compatibility with conventional X-ray data processing programs, the microED diffraction images were converted from tiff or TVIPS format to the SMV crystallographic format. We used XDS to index the diffraction images (47), and XSCALE for merging and scaling together data sets originating from different crystals. For NACore, data from four crystals were merged, while for PreNAC, data from three crystals were merged to assemble the final data sets (see Supplementary Table 5.1).

Structure Determination

The molecular replacement solution for SubNACore was obtained using the program Phaser (48). The search model consisted of a geometrically ideal β-strand composed of nine alanine residues. Crystallographic refinements were performed with the program Refmac (49). The molecular replacement solution for NACore was obtained using the program Phaser (48). The search model consisted of the SubNACore structure determined previously. Crystallographic refinements were performed with the program Phenix (50) and Buster (51). The molecular replacement solution for PreNAC was obtained using the program Phaser (48). The search model consisted of a geometrically ideal β-strand composed of six residues with

sequence GVTTVA. Crystallographic refinements were performed with the program Phenix (50) and Refmac (49).

Model building for all segments was performed using COOT (52). Data processing and refinement statistics are reported in Supplementary Table 5.1. The coordinates of the final models and the structure factors have been deposited in the Protein Data Bank with PDB code 4RIK for SubNACore, 4RIL for NACore, and 4ZNN for PreNAC. The structures were illustrated using Pymol (53).

Protein expression and purification

The human wild type α-syn construct was previously characterized (54) (pRK172, ampicillin, T7 promoter) with sequence:
MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHG
VATVAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILE
DMPVDPDNEAYEMPSEEGYQDYEPEA.

Full length α-syn was purified according to published protocols (34). The α-syn construct was transformed into *E.coli* expression cell line BL21 (DE3) gold (Agilent Technologies, Santa Clara, CA) for wild type α-syn protein expression. A single colony was incubated into 100 mL LB Miller broth (Fisher Scientific, Pittsburgh, PA) supplemented with 100 mg/mL ampicillin (Fisher Scientific, Pittsburgh, PA) and grown overnight at 37°C. One liter of LB (Miller) supplemented with 100 mg/mL ampicillin in 2 L shaker flasks was incubated with 10 mL of overnight culture and grown at 37°C until the culture reached an OD600 ~0.6-0.8 as measured by a BioPhotometer UV/VIS Photometer (Eppendorf, Westbury, NY). IPTG (Isopropyl b-D-1-thiogalactopyranoside) was added to a final concentration of 0.5 mM, and grown for 4-6 hours at 30°C. Cells were harvested by centrifugation at 5,500 x g for 10 minutes at 4°C. The cell pellet was frozen and stored at -80°C.

The cell pellet was thawed on ice and resuspended in lysis buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA pH 8.0) and lysed by sonication. Crude cell lysate was clarified

by centrifugation at 15,000 x g for 30 minutes at 4°C. The clarified cell lysate was boiled and cell debris was removed by centrifugation. Protein in the supernatant was precipitated in acid at pH 3.5 through addition of HCl by titration to protein solution on ice while stirring then centrifuged for an additional 15,000 x g for 30 minutes at 4 °C. Supernatant was dialyzed against buffer A (20 mM Tris-HCl, pH 8.0). After dialysis the solution was filtered through a 0.45 μm syringe (Corning, NY 14831) before loading onto a 20 mL HiPrep Q HP 16/10 column (GE Healthcare, Piscataway, NJ). The Q-HP column was washed with five column volumes of buffer A and protein eluted using a linear gradient to 100% in five column volumes of buffer B (20 mM Tris-HCl, 1M NaCl, pH 8.0). Protein eluted at around 50-70% buffer B; peak fractions were pooled. Pooled samples were concentrated approximately tenfold using Amicon Ultra-15 centrifugal filters. Approximately 5 ml of the concentrated sample was loaded onto a HiPrep 26/60 Sephacryl S-75 HR column equilibrated with filtration buffer (25 mM sodium phosphate, 100 mM NaCl, pH 7.5). Peak fractions were pooled from the gel filtration column and dialyzed against 5 mM Tris-HCl, pH 7.5, concentrated to 3 mg/ml. These were filtered through a 0.2 μm pore size filter (Corning, NY 14831) and stored at 4 °C.

Recombinantly expressed full-length α-syn with an N-terminal acetylation was prepared and purified in the following way based on a protocol by Der-Sarkissian *et al.* (16)The α-syn plasmid was co-expressed with a heterodimeric protein acetylation complex from S. pombe to acetylate the N-terminus (pACYC-DUET, chloramphenicol, T7 promoter) (55). The two vectors were co-transformed into E. coli BL21 (DE3) using media containing both ampicillin and chloramphenicol. Cell cultures were grown in TB media containing ampicillin and chloramphenicol and induced to express α-syn with 0.5 mM IPTG overnight at 25°C. Cells were harvested by centrifugation, the cell pellet then resuspended in lysis buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA pH 8.0, and 1 mM phenylmethylsulfonyl fluoride) and cells lysed using an Emulsiflex homogenizer (Avestin). The lysate was boiled and debris removed by centrifugation. A protein fraction was also removed by precipitation at low pH on ice followed by

centrifugation. The remaining supernatant was pH adjusted by titration and dialyzed against Buffer A (20 mM Tris-HCl, pH 8.0, 1 mM DTT, 1 mM EDTA, pH 8.0). The resulting protein solution was loaded onto a 5 mL Q-Sepharose FF column (GE Healthcare) equilibrated with Buffer A and eluted against a linear gradient of Buffer B (1M NaCl, 20 mM Tris-HCl, pH 8.0, 1 mM DTT, 1 mM EDTA, pH 8.0). Fractions containing α-syn were identified using SDS-PAGE, collected, concentrated and further purified by size exclusion (Sephacryl S-100 16/60, GE Healthcare) in 20 mM Tris pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA. Purity of fractions was assessed by SDS-PAGE.

Acetylated protein was characterized by LC-MS (30, 56). Expected average mass: 14460.1 Da for alpha-synuclein and 14502.1 for acetylated alpha-synuclein. Observed average mass: 14464.0 Da for alpha-synuclein and 14506.0 for acetylated alpha-synuclein (Extended Data Figure 6). The shift of 4 Da between observed and expected average masses is due to instrumental error.

Fibril formation and detection

Purified α-syn in 50mM Tris, 150 mM KCl pH 7.5 was shaken at a concentration of 500 μM at 37°C in Torey Pine shaker. To form the fibrillar samples of SubNACore and NACore, lyophilized peptides were dissolved to a final concentration of 500 μM in 5 mM lithium hydroxide, 20 mM sodium phosphate pH 7.5 and 0.1 M NaCl. All samples were shaken at 37°C in a Torey Pine shaker for 72 hours. Freshly dissolved samples were prepared by dissolving lyophilized peptides immediately prior to addition to cells for assays.

Turbidity measurements were used to compare NACore and SubNACore aggregation. Peptide samples were freshly dissolved to 1.6mM in a sample buffer with 5mM LiOH and 1% DMSO and then filtered through a PVDF filter (Millipore, 0.1μm). Measurements were performed using a black NUNC 96 well plate with 200μL of sample/well (3-4 replicates per sample). The plate was agitated at 37°C, with a 3mm rotation diameter, at 300rpm in a

Varioskan microplate reader (Thermo). Absorbance readings were recorded every 3-15 minutes at 340nm.

Negative Stain Transmission Electron Microscopy

Cytotoxicity samples were evaluated for presence of fibrils by electron microscopy. Briefly, 5 µL samples were spotted directly on freshly glow-discharged carbon-coated electron microscopy grids (Ted Pella, Redding, CA). After 4 min incubation, grids were rinsed twice with 5-µL distilled water and stained with 2% uranyl acetate for 1 min. Specimens were examined on an FEI T12 electron microscope.

Fibril Diffraction

Fibrils formed from purified α-syn with and without N-terminal acetylation were concentrated by centrifugation, washed, and oriented while drying between two glass capillaries. Likewise, NACore nanocrystals were also concentrated, washed with nanopure water, and allowed to orient while drying between two glass capillaries. The glass capillaries holding the aligned fibrils or nanocrystals were mounted on a brass pin for diffraction at room temperature using 1.54Å x-rays produced by a Rigaku FRE+ rotating anode generator equipped with an HTC imaging plate. All patterns were collected at a distance of 180 mm and analyzed using the Adxv software package (57). A simulated pattern from the full length α-syn model presented in Figure 3 was obtained by calculating structure factors from the model using the sfall module from CCP4, assigning the model a unit cell of 200 x 4.74 x 200 angstroms. Cylindrical averaging of these structure factors about the fiber axis (y axis) direction produced a set of simulated fibril diffraction intensities.

Cytotoxicity assays

Adherent PC12 cells were cultured in ATCC-formulated RPMI 1640 medium (ATCC; cat.# 30-2001) supplemented with 10% horse serum and 5% fetal bovine serum and plated at 10,000 per well to a final volume of 90 µl. All MTT assays were performed with Cell Titer 96 aqueous non-radioactive cell proliferation kit (MTT, Promega cat #4100). Cells were cultured in

96-well plates for 20h at 37°C in 5% CO_2 prior to addition of samples (Costar cat. # 3596). 10 µL of sample was added to each well containing 90 µL medium and incubated for 24h at 37°C in 5% CO_2. Then, 15 µL dye solution (Promega cat #4102) was added into each well, followed by incubation for 4h at 37°C in 5% CO_2. This was followed by the addition of 100µl solubilization Solution/Stop Mix (Promega cat #4101) to each well. After 12h incubation at room temperature, the absorbance was measured at 570nm. Background absorbance was recorded at 700nm. The data was normalized with cells treated with 1% (w/v) SDS to 0% reduction, and cells treated with sample buffer to 100% reduction.

Lactose dehydrogenase assays were done using CytoTox-ONE™ Homogeneous Membrane Integrity, (Promega, cat #G7890) as per manufacturers instructions. Briefly, cells were plated in 96 well black-wall, clear bottom (Fisher Cat #07-200-588) tissue culture plates at 10,000 cells per well to a final volume of 90µL. Cells were incubated for an additional 20h at 37°C in 5% CO_2 prior to addition of samples. Next, 10µl of sample was added to each well following which the cells were incubated for another 24 hours. 100µl of reagent was added to each well and incubated for 15mins at room temperature. The addition of 50µL of stop solution stopped the reaction. Fluorescence was measured in a Spectramax M5 (Molecular Devices) using excitation and emission wavelengths of 560 nm and 590 nm, respectively. Data was normalized using cells treated with buffer as 0% release and 0.1% triton X-100 as 100% release.

Construction of α-syn A53T fibril model

A model for full-length α-syn A53T mutant fibrils that are involved in the early onset of PD was constructed using a section of the NACore crystal packing as a scaffold. Figure 5.3 illustrates the four copies of the NACore segment used for the scaffold. The crystal structure of the two inner strands was adapted with minimal changes as the analogous segments 68-78. The structure of PreNAC was matched onto the weak interface of the NACore structure. Only four of the eleven sidechains in the segment 46-56 differ from those in the NACore segment 68-78 and residues V51-V55 can be closely matched to V71-V75. Hence the model for both the homotypic

interface and heterotypic interface in the full-length fiber model closely resemble those observed in the NACore structure. The regions outside these segments were adapted from the structure of the native α-syn fold, PDB ID 2KKW (58). These segments were spliced in manually using COOT. The models were energy minimized and temperature annealed using the program CNS (59) with hydrogen bonding potential (60). The simulated fiber diffraction pattern calculated from this model shows prominent reflections that agree with those observed in fiber diffraction patterns of NACore, α-syn, and N-acetyl α-syn (Supplementary Table 5.2).

ACKNOWLEDGEMENTS

We thank Cong Liu for supplying PC12 cells; APS staff for beam line help solving SubNACore: Malcom Capel, Kanagalaghatta Rajashankar, Narayanasami Sukumar, Jon Schuermann, Igor Kourinov and Frank Murphy at NECAT beam lines 24-ID at APS funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403) and the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We thank the LCLS injection staff support: Sabine Botha, Robert Shoeman, and Ilme Schlichting. A.S.B. and N.K.S. were supported by NIH grants GM095887 and GM102520 and by the Director, Office of Science, Department of Energy (DOE) under contract DE-AC02-05CH11231 for data-processing methods. This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under award number DE-FC02-02ER63421. We also thank the award MCB-0958111 from the National Science Foundation, award 1R01-AG029430 from the National Institutes of Health, award NIH-AG016570 from Alzheimer's Disease Research (ADRC) at UCLA, and HHMI for support. J.A.R. was supported by the Giannini Foundation.

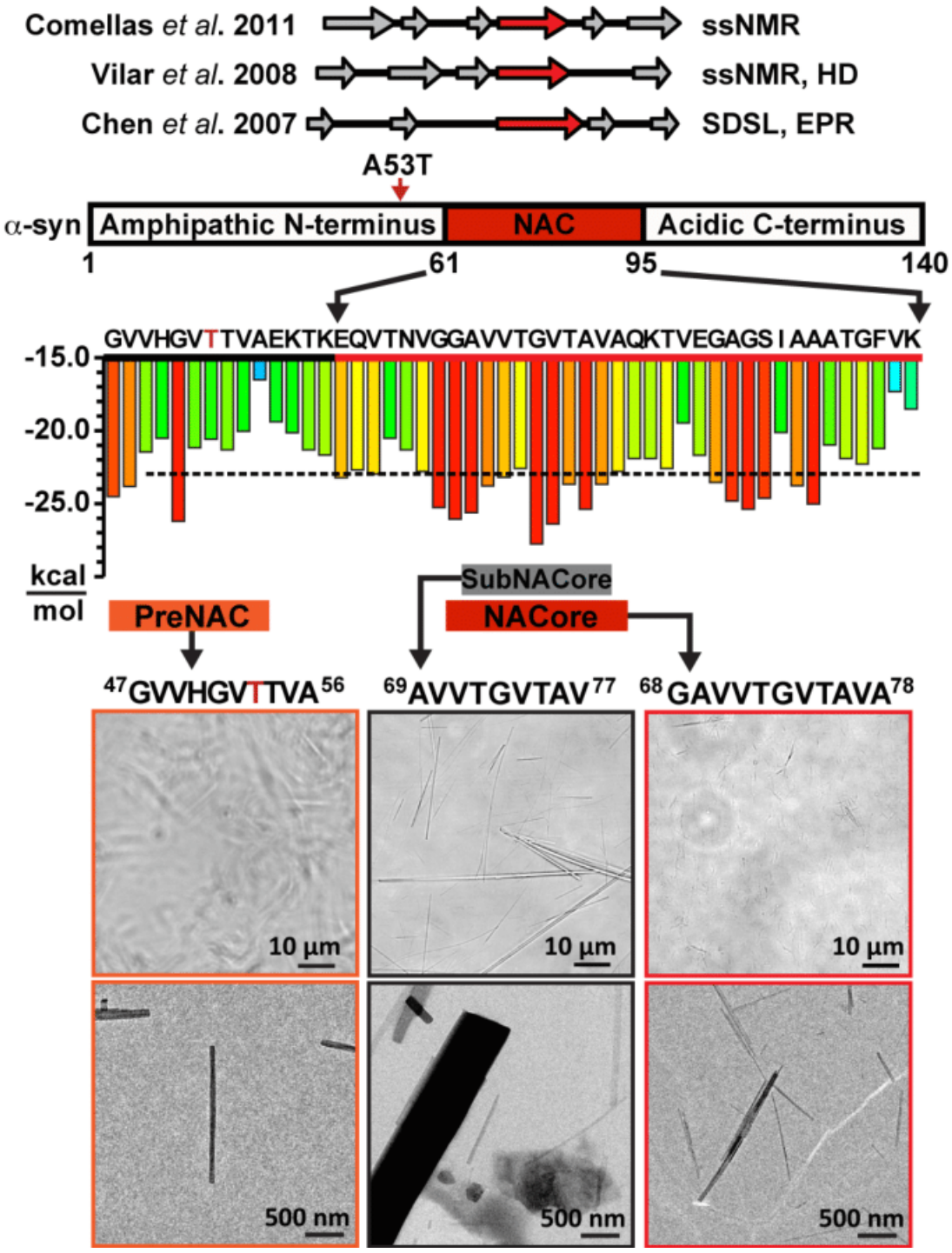


Figure 5.1. NACore (residues 68-78) is the fibril-forming core of the NAC domain of full-length α-syn. Top: beta strands indicated byNMR and EPR (17, 40, 41). Center: red 6-residue segments are predicted to form amyloid fibrils (42). The A53T early-onset Parkinson mutation is indicated by a red arrow and red letter T. Bottom: the miniscule size of the preNAC

and NACore crystals used for MicroED is illustrated by this comparison to SubNACore microcrystals (middle. Scale comparisons are illustrated on two magnifications using phase contrast light microscope images and electron micrographs , in which individual NACore and PreNAC nanocrystals are indistinguishable by light microscopy.

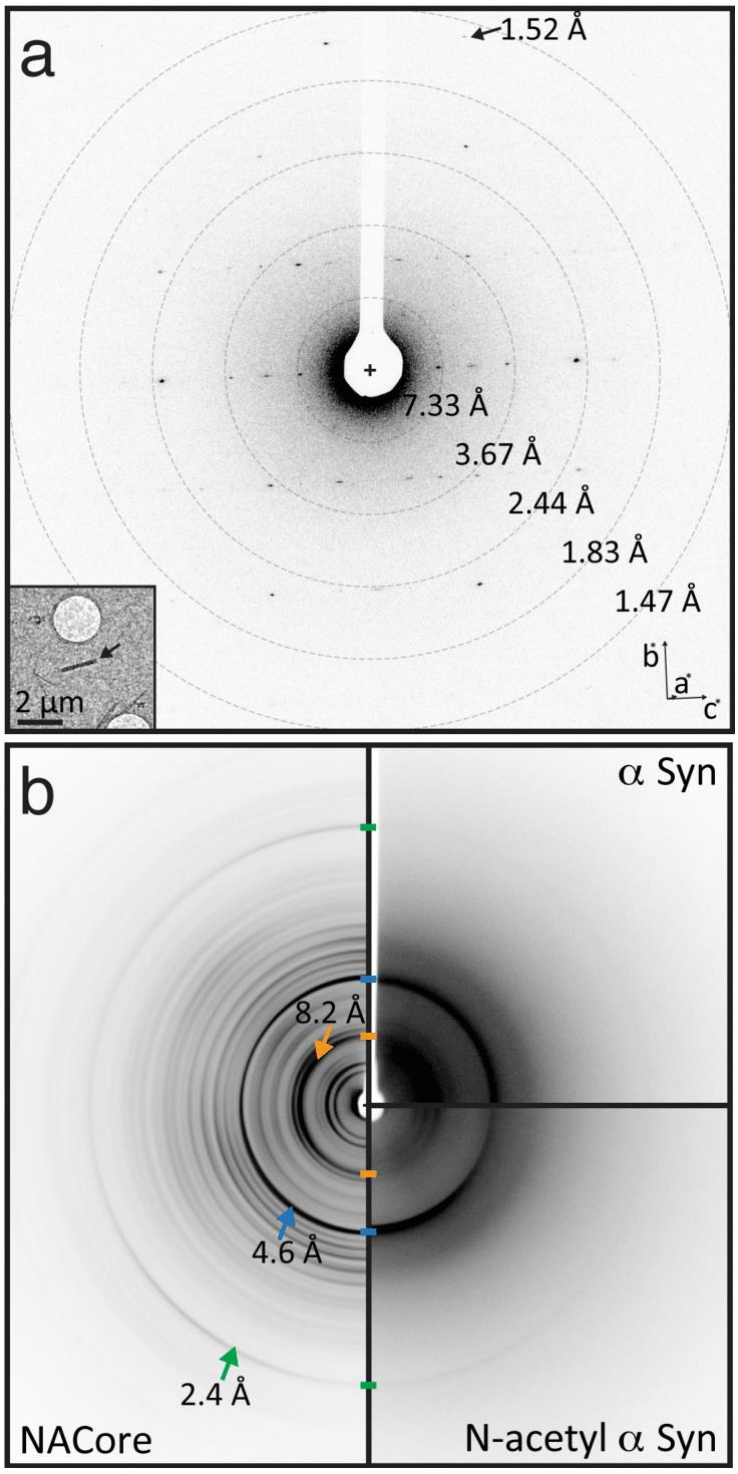


Figure 5.2. Diffraction from NACore nano crystals is similar to that from full length α-syn fibrils. (A) Single crystal electron diffraction pattern obtained duringMicroED data collection (see text). Equally spaced concentric rings denote resolution shells. An arrow

points to The highest resolution spot is at 1.52 Å (arrow). The inset shows the overfocused image of the diffracting crystal (arrow), which is ~1480 x 200 x 200 nm. (B) Composite of fibril diffraction patterns from α-syn preparations and NACore. Full-length α-syn reveals reflections that match those from NACore and N-terminally acetylated α-syn. The two patterns of full-length α-syn share with NACore three major peaks denoted by arrows: 8.2 Å (orange), 4.6 Å (blue), and 2.4 Å (green). The origin of these peaks can be traced to the (0 0 2), [(1,1,1),(-1 1 1)], and (0 2 0) planes in the NACore structure, respectively. The reflection at 8.2Å likely arises from adjacent pairs of β-sheets.

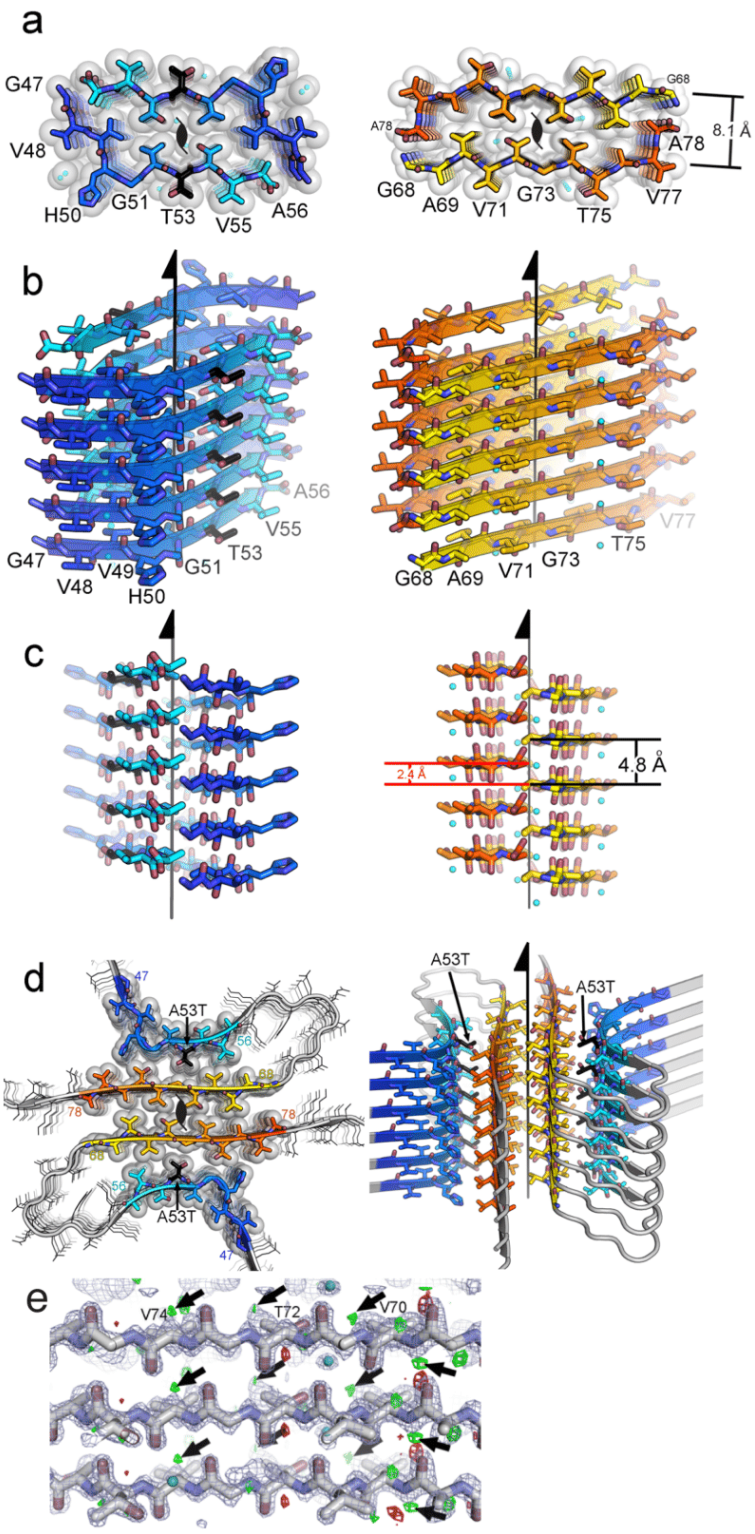


Figure 5.3. Structure of the amyloid core of α-syn. (A) The crystal structure of NACore (orange) reveals pairs of sheets as in the spines of amyloid fibrils. The A53T mutation in PreNAC

is shown in black. The sheets in both structures are related by the 2_1 fibril axes shown in black. The gaps left by the interface are filled with water molcules which hydrogen-bond to the threonine residues (partly showing aqua spheres). (B) and (C), are orthogonal views of the fibrillar assemblies. (D) A speculative model of an α-syn protofibril containing the A53T mutation (black), where the strong interface of NACore (orange) forms the core of the fibril and its weaker interface interacts with PreNAC (blue). (E) The locations of 5 out of a possible 73 protons are suggested by small, positive Fo-Fc density (green contoured at 2.8s, shown by arrows. The blue mesh is 2Fo-Fc density contoured at 1.4 s.

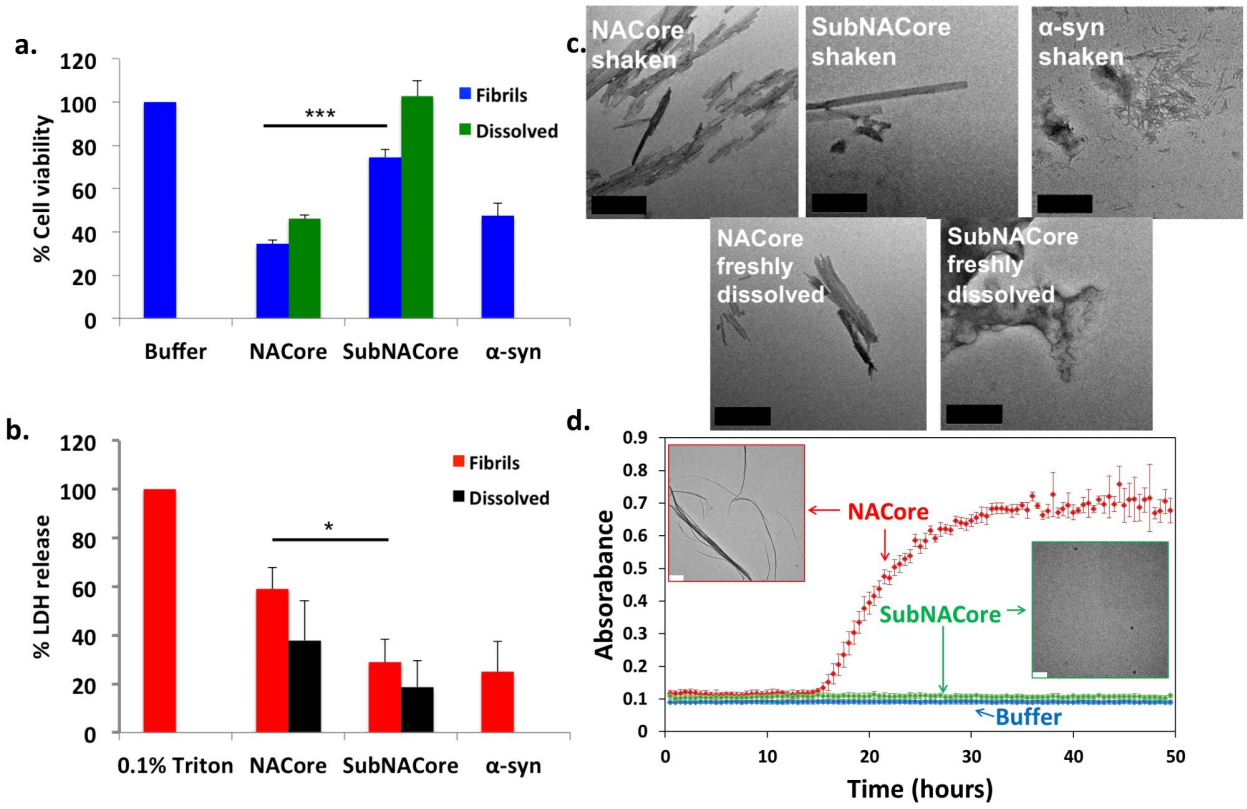
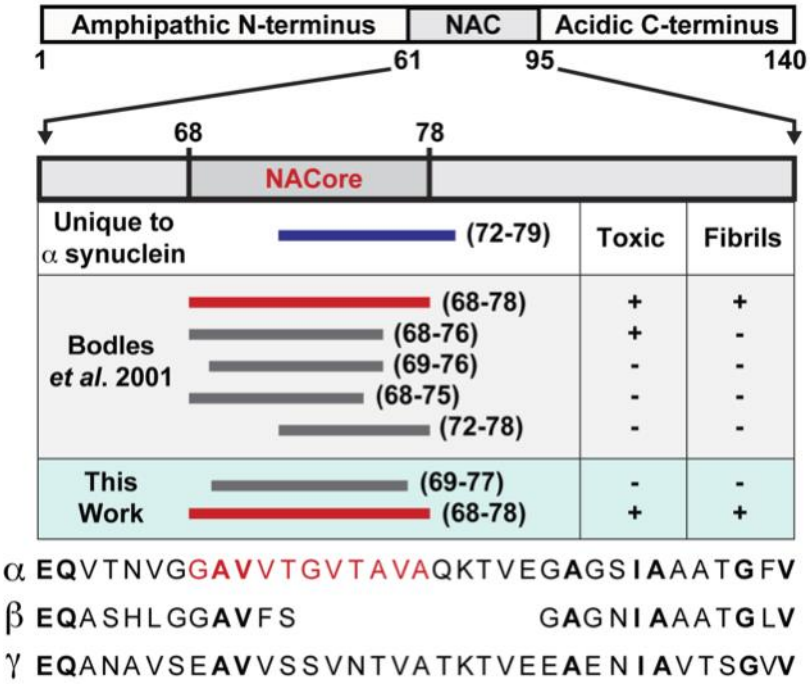
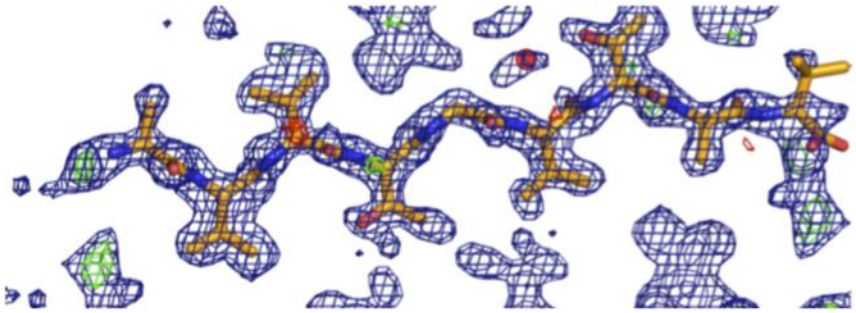


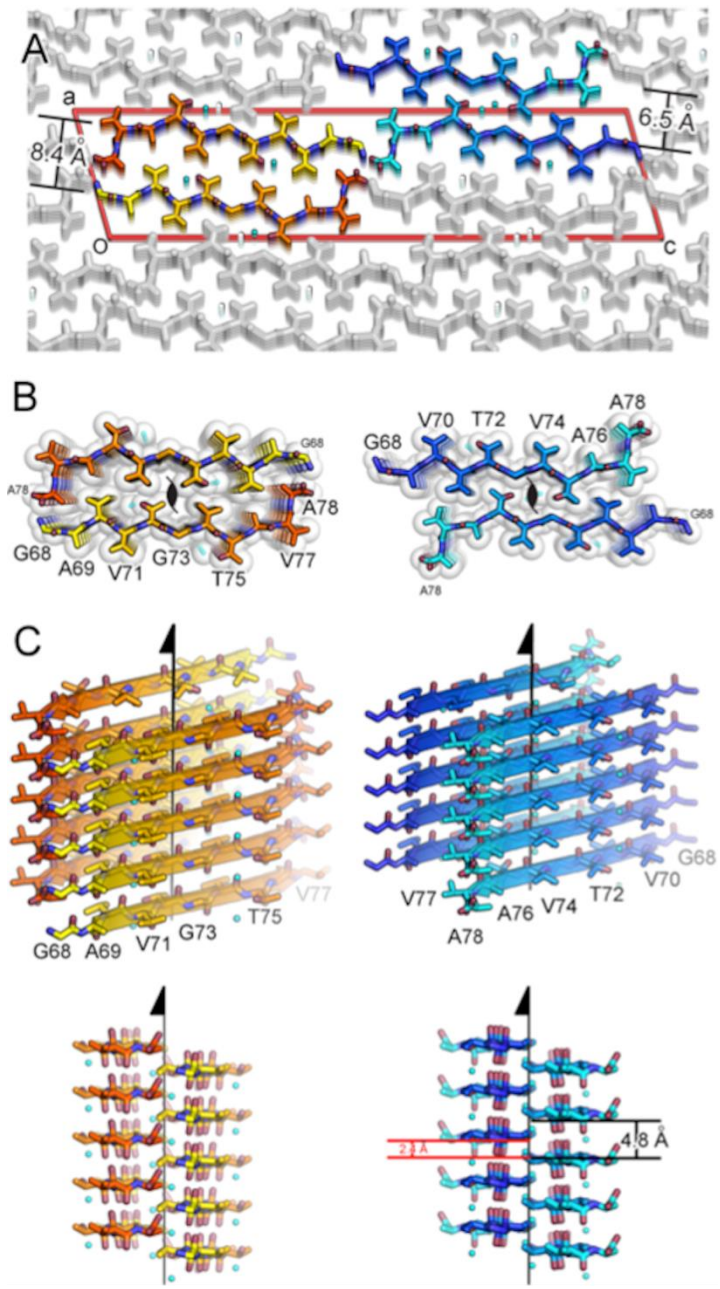
Figure 5.4. NACore aggregates faster than SubNACore and is more cytotoxic to cultured cells. Cytotoxicity of NACore, SubNACore and α-syn measured on PC12 cells using (A) MTT assay and (B) LDH release assay. In both assays NACore is more toxic than SubNACore. Also, shaken fibrils are more toxic than an equal concentration of freshly dissolved sample. Results shown as mean ± standard error of the mean from triplicate samples. A *t-test* was used to measure statistical significance * <0.05, **<0.01,***<0.001. (C) Representative electron micrographs of NACore, SubNACore and α-syn samples tested for cytotoxicity. NACore and α-syn show abundant fibrils but SubNACore few. NACore also forms fibrils immediately upon dissolving whereas SubNACore shows no fibers, but instead amorphous aggregates. Scale 500nm. (D) NACore and SubNACore were aggregated in identical conditions and monitored by turbidity. NACore begins to aggregate in 15 hours while SubNACore forms no aggregates for up to 50 hours. Electron microscopy of the samples at 50 hours confirmed the turbidity readings (insets, scale 2 μm), with error bars denoting standard deviation from triplicates.



Supplementary Figure 5.1. A schematic representation of α-syn, highlighting the NAC region (residues 61-95) and within it the NACore sequence (residues 68-78). A series of bars span regions of α-syn that are of interest to this work. Among the three synuclein paralogs (α, β, and γ), the region whose sequence is unique to α-syn is shown as a blue bar (residues 72-83) that overlaps with a large portion of NACore. Segments investigated by Bodles *et al.*[23] are also shown. These span a variety of regions within NACore. Two of the segments we now investigate, SubNACore and NACore, are shown in this context. Only one of the segments studied by Bodles *et al.* is an exact match to our NACore sequence, and only this segment is both toxic and fibrillar. The sequences of α-synuclein, β-synuclein, and γ-synuclein are shown as a reference with conserved residues in bold and the NACore sequence in red.

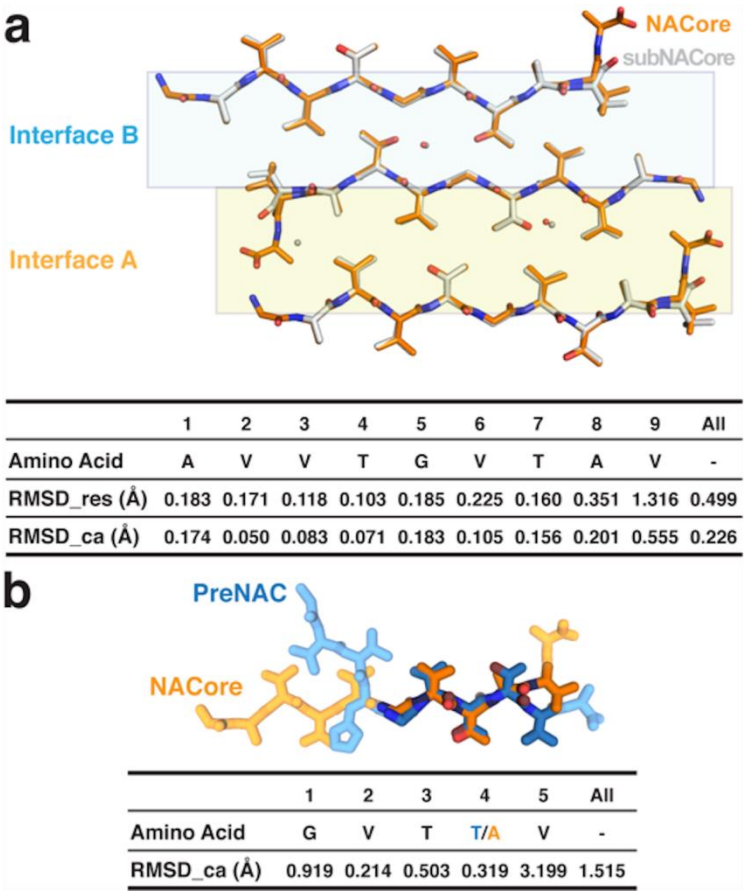


Supplementary Figure 5.2. **Difference density maps** calculated after successful molecular replacement using the SubNACore search model clearly revealed the positions of the missing residues (positive F_o-F_c density at N and C termini corresponding to G68 and A78) and one water molecule near a threonine side chain (red circle); a second water was located during the refinement process. The blue mesh represents $2F_o$-F_c density contoured at 1.2 s level. The green and red mesh represent F_o-F_c density contoured at 3.0 and -3.0 s, respectively. All maps were sa-weighted (61).



Supplementary Figure 5.3. The crystal structure of NACore reveals pairs of sheets as in the spines of amyloid fibrils. (A) NACore's two types of sheet-sheet interfaces: a larger interface (orange, 268 Å^2 of buried accessible surface area per chain) we call interface A, and a weaker interface (blue, 167 Å^2) we call interface B. The crystal is viewed along the hydrogen-bonding direction (crystal "b" dimension). The red lines outline the unit cell. (B) The van der Waals packing between sheets.The sheets are related by a 2_1 screw axis denoted in black.

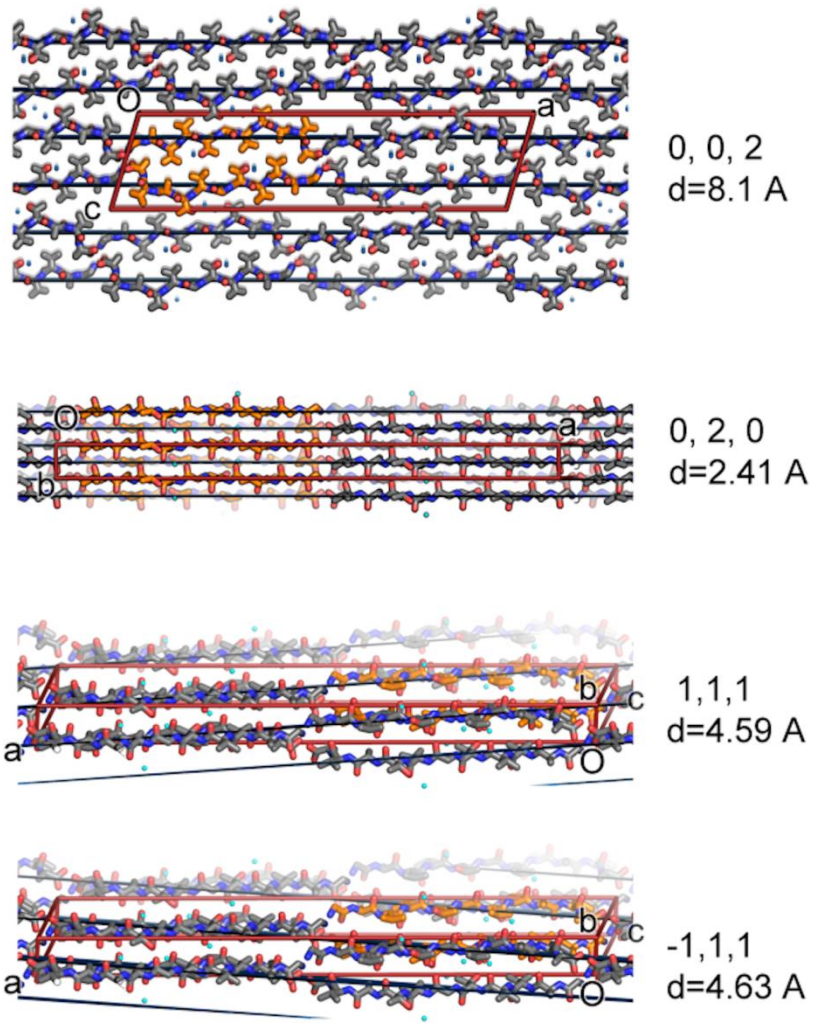
The only gaps left by the interface are filled with water molcules which hydrogen-bond to the threonine residues (partly showing aqua spheres). The shape complementarity of both interfaces is 0.7. The viewing direction is the same as in (A, C) Orthogonal view of the fibrillar assembly. The protofibril axis, coinciding with the 2_1 screw axis designated by the arrow, runs vertically between the pairs of sheets.



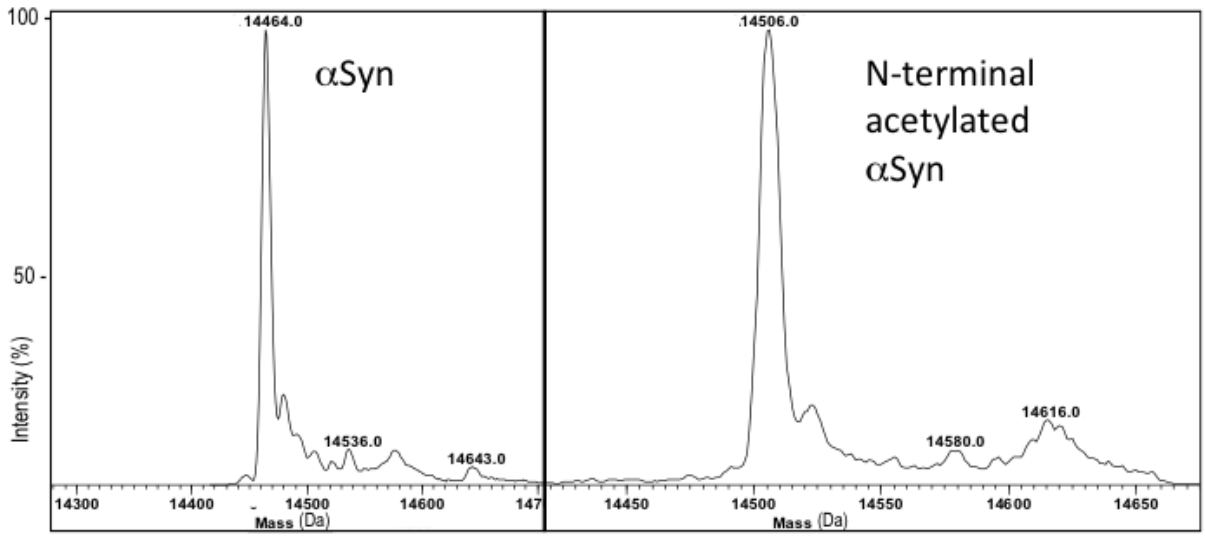
	1	2	3	4	5	6	7	8	9	All
Amino Acid	A	V	V	T	G	V	T	A	V	-
RMSD_res (Å)	0.183	0.171	0.118	0.103	0.185	0.225	0.160	0.351	1.316	0.499
RMSD_ca (Å)	0.174	0.050	0.083	0.071	0.183	0.105	0.156	0.201	0.555	0.226

	1	2	3	4	5	All
Amino Acid	G	V	T	T/A	V	-
RMSD_ca (Å)	0.919	0.214	0.503	0.319	3.199	1.515

Supplementary Figure 5.4. (A) **Comparison of the crystal packing for NACore (orange chain) and SubNACore (white chain).** The face-to-face interactions are virtually the same for the pairs of NACore peptides in its crystal structure and the SubNACore peptides in its structure (A and B shown in gold and blue, respectively). The table below shows a pairwise the RMSD comparing the nine residues shared in common between the structures. RMSD_res is an all-atom comparison between residue pairs, while RMSD_ca compares only alpha carbon pairs. (B) PreNAC (blue) is compared with NACore (orange). Five residues from each strand are shown in darker color and the RMS deviations between their alpha carbon pairs compared in the table below. Therefore the PreNAC-NACore interaction mimics weaker Interface B in the NACore structure.



Supplementary Figure 5.5. Intense reflections common among the NACore and the two polymorphs of full length α-syn suggest common structural features. These features are illustrated here on the crystal packing diagrams of NACore. The (0,0,2) planes approximate the separation between sheets in interface A (orange). The (0,2,0), (-1,1,1), and (1,1,1) reflections are intense because the corresponding Bragg planes recapitulate the staggering of strands from opposing sheets. The red lines correspond to the unit cell boundaries and all planes are shown in black.



Supplementary Figure 5.6. **Mass spectrometry analysis of recombinantly expressed, full length α-syn**, with and without N-terminal acetylation The mass profile of wild-type full length α-syn (left) is compared to that of an N-terminally acetylated form of the protein (right). The mass shift for the N-terminally acetylated form is appropriately shifted with respect to the native form of the protein (14464.0 Da for alpha-synuclein and 14506.0 for acetylated alpha-synuclein), within a margin of error of 4 Da.

Supplementary Table 5.1. Statistics of data collection and atomic refinement for NACore, its fragment SubNACore, and PreNAC. *Highest resolution shell is shown in parenthesis.

Segment	SubNACore $_{69}$AVVTGVTAV$_{77}$	NACore $_{68}$GAVVTGVTAVA$_{56}$	PreNAC $_{47}$GVVHGVTTVA$_{56}$
Data collection			
Radiation source	Synchrotron	Electron	Electron
Space group	C2	C2	P21
Cell dimensions			
a,*b*,*c* (Å)	61.9, 4.80, 17.3	70.8, 4.82, 16.79	17.9, 4.7, 33.0
α,*β*,*γ* (°)	90, 104.1, 90	90, 105.7, 90	90, 94.3, 90
Resolution (Å)	1.85 (1.95-1.85)	1.43 (1.60-1.43)	1.41 (1.56-1.41)
Wavelength (Å)	0.9791	0.0251	0.0251
R_{merge}	0.117 (0.282)	0.173 (0.560)	0.236 (0.535)
$R_{r.i.m.}$	0.135 (0.322)	0.199 (0.647)	0.264 (0.609)
$R_{p.i.m.}$	0.065 (0.154)	0.093 (0.311)	0.185 (0.305)
$I/\sigma I$	5.2 (2.7)	5.5 (2.5)	4.6 (1.8)
$CC_{1/2}$ (%)	99.5 (97.8)	99.4 (92.3)	96.7(74.0)
Completeness (%)	97.9 (98.3)	89.9 (82.6)	86.9 (69.6)
Multiplicity	4.1 (4.0)	4.4 (4.3)	3.7 (3.5)
Refinement			
Resolution (Å)	1.85 (2.07-1.85)	1.43 (1.60-1.43)	1.41 (1.41-1.57)
No. reflections	470 (125)	1073 (245)	1006 (239)
R_{work}	0.176 (0.248)	0.248 (0.253)	0.235 (0.336)
R_{free}	0.221 (0.286)	0.275 (0.331)	0.282 (0.329)
CC_{work}	0.964 (0.896)	0.947(0.618)	0.937(0.335)
CC_{free}	0.889 (0.993)	0.986(0.269)	0.967(0.361)
No. atoms			
Protein	57	66	66
Water	3	2	4
B-factors (Å^2)			
Protein	17.1	9.0	16.1
Water	27.6	2.7	24.6
Wilson B (Å^2)	11.8	10.3	13.8
R.m.s deviations			
Bond lengths (Å)	0.005	0.010	0.020
Bond angles (°)	1.1	1.6	2.0
PDB ID code	4RIK	4RIL	4ZNN
EMDB ID code		EMD-3028	EMD-3001

Segment	Reflections (Å)
NACore GAVVTGVTAVA	2.21, 2.26, **2.39**, 2.52, 2.61, 2.68, 2.78, 3.02, 3.12, 3.34, 3.56, 3.86, 4.34, **4.57**, 5.16, 5.98, 7.56, **8.19**, 10.46, 11.63, 13.29, 16.61
α-syn	**2.39**, **4.64**, 6.82, **8.29**, 10.06
N-acetyl α-syn	**2.38**, **4.62**, **8.18**, 9.80, 11.90
Simulated α-syn	2.23, 2.25, **2.35**, 3.29, 3.63, 3.70, 3.95, 4.08, 4.56, **4.68**, **8.36**, 8.69, 21.76, 24.47, 27.61, 31.67

Supplementary Table 5.2. Comparison of reflections observed in powder diffraction of fibrils of full-length α-syn, N-acetyl α-syn, and a synthetic pattern calculated from our α-syn model, to aligned nanocrystals of NACore. Bold reflections are strong and common to all three samples.

REFERENCES

1. Spillantini, M. G. *et al.* Alpha-synuclein in Lewy bodies. *Nature* **388,** 839–840 (1997).

2. Goedert, M., Spillantini, M. G., Del Tredici, K. & Braak, H. 100 years of Lewy pathology. *Nat. Rev. Neurol.* **9,** 13–24 (2013).

3. Polymeropoulos, M. H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276,** 2045–2047 (1997).

4. Krüger, R. *et al.* Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.* **18,** 106–108 (1998).

5. Zarranz, J. J. *et al.* The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **55,** 164–173 (2004).

6. Ibáñez, P. *et al.* Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364,** 1169–1171 (2004).

7. Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302,** 841 (2003).

8. Uéda, K. *et al.* Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 11282–11286 (1993).

9. Biere, A. L. *et al.* Parkinson's disease-associated alpha-synuclein is more fibrillogenic than beta- and gamma-synuclein and cannot cross-seed its homologs. *J. Biol. Chem.* **275,** 34574–34579 (2000).

10. Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q. & Lee, V. M.-Y. A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of α-Synuclein Is Essential for Filament Assembly. *J. Biol. Chem.* **276,** 2380–2386 (2001).

11. Du, H.-N. *et al.* A peptide motif consisting of glycine, alanine, and valine is required for the fibrillization and cytotoxicity of human alpha-synuclein. *Biochemistry (Mosc.)* **42,** 8870–8878 (2003).

12. Periquet, M., Fulga, T., Myllykangas, L., Schlossmacher, M. G. & Feany, M. B. Aggregated α-Synuclein Mediates Dopaminergic Neurotoxicity In Vivo. *J. Neurosci.* **27,** 3338–3346 (2007).

13. Han, H., Weinreb, P. H. & Lansbury, P. T. The core Alzheimer's peptide NAC forms amyloid fibrils, which seed and are seeded by beta-amyloid: is NAC a common trigger or target in neurodegenerative disease? *Chem. Biol.* **2,** 163–169 (1995).

14. El-Agnaf, O. M. *et al.* Aggregates from mutant and wild-type alpha-synuclein proteins and NAC peptide induce apoptotic cell death in human neuroblastoma cells by formation of beta-sheet and amyloid-like filaments. *FEBS Lett.* **440,** 71–75 (1998).

15. Crowther, R. A., Daniel, S. E. & Goedert, M. Characterisation of isolated alpha-synuclein filaments from substantia nigra of Parkinson's disease brain. *Neurosci. Lett.* **292,** 128–130 (2000).

16. Der-Sarkissian, A., Jao, C. C., Chen, J. & Langen, R. Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. *J. Biol. Chem.* **278,** 37530–37535 (2003).

17. Chen, M., Margittai, M., Chen, J. & Langen, R. Investigation of alpha-synuclein fibril structure by site-directed spin labeling. *J. Biol. Chem.* **282,** 24970–24979 (2007).

18. Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. Biochemical Characterization of the Core Structure of α-Synuclein Filaments. *J. Biol. Chem.* **277,** 19213–19219 (2002).

19. Conway, K. A., Harper, J. D. & Lansbury, P. T. Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to early-onset Parkinson disease. *Nat. Med.* **4,** 1318–1320 (1998).

20. Polymeropoulos, M. H. *et al.* Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276,** 2045–2047 (1997).

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

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

23. Bodles, A. M., Guthrie, D. J., Greer, B. & Irvine, G. B. Identification of the region of non-Abeta component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity. *J. Neurochem.* **78,** 384–395 (2001).

24. Nannenga, B. L. & Gonen, T. Protein structure determination by MicroED. *Curr. Opin. Struct. Biol.* **27C,** 24–31 (2014).

25. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11,** 927–930 (2014).

26. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2,** e01345 (2013).

27. Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E. & Gonen, T. Structure of catalase determined by MicroED. *eLife* **3,** e03600 (2014).

28. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3368–3373 (2015).

29. Doyle, P. A. & Turner, P. S. Relativistic Hartree–Fock X-ray and electron scattering factors. *Acta Crystallogr. Sect. A* **24,** 390–397 (1968).

30. Sarafian, T. A. *et al.* Impairment of mitochondria in adult mouse brain overexpressing predominantly full-length, N-terminally acetylated human α-synuclein. *PloS One* **8,** e63557 (2013).

31. Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26,** 267–298 (2003).

32. Danzer, K. M., Schnack, C., Sutcliffe, A., Hengerer, B. & Gillardon, F. Functional protein kinase arrays reveal inhibition of p-21-activated kinase 4 by alpha-synuclein oligomers. *J. Neurochem.* **103,** 2401–2407 (2007).

33. Karpinar, D. P. *et al.* Pre-fibrillar alpha-synuclein variants with impaired beta-structure increase neurotoxicity in Parkinson's disease models. *EMBO J.* **28,** 3256–3268 (2009).

34. Winner, B. *et al.* In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 4194–4199 (2011).

35. Chen, S. W. *et al.* Structural characterization of toxic oligomers that are kinetically trapped during α-synuclein fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E1994–2003 (2015).

36. Bousset, L. *et al.* Structural and functional characterization of two alpha-synuclein strains. *Nat. Commun.* **4,** 2575 (2013).

37. Auluck, P. K., Caraveo, G. & Lindquist, S. α-Synuclein: membrane interactions and toxicity in Parkinson's disease. *Annu. Rev. Cell Dev. Biol.* **26,** 211–233 (2010).

38. Lee, J. C., Langen, R., Hummel, P. A., Gray, H. B. & Winkler, J. R. Alpha-synuclein structures from fluorescence energy-transfer kinetics: implications for the role of the protein in Parkinson's disease. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 16466–16471 (2004).

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

40. Comellas, G. *et al.* Structured regions of α-synuclein fibrils include the early-onset Parkinson's disease mutation sites. *J. Mol. Biol.* **411,** 881–895 (2011).

41. Vilar, M. *et al.* The fold of alpha-synuclein fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 8637–8642 (2008).

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

43. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* (ed. Charles W. Carter, J.) **Volume 276,** 307–326 (Academic Press, 1997).

44. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83,** 035108 (2012).

45. Hattne, J. *et al.* Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat. Methods* **11,** 545–548 (2014).

46. Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. New Python-based methods for data processing. *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1274–1282 (2013).

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

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

49. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).

50. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 352–367 (2012).

51. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2210–2221 (2004).

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

53. Delano, W. *The PyMOL Molecular Graphics System*. (Schrödinger LLC). at http://www.pymol.org

54. Jakes, R., Spillantini, M. G. & Goedert, M. Identification of two distinct synucleins from human brain. *FEBS Lett.* **345,** 27–32 (1994).

55. Johnson, M., Coulton, A. T., Geeves, M. A. & Mulvihill, D. P. Targeted amino-terminal acetylation of recombinant proteins in E. coli. *PloS One* **5,** e15801 (2010).

56. Whitelegge, J. P., Zhang, H., Aguilera, R., Taylor, R. M. & Cramer, W. A. Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein: cytochrome b(6)f complex from spinach and the cyanobacterium Mastigocladus laminosus. *Mol. Cell. Proteomics MCP* **1,** 816–827 (2002).

57. Arvai, A. *Adxv - A Program to Display X-ray Diffraction Images*. (2015). at http://www.scripps.edu/tainer/arvai/adxv.html

58. Rao, J. N., Jao, C. C., Hegde, B. G., Langen, R. & Ulmer, T. S. A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *J. Am. Chem. Soc.* **132,** 8657–8668 (2010).

59. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2,** 2728–2733 (2007).

60. Fabiola, F., Bertram, R., Korostelev, A. & Chapman, M. S. An improved hydrogen bond potential: impact on medium resolution protein structures. *Protein Sci. Publ. Protein Soc.* **11,** 1415–1423 (2002).

61. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42,** 140–149 (1986).