UNIVERSITY OF CALIFORNIA

Los Angeles

Molecular Snapshots of the Mammalian Prion CPEB3

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

by

Maria Flores

2023

ABSTRACT OF THE DISSERTATION

Molecular Snapshots of the Mammalian Prion CPEB3

by

Maria Flores

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2023

Professor Jose A. Rodriguez, Chair

The cytoplasmic polyadenylation element-binding protein 3 (CPEB3) is a functional prion thought to modulate protein synthesis at synapses and enable consolidation of long-term memory in neurons. The biophysical characteristics of the prion state may facilitate the endurance of CPEB3 in environments with high molecular turnover and enable rapid aggregation and localization of targets to subcellular sites. Truncations to the CPEB3 prion-like domain cause loss of stimulation induced protein upregulation and impede long-term potentiation in mice, highlighting the critical role of its putative prion conformation.

The prion state often requires polypeptides to adopt beta-sheet rich conformations of amyloid to generate and propagate faithful conformations in fluctuating environments,

an alluring feature if controlled properly. Because of their association with neurodegenerative diseases, functional amyloid proteins and their regulation in the refined structure of the brain are not entirely understood. CPEB3 undergoes extensive post-translational modifications that control its subcellular localization and activity. Proper shuttling of the prion-like protein from RNA storage granules to polysomes is thought to be a critical component in its ability to exist within the homeostasis of neurons.

The work presented in this dissertation aims to determine the ability of CPEB3's prion-like domain to form amyloid conformations and the role of this conformation in translation regulation. Through in vitro experiments investigating the ordered amyloid core of CPEB3, reversibility and lack of stability are observed as key differences in CPEB3 fibrils when compared to pathogenic counterparts. The data provide insights into the reversible amyloid core and show its requirement for proper storage and translation regulation in neurons. These results bridge the studies from in vitro generated structure to supramolecular assemblies observed inside of neurons. The native environments and structured complexes begin to show potential conformations of CPEB3 that are complementary to its function.

The dissertation of Maria Flores is approved.

David S. Eisenberg

Joseph Ambrose Loo

Alcino J. Silva

Jose Alfonso Rodriguez, Committee Chair

University of California, Los Angeles

2023

DEDICATION

Para mi mamá y mi papá.

Por el gran ejemplo de la perseverancia, por sus consejos, apoyo, y amor durante todas las épocas de mi vida.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am indebted to all of those who have in any way contributed and supported me during the period of this work.

First, I would like to thank my advisor, Prof. Jose Rodriguez, for the gift of independence and his laissez-faire approach on my thesis work that taught me critical lessons on failure and perseverance throughout my graduate career. The members of my committee, Prof. Alcino Silva, Prof. David Eisenberg, and Prof. Joseph Loo for their time, advice and feedback during the course of this work.

I would like to thank all my lab mates, past and present for their many fruitful scientific discussions. In particular, Logan Richards, for his support throughout as my cellmate and intellectual contributions in phasing MicroED data. Ambarneil Saha, for his support throughout as my cellmate and for his enthusiastic and brilliant explanations regarding electron microscopes and all things reciprocal space. Shervin Nia, for his engaging discussions as I was learning about tomographic reconstruction. Sam Zink, for her support throughout as my cellmate and intellectual contributions in MD simulations.

I would like to thank our collaborator Dr. Luana Fioriti, for teaching me so much more about CPEB and how to approach biological questions. For her intuitive suggestions, guidance and refreshing and consistent interest for all results, positive or negative.

I would like to thank our collaborator, Dr. Gustavo Helguera, for welcoming me into the OKT9 transferrin receptor complex studies that served as a much-needed boost during a low and difficult period of my thesis work. For his ecstatic scientific explanations, advice and confidence in my abilities.

At UCLA, I would like to start by thanking Dr. Diana Azurdia, for teaching me how to advocate for myself in academia, and for her support and guidance during my leadership of SACNAS at UCLA. Prof. Jorge Torres, for his mentorship and guidance during my rotations and early graduate career.

I would like to thank Dr. Duilio Cascio and Dr. Michael Sawaya, first and foremost, for training me as a crystallographer when I arrived to UCLA with little knowledge of structural biology. Dr. Micheal Sawaya for his elaborate explanations, patience during data processing and open access policies on all his amazing resources, and for his intellectual contributions to analysis of the single-particle structure of CPEB3.

Dr. Duilio Cascio, for his unwavering support throughout every single era of my graduate career. For his lively demonstrations of diffraction data processing, guidance in seeking training and unprompted reassurances pertaining to progress in my academic career. I am incredibly grateful to have found such an exceptionally supportive and kind mentor.

I would like to thank Mike Collazo for his assistance during crystallization attempts of amyloidotic peptides and for his uplifting discussions on science and life. Wong Hoi Hui, for his guidance, advice and patience while teaching me cryo-EM sample preparation and insertion on the TF20. Dr. Peng Ge, who was the first person to walk me through and teach me how to properly align an electron microscope, for assistance during data collection on the Krios and for his priceless feedback and discussions regarding helical reconstruction. Dr. David Boyer, for his willingness to teach me all aspects of cryo-EM data processing and helical reconstruction and for countless discussions and intellectual

contributions throughout the (many) datasets that led to the eventual determination of the structure of CPEB3 fibrils.

At SLAC, I am incredibly grateful for the opportunities and access the training programs have provided me. I would like to thank Prof. Wah Chiu, for access to his microscopes at SLAC and providing opportunities to present my work. I would like to thank Dr. Corey Hecksel, for teaching me how to FIB mill and for his enthusiasm and for his assistance during single particle and cryo-ET data collection. Dr. Chengsong Zhang, for training me on the cryo-confocal instrument and for countless support and helpful FIB milling discussions during my trips. Dr. Lydia-Marie Joubert, for her support during my monthly trips and her awe-inspiring management and organization of all on-site activities. For including me in iFLM TFS training sessions and providing opportunities for me to present and discuss my work.

I would like to thank Prof. Eva Nogales, for her generous support and guidance during writing and planning of my proposed post doc work, as well as her patience, as I wrapped up my thesis work at UCLA.

Finally, I would like to thank my friends and family for their everlasting love and support. My sisters, Clementina, Jessica, Bianca and my nieces, Meena and Elena for their uplifting daily calls and messages. My parents, Minerva and Federico, who endured and persisted through unimaginable struggles in this country for me to have the opportunities that they were never granted. These achievements are ours.

Chapter 2 in this thesis is presented (Richards, Flores, Milan et al., 2023) with permission from the authors.

Appendix Chapter 1 in this thesis is presented (Ferrero & Flores et al., 2021) with permission from the authors.

Richards, L. S., **Flores, M. D.**, Millán, C., Glynn, C., Zee, C.-T., Sawaya, M. R., Gallagher-Jones, M., Borges, R. J., Usón, I., & Rodriguez, J. A. (2023). Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio & Med Chem Au. https://doi.org/10.1021/ACSBIOMEDCHEMAU.2C00082

Ferrero, S., **Flores, M. D**., Short, C., Vazquez, C. A., Clark, L. E., Ziegenbein, J., Zink, S., Fuentes, D., Payes, C., Batto, M. V., Collazo, M., García, C. C., Abraham, J., Cordo, S. M., Rodriguez, J. A., & Helguera, G. (2021). Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain. Journal of Virology, 95(17), 1868–1888. https://doi.org/10.1128/JVI.01868-20/FORMAT/EPUB

CURRICULUM VITAE

Maria Flores

Education

2017-2023	University of California, Los Angeles
	Doctor of Philosophy in Biochemistry, Molecular and Structural Biology
	Expected
2013-2017	Indiana University Bloomington
	Bachelor of Science in Biotechnology

Honors and Awards

2023	Finalist, HHMI Hanna H. Gray Fellows Competition
2021	Audree V. Fowler Fellow in Protein Science, UCLA
2020	Whitcome Pre-doctoral Fellowship, UCLA
2019-2022	National Science Foundation Graduate Research Fellowship
2019-2020	Ruth L. Kirschstein National Research Service Award GM007185
2017-2023	Eugene V. Cota-Robles Fellowship, UCLA
2017	Competitive Edge Summer Research Program, UCLA

Publications: Original Research

1. Li, Y. L., Zee, C.-T., Lin, J. B., Basile, V. M., Muni, M., **Flores, M. D.**, Munárrizmunárriz, J., Kaner, R. B., Alexandrova, A. N., Houk, K. N., Tolbert, S. H., & Rubin, Y. (2020). Fjord-Edge Graphene Nanoribbons with Site-Specific Nitrogen Substitution. *J. Am. Chem. Soc*, *142*, 18093–18102. https://doi.org/10.1021/jacs.0c07657

2. Ferrero, S.†, **Flores, M. D**.†, Short, C., Vazquez, C. A., Clark, L. E., Ziegenbein, J., Zink, S., Fuentes, D., Payes, C., Batto, M. V., Collazo, M., García, C. C., Abraham, J., Cordo, S. M., Rodriguez, J. A., & Helguera, G. (2021). Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain. Journal of Virology, 95(17), 1868–1888. https://doi.org/10.1128/JVI.01868-20/FORMAT/EPUB

3. Richards, L. S.†, **Flores, M. D.**†, Millán, C.†, Glynn, C., Zee, C.-T., Sawaya, M. R., Gallagher-Jones, M., Borges, R. J., Usón, I., & Rodriguez, J. A. (2023). Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio & Med Chem Au. https://doi.org/10.1021/ACSBIOMEDCHEMAU.2C00082

4. **Flores, M. D.,** Sawaya, M. R., Boyer, D. R., Zink, S., Tovmasyan, S., Saucedo, A., Zee, C.-T., Cardenas, J., Fioriti, L., & Rodriguez, J. A. (2022). Structure of a reversible amyloid fibril formed by the CPEB3 prion-like domain reveals a core sequence involved in translational regulation. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2022.12.07.519389

5. Richards, L. S., **Flores, M. D.**, Zink, S., Schibrowsky, N. A., Sawaya, M. R., & Rodriguez, J. A. (2023). Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at its Core. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2023.02.08.527771

†*denotes equal contribution*

Chapter 1

PROLOGUE

Prions

Prions, defined as proteinaceous infectious particles[1], are proteins that can form stable, long-lived pathogenic or functional assemblies, and present in organisms throughout the tree of life[2]. Prion assemblies are unique in their ability to propagate by self-templating, recruiting soluble proteins into self-assembly. Aggregation prone prion-like proteins are a liability in the human brain and the major cause of progressive neurodegenerative diseases[3]. By nature of their structural transitions and robust conformations, functional prions may appear as an anomaly in the delicate framework of neurons. Thus, there remains a need to explore and identify structural and molecular hallmarks of functional prions as well as their regulation in the mammalian nervous system.

The prion state enables diverse functional roles

Protein folding is the pathway connecting genotypes and phenotypes[4], oftentimes altering functional traits and evolutionary trajectories[2,5]. A protein equipped with a multiplicity of structural conformations is likely to conduct a variety of functional tasks providing a means of adaptation in fluctuating environments[2]. Trailblazing work from Susan Lindquist's group showed that the exceptional structural plasticity of prions is essential for various genetic pathways[6–8]. Initial reports demonstrated that these non-mendelian mechanisms of inheritance are extensively exploited in fungal organisms[9]. Transmission of structures relies on their prion-like domain, enriched in glutamine and asparagine residues, prone to self-assembly and crucial for propagation[2].

Prions as a putative substrate for memory

The synaptic tagging and capture hypothesis proposes that synapses that undergo plasticity can be tagged with molecular markers that enable them to capture and retain specific proteins[10]. When a synapse is activated by a learning event, it triggers the synthesis of new proteins required for long-term changes in synaptic strength. CPEBs bind to mRNA that encode these proteins and promote cytoplasmic polyadenylation, enhancing translational rates[10,11]. The increase in local protein synthesis enables molecular markers to tag a synapse, priming it for synaptic changes. CPEBs coordinate mRNA translation, allowing for the capture and retainment of proteins necessary for synaptic strengthening and formation of new synaptic connections.

The nature of the synaptic mark calls for a protein capable of activation via a conformational switch, a spatially controlled functionality and efficient signal propagation[10,12]. The molecular basis of memory formation thus requires a molecular state that relies solely on self-propagation as opposed to the traditionally documented roles of nucleic acids. This mechanism's endurance must be robust enough to withstand molecular turnover and sporadic environments within the neuronal synapses[10]. Eric Kandel’s group began providing evidence for prion-based mechanisms operating in neurons in *Aplysia*[13], *Drosophila*[14] and, mammals[15–17], all of which were ideal candidates as components of the synaptic tag[12]. These reports showcased a group of proteins belonging to a family of RNA-binding, cytoplasmic polyadenylation element binding proteins (CPEB), that control polyadenylation-induced translation and are highly conserved across species[11]. Early biochemical and functional studies of the neuronal isoform of CPEB in aplysia revealed an n-terminal domain sequence enriched in polar

residues, with striking similarity to those reported in yeast prion domains[13,18]. In addition, these experiments revealed that CPEB allows for localized translation at the synapses, utilizing its prion-like domain to convert between a translational repressor to a translational de-repressor [13]. Fluorescence experiments captured bright, coalesced aggregates, puncta, in neurons expressing CPEB-GFP[13].

The mammalian prion, CPEB3

A mouse isoform of CPEB, CPEB3, has also been shown to be a translational regulator that is implicated in long-term memory maintenance[15–17,19]. However, CPEB3's n-terminal prion-like region displays a greater degree of sequence complexity than that of its invertebrate orthologues. Previous work has established a tri-partite organization of the prion-like domains, splitting the 449 span of polar and hydrophobic residues into three main domains[16]. Here, two prion-like domains (1 and 2) flank a middle, regulatory domain, which is shown to rely on the actin cytoskeleton for proper aggregation and localization[16]. Experiments in primary neurons demonstrate the essential nature of the first 217 residues, prion-like domain 1 (PRD1), in localization and stimulation based translational activation[16]. These studies are complemented by observations that knockout CPEB3 mouse lines experience learning recovery after introduction of the full-length protein, but not with a truncated construct missing PRD1[17]. Thus, the ability of CPEB3 to regulate translation and consolidate memory formations, appears to rely entirely on PRD1. These observations focus primarily on CPEB3 and its ability to self-assemble and regulate translation as a protein. Further work has also demonstrated a

self-cleaving ribozyme from CPEB3 gene is capable of regulating transcriptional and translational processes.

Amyloid as a structural scaffold

While their functions vary across organisms, the majority of prion biophysical features can be consolidated to their ability to assemble into amyloid and use that state as a structural scaffold[3,20,21]. Amyloid refers to non-branched fibrillar assemblies that are aggregation prone, protease-resistant, and consist of a β-sheet rich structures[3]. In the amyloid fold, β-sheet residues interdigitate and allow for compact side by side interactions, typically forming tight and dry interfaces. Sheets are the one-dimensional layer of the assembly that stack on top of one another, stacked 4.8 Å, spaced ~ 8-12 Å apart. Layers run perpendicular to the fibril axis and are reinforced by the backbone hydrogen bonding with sheet layers above and below. At the time of their discovery and throughout many decades following, their biophysical characteristics and presence in disease and function became more clear, yet, the atomic structures of cross-β cores of amyloid fibrils remained out of reach.

The use of scattering-based methods to reveal amyloid structures

Structural biology has benefitted greatly from the advent of technological advancements from x-ray beamlines at synchrotrons to the rapid maturation of cryogenic electron microscopy (cryo-EM). However, for the first several decades of EM, amyloid protein depositions remained underrepresented in the protein data bank for many reasons. As previously described, the two-dimensional layers of amyloids stack perpendicular to the

fibril axis. To accommodate the dense packing of strands, layers twist slightly, favoring formation of elongated, fibrillar structures[3,20]. These constraints limit the ability of amyloid assemblies to form crystals and if formed, significantly deter crystal size to a few microns or less. David Eisenberg's group pioneered alternative approaches to interrogate these structures with the discovery that much shorter peptidic segments derived from fibril forming proteins showed identical cross-β diffraction patterns. Using a small 7 residue segment from the yeast prion, Sup35, crystals amenable to the newly available microfocus beamlines revealed the first atomic structure of the amyloid scaffold, an 'amyloid spine'[22]. These attempts were quickly followed by a collection of structures derived from various fibril-forming proteins[23], mapping the atomic structures of cross-β spines across various amyloids. Despite these breakthroughs, the majority of amyloid segments investigated formed crystals too small for x-ray crystallography. It was not until the development of continuous rotation micro-electron diffraction (Micro-ED), that micron sized crystals would be within reach.

A brief history of cryogenic electron microscopy

Electron microscopy's (EM) developments and feats stalled behind x-ray based methods for several decades due to effects of the high vacuum chamber and radiation damage on specimen but mostly, extremely noisy images[24]. In brief, when electrons interact with a sample, two types of interactions can occur. Interaction of scattered electrons with specimen without energy loss will subsequently contribute to signal, called elastic scattering events, but deposition of some of their energy on the specimen will cause an inelastic scattering event. Biological samples in particular are susceptible

to inelastic scattering events due to the large inelastic cross section of light atoms, and the inherently weak bonds that are easily broken by the high energy electrons, resulting in irreversible damage to specimen[25]. Thus, imaging biological specimen with electrons requires intentional reduction in flux ('dose') to avoid irreversible damage, producing images with low signal-to-noise ratio (SNR). Eventually, Kenneth Taylor and Bob Glaeser demonstrated that catalase crystals could be preserved with frozen hydrated approaches[26], achieving protein crystallography with electrons. Then, Jacques Dubochet reported the revolutionary application of vitrification to create thin films of vitreous ice, providing protection of biological specimen from the brutal electron beam[27], bearing the field of cryogenic electron microscopy (cryo-EM). However, the technique still required limited flux and produced blurry images. The initial application of cryo-EM began with Richard Henderson's structural determination of bacteriorhodopsin from 2D crystals[28] and slowly expanded to large, symmetrical and homogeneous particles, such as filaments[29] and viruses[30]. It took seminal work in hardware improvements, like the switch from charged-coupled devices (CCD) to direct electron detectors (DED)[31], to boost SNR and readout time, and revolutionize cryo-EM data collection[24]. In parallel, improvements in processing of images acquired with DED, like beam-induced motion correction[32] and reliable computational averaging of heterogenous particles[33], that finally led to the resolution revolution of single particle cryo-EM.

Amyloid and cryo-EM

The determination of amyloid structures benefitted greatly from cryo-EM strides, starting with the development of continuous rotation Micro-ED[34], or 3D ED. In fact, two core

segments from alpha-synuclein provided the first novel structure from MicroED data to near atomic resolution[35]. This study paved the way for the determination of what are now dozens of peptidic amyloid segments from tiny micron sized crystals[20]. Then, with the enhancement of EM hardware and growing accessibility of helical reconstruction software[24,33], a full-length tau fibril derived from diseased brain tissue was reported[36]. Of note, nearly two decades after Eric Kandel, Susan Lindquist and Kausik Si reported and characterized the prion-like domain and its role in memory storage in Aplysia CPEB[13], the physical entity of the putative memory conduit was isolated from *Drosophila* brain[37]. By using over four million *Drosophila* brains, single particle cryo-EM and helical reconstruction, the Orb2 prion-like domain conclusively showed that CPEB utilizes a canonical amyloid scaffold to regulate translation. The determination of high-resolution structures of amyloid fibrils is mainstream[38].

Cryo-electron tomography

While single-particle cryo-EM allows proteins or complexes to be studied in solution, they remain in isolation away from their native functional environments. Owing to the same technological advances that the resolution revolution of single-particle cryo-EM took off, so too did the field of cryo-electron tomography (cryo-ET)[39]. Tomography allows for specimen to be imaged through a collection of projections as the sample is tilted from different angles[40]. The resulting images are then aligned and back-projected to create a 3-dimensional volume, a tomogram[40]. Therefore, the triumph of cryo-ET comes when whole cells are vitrified in a near native state, free of chemical fixatives and artifacts from dehydration, revealing 3-dimensional information of cellular structures at

nanometer scale resolution, without the need for averaging[41,42]. However, seminal work and achievements showcasing the ability of taking subvolumes of the protein of interest and aligning and averaging, or subtomogram averaging[43], have made molecular resolution accessible to cryo-ET[44]. Cryo-ET is thus capable of imaging whole cells and determining high-resolution structures of proteins in their native environments. Cryo-ET is subject to the major limitations shared among all cryo-EM techniques[41]. First, in order to minimize radiation damage to specimen, each tilt image must be acquired at much lower electron dose (0.3 – 1 $e^-/Å^2$) than those of single particle cryo-EM. This exaggerates the low SNR described before but has been circumvented with the development of phase plates[45,46] that enhance low-spatial frequency information and thus, increase contrast. The next obvious limitation relates to sample thickness, where increased inelastic scattering events of electrons are seen as samples near > ~300 nm, resulting in limited resolution and restriction of samples amenable to cryo-ET. Zero loss energy filtering aids in removal of inelastically scattered electrons[47], further improving SNR. Lastly, resulting tomograms often reveal the crowded and chaotic insides of cells, with no true grasp of what molecules are being imaged. Prior knowledge of subcellular structures serves as the only means to help identify your features of interest, making identification and targeting a major bottleneck of cellular cryo-ET.

Correlative cryo-focused ion beam milling

To overcome the inherent limitations in sample thickness for cryo-ET, significant specimen thinning is required. Michael Marko first illustrated the use of focused-ion

beam (FIB) milling[48], demonstrating that gallium ions could be used to precisely thin frozen hydrated bacterial cells. Here, all but a thin region of the cell was ablated by high energy ions while monitoring of the process was enabled by a scanning electron microscope (SEM). The use of the dual beam FIB-SEM instrument results in a slice of a cell, a lamella, that is amenable to cryo-ET data collection. Further developments and refinements to cryo-FIB milling as a technique for thinning of biological specimen were led by Wolfgang Baumeister's group[49,50], eventually reaching the holy-grail of eukaryotic structural cell-biology, a well-preserved nuclear periphery[51].

As the feat of generating lamellae from cells was accomplished, the daunting task of labeling and targeting of specific molecules within a cell had already begun[52]. Correlative light and electron microscopy (CLEM) remains the most viable approach to date, where proteins of interest are tagged with a fluorescent marker or probed with dyes. Initially, the use of cryo-confocal instrumentation was described, where fluorescence data is acquired before cryo-ET[53], and correlation of features of interest (FOIs) are then aligned or traced using overlay techniques. This method proves problematic with fragile and precious lamellae, as cryo-confocal instruments do not operate under high vacuum and expose the specimen to ice contamination and potential warping from temperature changes associated with imaging and handling, further hindering correlative efforts down the line[52,54]. In addition, the objective cannot touch the vitrified sample, and resolution limitations due to long-working distances arise. An alternative approach that has very recently become commercially available is to have an integrated fluorescence objective in the FIB-SEM instrument. Even then, the low angle FIB requires precise targeting of a molecule in the z-direction, to ~500 nM

accuracy, this remains a significant challenge. As briefly introduced, the resulting workflow for correlative cryo-FIB milling currently requires specimen to be handled and transferred through various instruments while requiring the user to operate various softwares for an efficient session and proper correlation[55], though, the resulting specimen have provided unprecedented views of cellular environments. The developmental stage of correlative and thus targeted lamella tomography continues at present, but of note, great interest in pathogenic amyloid structures has allowed for early efforts to explore toxic inclusions in their native environments[56,57].

Overview

Chapter 2 of this thesis begins with a classical approach of interrogating amyloids in CPEB3, through peptidic segments. In parallel, the work demonstrates the need and development of novel *ab initio* fragment-based phasing approaches for MicroED data. Chapter 3 then dives into the use of single particle cryo-EM and helical reconstruction to determine the structure of reversible CPEB3 fibrils and validate their functional significance in translational regulation. Chapter 3 concludes with the exploration of neurons expressing CPEB3-GFP, using cutting edge cryo-ET techniques to search for the shadows of the functional prion-like protein in its native environment.

In addition, an appendix describes my application of a novel approach to neutralizing new world hemorrhagic fever viruses and the characterization of broadly neutralizing agents.

References

1. Prusiner, S. B. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science (80-.).* **216**, 136–144 (1982).
2. Shorter, J. & Lindquist, S. Prions as adaptive conduits of memory and inheritance. *Nature Reviews Genetics* vol. 6 435–450 (2005).
3. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
4. James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution – a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
5. Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
6. Sondheimer, N. & Lindquist, S. Rnq1: An Epigenetic Modifier of Protein Function in Yeast. *Mol. Cell* **5**, 163–172 (2000).
7. Wickner, R. B. [URE3] as an Altered URE2 Protein: Evidence for a Prion Analog in Saccharomyces cerevisiae. *Science (80-.).* **264**, 566–569 (1994).
8. Lindquist, S. Mad cows meet psi-chotic yeast: The expansion of the prion hypothesis. *Cell* **89**, 495–498 (1997).
9. Wickner, R. B. Yeast and Fungal Prions. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).
10. Kandel, E. R. The molecular biology of memory storage: A dialogue between genes and synapses. *Science (80-.).* **294**, 1030–1038 (2001).
11. Richter, J. D. CPEB: a life in translation. *Trends Biochem. Sci.* **32**, 279–285

(2007).

12. Rayman, J. B. & Kandel, E. R. Functional Prions in the Brain. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
13. Si, K., Lindquist, S. & Kandel, E. R. A Neuronal Isoform of the Aplysia CPEB Has Prion-Like Properties. *Cell* **115**, 879–891 (2003).
14. Keleman, K., Krüttner, S., Alenius, M. & Dickson, B. J. Function of the Drosophila CPEB protein Orb2 in long-term courtship memory. (2007) doi:10.1038/nn1996.
15. Pavlopoulos, E. *et al.* Neuralized1 activates CPEB3: A function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell* **147**, 1369–1383 (2011).
16. Stephan, J. S. *et al.* The CPEB3 Protein Is a Functional Prion that Interacts with the Actin Cytoskeleton. *Cell Rep.* **11**, 1772–1785 (2015).
17. Fioriti, L. *et al.* The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. *Neuron* **86**, 1433–1448 (2015).
18. Raveendra, B. L. *et al.* Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. *Nat. Struct. Mol. Biol.* **20**, (2013).
19. Drisaldi, B. *et al.* SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep.* **11**, 1694–1702 (2015).
20. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R. & Eisenberg, D. S. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873 (2021).
21. Prusiner, S. B. *et al.* Scrapie prions aggregate to form amyloid-like birefringent rods. *Cell* **35**, 349–358 (1983).

22. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nat. 2005 4357043* **435**, 773–778 (2005).

23. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nat. 2006 4477143* **447**, 453–457 (2007).

24. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **58**, 677 (2015).

25. Saha, A., Nia, S. S. & Rodríguez, J. A. Electron Diffraction of 3D Molecular Crystals. *Chem. Rev.* **122**, 13883–13914 (2022).

26. Taylor, K. A. & Glaeser, R. M. Electron diffraction of frozen, hydrated protein crystals. *Science* **186**, 1036–1037 (1974).

27. Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A. & Homo, J. -C. Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* **128**, 219–237 (1982).

28. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).

29. Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. Nicotinic acetylcholine receptor at 4.6 Å resolution: Transverse tunnels in the channel wall. *J. Mol. Biol.* **288**, 765–786 (1999).

30. Liu, H. *et al.* Atomic structure of human adenovirus by cryoEM reveals interactions among protein networks. *Science* **329**, 1038 (2010).

31. McMullan, G., Faruqi, A. R. & Henderson, R. Direct Electron Detectors. *Methods Enzymol.* **579**, 1–17 (2016).

32. Li, X. *et al.* Electron counting and beam-induced motion correction enable near

atomic resolution single particle cryoEM. *Nat. Methods* **10**, 584 (2013).

33. Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
34. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *Elife* **2013**, (2013).
35. Rodriguez, J. A. *et al.* Structure of the toxic core of α-synuclein from invisible crystals HHS Public Access. *Nature* **525**, 486–490 (2015).
36. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nat. 2017 5477662* **547**, 185–190 (2017).
37. Hervas, R. *et al.* Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science (80-.).* **367**, 1230–1234 (2020).
38. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *bioRxiv* 1–9 (2019) doi:10.1101/823310.
39. Gan, L. & Jensen, G. J. Electron tomography of cells. *Q. Rev. Biophys.* **45**, 27–56 (2012).
40. Tocheva, E. I., Li, Z. & Jensen, G. J. Electron Cryotomography. *Cold Spring Harb. Perspect. Biol.* **2**, (2010).
41. Wan, W. & Briggs, J. A. G. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol.* **579**, 329–367 (2016).
42. Wagner, J., Schaffer, M. & Fernández-Busnadiego, R. Cryo-electron tomography-the cell biology that came in from the cold. *FEBS Lett.* **591**, 2520–2533 (2017).
43. Briggs, J. A. Structural biology in situ-the potential of subtomogram averaging. *Curr. Opin. Struct. Biol.* **23**, 261–267 (2013).

44. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods 2019 1611* **16**, 1146–1152 (2019).

45. Danev, R., Kanamaru, S., Marko, M. & Nagayama, K. Zernike phase contrast cryo-electron tomography. *J. Struct. Biol.* **171**, 174–181 (2010).

46. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15635–15640 (2014).

47. Grimm, R. *et al.* Energy filtered electron tomography of ice-embedded actin and vesicles. *Biophys. J.* **72**, 482 (1997).

48. Marko, M., Hsieh, C., Schalek, R., Frank, J. & Mannella, C. Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy. *Nat. Methods 2007 43* **4**, 215–217 (2007).

49. Rigort, A. *et al.* Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4449–4454 (2012).

50. Schaffer, M. *et al.* Cryo-focused Ion Beam Sample Preparation for Imaging Vitreous Cells by Cryo-electron Tomography. *BIO-PROTOCOL* **5**, (2015).

51. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science (80-.).* **351**, 969–972 (2016).

52. Arnold, J. *et al.* Site-Specific Cryo-focused Ion Beam Sample Preparation Guided by 3D Correlative Microscopy. *Biophys. J.* **110**, 860–869 (2016).

53. Schorb, M. & Briggs, J. A. G. Correlated cryo-fluorescence and cryo-electron microscopy with high spatial precision and improved sensitivity. *Ultramicroscopy* **143**, 24–32 (2014).

54. Klumpe, S. *et al.* A Modular Platform for Automated Cryo-FIB Workflows. *Elife* **10**, (2021).

55. Wagner, F. R. *et al.* Preparing samples from whole cells using focused-ion-beam milling for cryo-electron tomography. *Nat. Protoc. 2020 156* **15**, 2041–2070 (2020).

56. Bäuerlein, F. J. B. *et al.* In Situ Architecture and Cellular Interactions of PolyQ Inclusions. *Cell* **171**, 179-187.e10 (2017).

57. Guo, Q. *et al.* In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. *Cell* **172**, 696-705.e12 (2018).

Chapter 2

This chapter is based on the published article “Fragment-based ab initio phasing of peptidic nanocrystals by MicroED.”

Summary

The work described here highlights the development of fragment-based phasing of MicroED data. The approach that was applied to a short peptidic segment derived from CPEB3’s prion-like domain and other amyloid segments. The application of *ab initio* phasing using ARCIMBOLDO_BORGES revealed an alternative conformation of the CPEB3 peptide, QIGLAQTQ. These data showcase the first high-resolution structures of CPEB3 and are a proof of principle that small segments from this prion-like domain are capable of forming amyloid, a finding not known at the time. The MicroED polymorphs illustrate the prion-like, polymorphic nature of the sequence, as the same residues were capable of forming two distinct amyloid conformations. These results were the primary motivation and groundwork for experiments summarized in Chapter 3.

Richards, L. S., **Flores, M. D.**, Millán, C., Glynn, C., Zee, C.-T., Sawaya, M. R., Gallagher-Jones, M., Borges, R. J., Usón, I., & Rodriguez, J. A. (2023). Fragment-Based Ab Initio Phasing of Peptidic Nanocrystals by MicroED. ACS Bio & Med Chem Au. https://doi.org/10.1021/ACSBIOMEDCHEMAU.2C00082

ACS BIO & MED CHEM Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/biomedchemau

Article

Fragment-Based *Ab Initio* Phasing of Peptidic Nanocrystals by MicroED

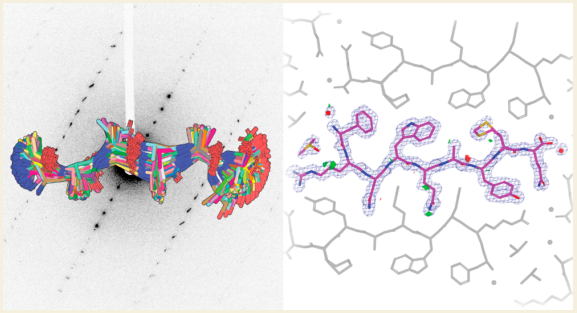
Logan S. Richards,[#] Maria D. Flores,[#] Claudia Millán,[#] Calina Glynn, Chih-Te Zee, Michael R. Sawaya, Marcus Gallagher-Jones, Rafael J. Borges, Isabel Usón,[*] and Jose A. Rodriguez[*]

Cite This: https://doi.org/10.1021/acsbiomedchemau.2c00082

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electron diffraction (MicroED/3DED) can render the three-dimensional atomic structures of molecules from previously unamenable samples. The approach has been particularly transformative for peptidic structures, where MicroED has revealed novel structures of naturally occurring peptides, synthetic protein fragments, and peptide-based natural products. Despite its transformative potential, MicroED is beholden to the crystallographic phase problem, which challenges its *de novo* determination of structures. ARCIMBOLDO, an automated, fragment-based approach to structure determination, eliminates the need for atomic resolution, instead enforcing stereochemical constraints through libraries of small model fragments, and discerning congruent motifs in solution space to ensure validation. This approach expands the reach of MicroED to presently inaccessible peptide structures including fragments of human amyloids, and yeast and mammalian prions. For electron diffraction, fragment-based phasing portends a more general phasing solution with limited model bias for a wider set of chemical structures.



KEYWORDS: *fragment-based phasing, ab initio, MicroED, nanocrystal, peptide, cryo-EM, ARCIMBOLDO*

INTRODUCTION

Crystallography has played a momentous role in our understanding of peptidic structures.[1] Microcrystal electron diffraction (MicroED) is expanding its scope by delivering atomic structures from peptide crystals less than a micrometer in thickness.[2−4] Electron diffraction leverages the strong interaction of electrons with matter, capturing diffraction signal that would be missed by conventional X-ray crystallography.[5] Some molecules of high biological or chemical importance are only known to grow nanocrystals, demanding structural methods of extreme sensitivity, as seen in the amyloid peptide structures of the toxic core of the Parkinson's-associated protein α-synuclein[3] or the ultrahigh-resolution structure of a prion protofibril.[4] Likewise, the technique has determined structures of complex bioderived or post-translationally modified peptides such as an amyloid-β core with a racemized residue,[6] the cyclic peptide antibiotic thiostrepton[7] and a synthetic tetrapeptide natural product analogue.[8]

Determination of MicroED structures presently follows one of two routes: *ab initio* phasing through direct methods[9] if data resolution is atomic, or molecular replacement (MR)[10] when a highly similar structure is known.[3,11−13] MR is challenged by unknown peptide structures that contain uncharacterized backbone geometries or a substantial fraction of unnatural amino acids. Without atomic resolution data, novel phasing solutions are still needed for MicroED targets of uncertain geometry, identity, or chemical connectivity.

Fragment-based phasing (FBP)[14] yields accurate solutions relying on the computational search for defined subsets of a target structure to obviate the need for atomic resolution data. Fragments are located by likelihood-based molecular replacement[15] and expanded through density modification and map interpretation.[16] The ARCIMBOLDO programs substitute the atomicity constraint underlying direct methods with stereochemical constraints.[17] For a structure containing defined fragments of constant geometry a single model fragment is appropriate, and model alpha-helices have been particularly successful.[18] General cases require joint evaluation of libraries of fragments, representing variations of a structural hypothesis. Relying on secondary and tertiary structure fragments extracted from the Protein Data Bank (PDB)[19] or from distant homologues,[20] ARCIMBOLDO_BORGES has been

Received: December 15, 2022
Revised: January 25, 2023
Accepted: January 30, 2023

Table 1

	GSTVYAPFT (7N2I)	QIGLAQTQ plate polymorph (7N2F)	QIGLAQTQ needle polymorph (7N2G)	NYNNYQ (7N2K)	QYNNENNFV (7N2J)	FRNWQAYMQ (7N2D)
			Data Collection and Processing			
resolution (A)	7.67−1.40(1.40)	7.09−1.2(1.2)	7.61−1.20(1.20)	8.91−1.30(1.30)	7.42−1.5(1.5)	19.39−1.50(1.50)
no. crystals	3	3	1	4	6	4
electron dose ($e^-/Å^2$)	<5	<5	<5	<5	<5	<5
space group	*C*2	$P2_1$	$P2_12_12_1$	$P3_1$	*P*1	*C*2
a, *b*, *c* (Å)	58.4, 4.73, 19.63	4.83, 16.29, 29.02	4.82, 20.48, 45.61	27.2, 27.2, 4.83	4.87, 10.06, 30.66	43.12, 4.84, 34.9
α, *β*, *γ* (degrees)	90.00, 105.01, 90.00	90.00, 94.61, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 120.00	94.85, 90.26, 99.98	90, 115.934, 90
no. total reflections	4159	6437	3963	6870	4236	4017
no. unique reflections	985(106)	1129 (105)	1357 (123)	969(84)	730(74)	961(70)
R_{merge}	15.9	20.0	12.8	21.2	18.5	20.2
CC1/2	98.7 (89.5)	98.1 (97.6)	98.0 (53.3)	98.5 (60.4)	98.8 (81.3)	97.7 (78.0)
$\langle I/\sigma I \rangle$	5.16	8.31	4.79	6.16	7.01	4.32
completeness	80.3 (85.5)	78.2 (78.52)	81.6 (83.1)	97.68 (87.9)	97.8 (82.7)	75.8 (51.2)
multiplicity	4.22	5.70	2.92	7.09	5.71	4.18
phasing success using alternative libraries[a]	RL: 45% correct solutions	CL: 16% correct solutions	CL: 10% correct solutions	CL:0.4% correct solutions	CL: 3% correct solutions	RL: 1% correct solutions
	GL: 2% correct solutions	GL: 2% correct solutions	GL: 5% correct solutions	GL: 2% correct solutions	GL: 4% correct solutions	GL: 3% correct solutions
residues placed	6	7	6	6	7	8
fragments placed	1	1	1	1	1	1
LLG	47.10	175.90	62.80	27.43	60.30	46.9
TFZ	4.60	4.80	6.90	5.30	7.90	5.3
final CC (%)	21.85	57.14	30.54	20.21	34.12	30.29
			Refinement			
R_{work} (%)	19.36 (26.52)	19.8 (21.2)	19.24 (32.32)	16.14 (22.74)	17.44 (22.90)	21.20(38.46)
R_{free} (%)	19.23 (45.49)	22.9 (18.1)	23.65 (31.41)	18.54 (30.74)	22.65 (18.79)	23.87(25.69
			RSCC			
no. waters ligand atoms	1	0	1	3	1	2
average B-factor:						
protein	8.78	3.22	3.85	7.05	4.68	8.07
water	13.05		10.66	20.01	6.15	12.7
ligand						18.35
r.m.s.d. bonds (A)	0.033	0.027	0.008	0.099	0.012	0.015
r.m.s.d. angles (degrees)	2.13	2.69	0.91	0.55	1.18	1.41
Ramachandran (ouliers,favored) (%)	0.00	0.00	0.00	0.00	0.00	0.00
Clashscore	0.00	0.00	0.00	0.00	6.76	0

[a]GL: general library CL: custom library RL: modeled library. Correct solutions are defined as those with an initial wMPE below 60°.

broadly used in phasing protein structures determined by X-ray crystallography. ARCIMBOLDO has also been used on MicroED data, to phase a 1.6 Å structure of Proteinase K from distant homologues.[21] Fragments placed accurately contribute to solutions despite accounting for a few percent of the scattering atoms in a structure.[22] Since any experimental or calculated fragment may be used as input, fragment-based phasing could prove powerful for the general determination of peptidic or other chemical structures by electron diffraction.

Here, we demonstrate fragment-based phasing for *ab initio* structure determination of novel peptide structures from MicroED data in the absence of atomic resolution. Our approach is based on the development of new fragment library methods tailored to sample structural variability, while profiting from the reduced size of active peptide structures to preclude model bias. We validate its success on known and novel structures obtained from nanocrystallites formed by diverse amyloid peptides.

RESULTS

High-Resolution MicroED Data from Peptide Nanocrystals

The limited crystal size and directional growth exhibited by some peptides of high biological or chemical interest renders electron diffraction a necessary choice for structure determination. However, faced with a growing number of MicroED data sets from peptide crystals, for which direct methods and molecular replacement solutions were unavailable, we set out to develop dedicated fragment phasing approaches for this set of substrates.

Nanocrystals from each of five peptide segments summarized in Table 1 were preserved on grids in a frozen-hydrated state; crystals of each were visually identified and diffracted as previously described.[3] Ideal candidates for MicroED yielded better than 2 Å diffraction. Diffraction data from several crystals were merged to improve completeness. Structural determination via direct methods with SHELXD[9] succeeded for peptides whose crystals diffracted to atomic resolution: a synthetic mammalian prion segment (QYNNENNFV) [**1**] and

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

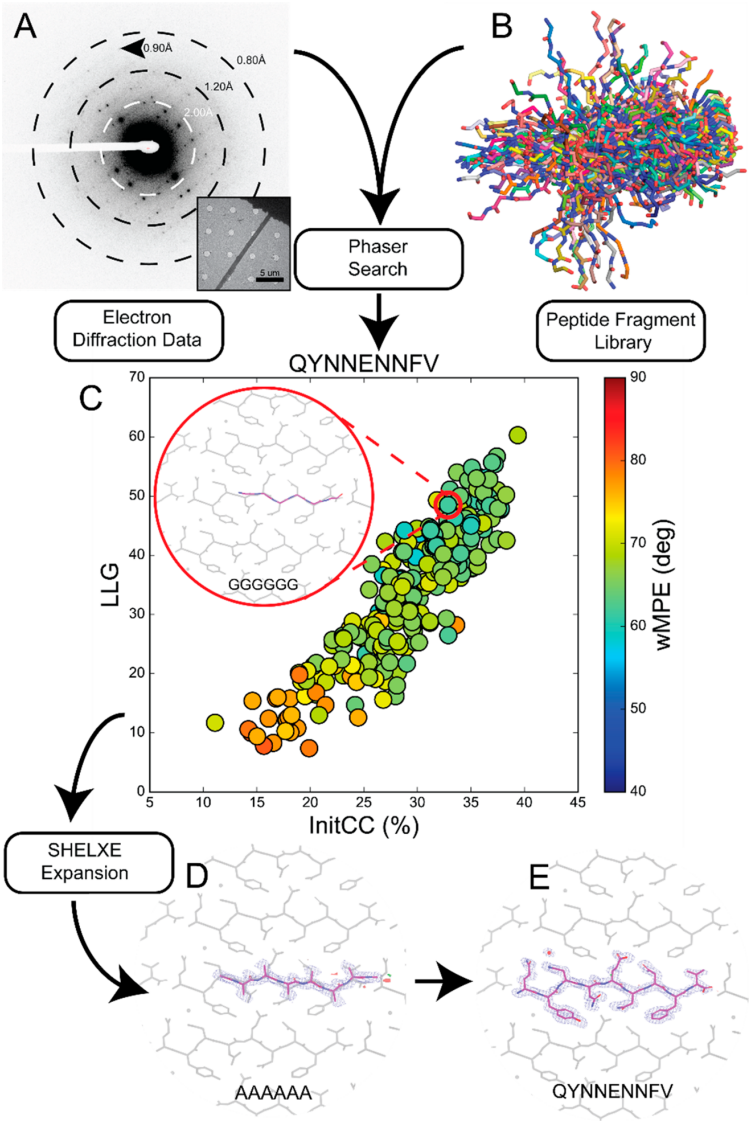


Figure 1. Workflow for using peptide fragments for phasing MicroED data. A) Electron diffraction pattern reaching atomic resolution for QYNNENNFV. Rings designate resolution ranges while arrowhead designates highest resolution spot. B) All fragments comprising the polyglycine hexapeptide library aligned in pymol. C) LLG vs InitCC plot for the fragments screened in ARCIMBOLDO-BORGES. Color indicates the wMPE of the fragment relative to the phases calculated from the final structure. Inset shows the fragment chosen for SHELXE expansion overlaid on the final structure. D) Output solution from ARCIMBOLDO-BORGES following SHELXE expansion is shown overlaid on the final structure. Maps are shown after one round of refinement in Phenix. E) Final structure and potential map for the QYNNENNFV peptide with symmetry related chains shown in gray.

a sequence variant of a repeat segment of the yeast prion New1p (NYNNYQ) [**2**] as well as a plate polymorph of the functional mammalian prion, CPEB3 (QIGLAQTQ) [**3**]. Direct methods solutions were unattainable for the needle polymorph of [**3**], a segment of the human amyloid protein LECT2 (GSTVYAPFT) [**4**], and a segment from the human zinc finger protein (ZFP) 292 (FRNWQAYMQ) [**5**].

Phasing with ARCIMBOLDO Using General Local Fold Libraries

For *ab initio* macromolecular phasing, ARCIMBOLDO_-BORGES exploits fragment libraries representing a common local fold as found in a vast number of PDB structures. Such libraries can be derived from millions of fragments clustered to describe the geometrical variation within the radius of convergence of the method. Fragment superposition allows joint statistical analysis of all phasing attempts as a single experiment. Since typical fragments in macromolecular libraries contain more residues than our peptide structures,[23] we devised dedicated libraries that handled the high abundance of motifs exhibited by short peptides; for example, two short antiparallel beta strands. Weighing overall and local properties when superposing such small models also presented a challenge that was solved experimentally, simulating data

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

from a template and refining the location of all library models against the calculated data.

Single-strand libraries yielded partial solutions for all peptides in Table 1. However, the best solutions identified by this approach did not benefit from the statistical or phase combination algorithms enabled by prior superposition.[24] Instead, optimal solutions were obtained by correct placement of the single most accurate fragments. As peptidic MicroED data sets tend to be small and thus amenable to a large number of calculations from different starting fragments, we decided to exploit competing hypotheses to address the more pressing concern of model bias. With that in mind, we devised libraries encompassing dissimilar, nonsuperimposable models.

Broad, Knowledge-Based Libraries to Phase Nanocrystals

To develop diverse, knowledge-based libraries, we began with ideal cases: atomic resolution data from crystals of two peptide segments that yielded structures by direct methods. The first data set represented five crystals of mammalian prion segment [**1**]. The full 0.9 Å data were included in our set as a gold standard (Figure 1A). When intentionally truncated to a resolution of 1.5 Å, both direct methods and molecular replacement using models of closely related prion sequences[4] were unsuccessful. A library of 249 polyglycine hexapeptides derived from previously determined amyloid peptide structures were used as inputs for ARCIMBOLDO_BORGES (Figure 1B). This library contained models that could be clearly discriminated as potential solutions, scoring above 25% in their initial correlation coefficient[22] (Figure 1C). These same solutions would later be found to exhibit low errors relative to the phases calculated from a final structure (Table S2). The best solution placed six residues and was sufficient to build the remainder of the peptide based on difference density (Table 1, Figure 1D,E).

Data collected from crystals of the New1p segment [**2**] presented an increased challenge. Although microcrystals of [**2**] yielded an X-ray crystallographic structure (Figures S2 and S3B), the same condition also produced nanocrystals requiring MicroED. The 1.1 Å data set obtained combining four crystals of the latter polymorph rendered a direct methods structure different from that originally determined by X-ray diffraction (Figure S3A,B). The data set truncated to 1.3 Å served as a second test for ARCIMBOLDO_BORGES. A library holding polyglycine pentapeptides yielded a single promising solution that could be fully extended and matched the direct methods solution (Table 1, Figure S4).

Structure from a Segment of the Prion Domain of CPEB3 [3]

We next sought to determine novel peptide structures from data that were not suitable for direct methods. Crystallization of a segment from the prion domain of CPEB3 [**3**] produced crystal slurries ideal for MicroED (Figure 2). Screening them in overfocused diffraction mode revealed two distinct morphologies, suggesting the presence of multiple structures (Figure 2A,E). Crystals of plate morphology belonged to $P2_1$, while those of the relatively rare needle morphology presented space group $P2_12_12_1$ (Table 1, Figure 2E).

Merged data from 3 crystals of the plate polymorph were phased by direct methods. Its structure, an unkinked beta strand, was refined at 1.0 Å (Figure S3C). In contrast, a single crystal of the rare needle polymorph diffracted to ∼1.1 Å resolution, generated a data set that was 81.5% complete at 1.2 Å (Figure 2). Despite its high resolution, neither SHELXD nor

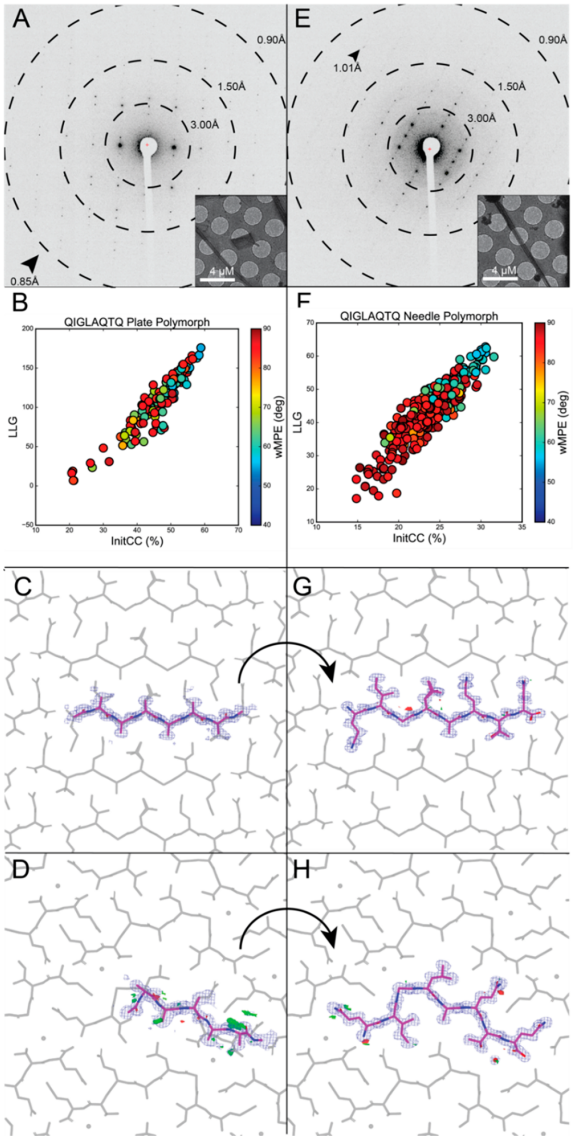


Figure 2. FBP of MicroED data from CPEB3 peptide QIGLAQTQ reveals two polymorphs. A) Electron diffraction pattern reaching atomic resolution for QIGLAQTQ plate morphology. Rings designate resolution ranges while arrowheads designate the highest resolution spot. Overfocused diffraction image of plate morphology crystal. B) Postmortem analysis from ARCIMBOLDO-BORGES plotting LLG vs InitCC ($P2_1$). C) Initial output potential maps following SHELXE expansion by ARCIMBOLDO-BORGES for plate polymorph overlaid on final solution (gray). Buildable density visible on several residues. D) Final density maps of QIGLAQTQ plate polymorph asymmetric unit with symmetry mates shown in gray. E) Diffraction pattern for QIGLAQTQ needle morphology. Rings designate resolution ranges while arrowhead designates highest resolution spot. F) Postmortem analysis from ARCIMBOLDO-BORGES plotting LLG vs IniCC ($P2_12_12_1$). G) Initial output potential maps following SHELXE expansion by ARCIMBOLDO-BORGES for needle polymorph overlaid on final solution (gray). H) Final density maps of QIGLAQTQ needle polymorph asymmetric unit with symmetry mates shown in gray and water molecule displayed as red sphere.

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

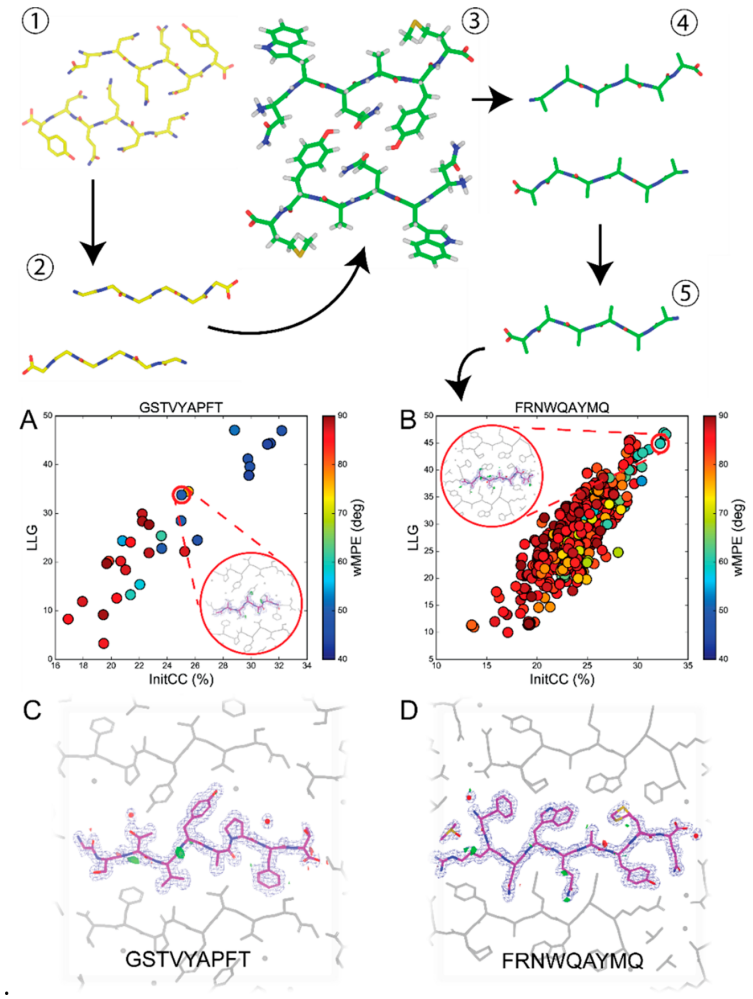


Figure 3. Rosetta library generation and wMPE analysis for GSTVYAPFT and FRNWQAYMQ peptide structures. 1) NNQQNY peptide structure used as a template steric zipper for Rosetta modeling. 2) NNQQNY peptide stripped to glycine residues in preparation for threading. 3) One example of a Rosetta-generated steric zipper structure after threading, repacking, and relaxing in iterative cycles to reach a calculated energy minimum. 4) Rosetta-generated structures are stripped to alanine residues and have hydrogen atoms removed. 5) Individual chains are isolated and are used as the fragment library for phasing with ARCIMBOLDO-BORGES. The LLG vs InitCC plots for A) GSTVYAPFT and B) FRNWQAYMQ are shown below. Inset shows the fragment that was chosen for SHELXE expansion, leading to the correct solution, overlaid with the final structures. C) Final potential maps of GSTVYAPFT asymmetric unit with symmetry mates shown in gray. D) Final potential maps of FRNWQAYMQ asymmetric unit with symmetry mates shown in gray.

molecular replacement with the structure derived from the plate polymorph truncated to poly alanine yielded a solution. Instead, the needle polymorph data set was successfully phased using a 270-fragment polyglycine library of tetrapeptides in ARCIMBOLDO-BORGES (Table 1, Figure S1A,B). An initial solution containing four alanine residues led to a fully refined model (Figure 2G,H). Alternatively, applying the same procedure with an 89-fragment polyglycine library of pentapeptides (Figure S1C,D) to data from the plate polymorph also resulted in a number of possible solutions. In both cases, nonrandom solutions with the highest LLG and Initial CC scores were identified by ARCIMBOLDO-BORGES (Table S2, Figure 2B,F).

Structures Determined by Modeled Fragment Libraries

To overcome the limitation of requiring prior structural knowledge for fragment libraries, we generated atomic models computationally. Such libraries would be ideally suited for determining structures with unanticipated local geometries and could be broadly applicable to a variety of small molecules. We computed fragments using PyRosetta[25,26] starting from a known peptide backbone as a template onto which sequences of interest were threaded and modeled[27] (Figure 3). This scheme was parallelized to generate libraries containing hundreds of fragments, which successfully facilitated FBP of several unknown structures.

Crystals grown from a segment of the LECT2 protein [**4**] diffracted to only 1.4 Å by MicroED, and the data set combined from three such crystals failed to yield solutions from direct methods or molecular replacement with prior

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

fragment libraries or a closely related peptide structure (Table 1). We generated a new hexapeptide poly alanine library using Rosetta in an attempt to approximate the structure of [**4**] in close packing while including the internal proline residue, an uncommon feature in amyloid peptides (Figure S1G,H). The full sequence of the LECT2 protein was used to generate hexapeptide models, that were subsequently threaded pairwise onto the two backbones of the known structure of the peptide NNQQNY.[27] Models were allowed to repack and relax to energy minima in Rosetta. Then, 111 models containing the GSTVYAPFT sequence were truncated to alanine and used in ARCIMBOLDO_BORGES (Figure 3), yielding a discriminated solution (Table S2).

A segment from the human protein ZFP-292 [**5**] presented the most severe challenge to conventional phasing. Due to a high degree of orientation bias of crystals on EM grids, merging data from four crystals achieved an overall completeness of 75.8% at 1.5 Å resolution (Table 1). As in the case of [**4**], this segment could not be phased by direct methods or standard molecular replacement, and no solutions were found when attempting FBP using our fragment libraries from known structures. We again turned to Rosetta in this case to populate a library of fragments that approximated the structure of [**5**], relying only on the 9-residue sequence of the peptide to generate paired hexapeptides for threading. These threaded segments were then evaluated in Rosetta to generate 20 models per pair (Figure 3). This library of 640 models produced in ARCIMBOLDO_BORGES (Figure S1) yielded low wMPE solutions, one of which facilitated a refined solution (Figure 3B,D, Table S2).

OUTLOOK AND CONCLUSIONS

To satisfy the need for new *ab initio* phasing solutions for MicroED, we have developed and deployed new fragment-based phasing strategies using ARCIMBOLDO_BORGES and determined five novel structures. The variation in diffraction quality we observed is representative of the spectrum typically encountered in chemical structure analysis, including examples of relatively low completeness, crystals with low solvent content, and lack of atomic resolution. While the six structures determined here represent a small sampling of the greater universe of peptidic molecules, each of these structures revealed challenges that could be generalized. All analyzed peptides had a high aggregation propensity, contained little to no disordered solvent, and naturally produced nanocrystallites instead of larger crystals. Polymorphism was encountered and in one case revealed differences in atomic structure.

In these cases, and particularly where atomic resolution is not available, information from pre-existing solutions is a lifeline for the phasing process. However, preventing the propagation of errors derived from model bias becomes even more pressing in such cases. Hence, ARCIMBOLDO jointly evaluates large libraries, where competing hypotheses are compared to provide a safeguard against erroneous solutions. The fragments used in our approach were successfully selected by Phaser, based on their LLG, identified with SHELXE CC scores and subsequently expanded by SHELXE into accurate initial solutions. While we observe examples of high phase error models scoring well in preliminary steps, the discrimination of competing potential solutions revealed an unambiguous solution in all cases.

Verification through competition[28] is particularly promising in chemical crystallography when exhaustive searches of solution space are manageable. The libraries we exploited for macromolecular phasing hold variations around a common local fold and yielded correct solutions in all instances. In all cases, single fragment solutions were found to outperform combined solutions. Broadening the base of hypotheses through the use of heterogeneous libraries helped address model bias. All solutions were verified through clear discrimination between conflicting models and similarly high scores for structurally compatible hypotheses. Knowledge-based and modeled libraries rendered higher Z-scores than general libraries, discriminating as illustrated in Figure 2B,F. This makes the present strategy amenable to exploring distinct secondary structure motifs, including primarily α-helical peptides or structures with more than one type of secondary structure.[29,30]

Our trials demonstrate that computed libraries were beneficial when applied to our most challenging cases [**4**–**5**] and could further benefit from new advancements in machine learning. Recently, AlphaFold harnessed the vast structural diversity available in the PDB using deep neural networks to achieve correct prediction of protein folds with unexpectedly high accuracy.[31] However, small, chemically and geometrically diverse structures still require dedicated development. Exploration of the rich structural expanse of chemical space will require methods that accurately select structural fragments while excluding bias artifacts to achieve structural solutions.

Summary

We expand the *ab initio* phasing toolkit for electron diffraction, overcoming the need of atomic resolution diffraction to produce *de novo* solutions. Using ARCIMBOLDO-BORGES and libraries of both known and computed structures, we determine six novel atomic structures of peptide segments. The structures determined using this method are accurate and represent varied geometries and sequences. Model bias is precluded by parallel assessment of a large collection of structural hypotheses providing a baseline. These methods successfully establish a three-dimensional structure from samples that were previously intractable and open a road to structural solutions for small molecules from near atomic-resolution MicroED/3DED data.

EXPERIMENTAL CONSIDERATIONS

Collection of Microfocus X-ray Data and Structure Determination

Crystal clusters of NYNNYQ were grown at room temperature in a 96-well Wizard screen, using a nominally 24.5 mM aqueous solution of the peptide. The crystallization condition chosen for further optimization in 24-well, hanging drop trays consists of 20% 2-methyl-2,4-pentanediol (MPD), buffered by 0.1 M sodium acetate pH 4.5). The peptide crystals were harvested from hanging drops using CryoLoops from Hampton Research with no additional cryoprotectant other than the MPD already present and flash-frozen in liquid nitrogen. Diffraction data sets were collected under cryogenic conditions (100 K) on beamline 24-ID-E at the Advanced Photon Source (APS) equipped with an ADSC Q315 CCD detector, using a 5 μm beam with a wavelength of 0.979 Å. The data were collected via manual vector scanning. 56 diffraction images were collected over three scans from one crystal and one scan from a different crystal. All images have an oscillation range of 5° and were indexed and integrated by XDS.[32] The reflection list outputted by XDS was sorted and merged in XSCALE. SHELXD[33] was able to reach an *ab initio* solution. The atomic coordinates from SHELXD were used to generate a F_{calc} map with SHELXL.[34] An atomic model commensurate with the generated electron density was built in Coot and refined in

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

PHENIX against measured data. The refinement statistics of the final structure are listed in Table S1.

Preparation of Peptide Nanocrystals

Lyophilized, synthetic peptides were purchased from Genscript. Crystals of each peptide were grown as follows: the QYNNENNFV peptide was dissolved in water at 0.88 mM. Crystals were grown using the hanging drop method where 1.5 μL of peptide was added to 1.5 μL of well solution (0.1 M Li_2SO_4, 2.5 M NaCl, 0.1 M NaOAc adjusted to pH 4.5 with acetic acid) in 24-well trays over 500 μL of well solution. QIGLAQTQ was prepared at 64.5 mM in water. Crystals were grown via 24-well hanging-drop vapor diffusion at 27 mM in 14% polyethylene glycol (PEG) 20,000, 100 mM MES, and 3% DMSO. The crystal slurry was briefly sonicated and 0.5 μL was used to seed a new 3 μL drop and repeated three times. FRNWQAYMQ peptide from ZFP-292 was prepared by dissolving at 1.61 mM in water and 3% DMSO. Crystals were grown in batch containing 35% MPD, 100 mM MES, and 200 mM Li_2SO_4, in 1:1 ratio of peptide to buffer. GSTVYAPFT peptide was prepared at 21.3 mM concentration dissolved in water. Crystals were grown in a 10 μL batch in 0.7 M sodium formate with 100 mM sodium acetate pH 4.6. All crystals appeared within 24 h and were identified by light microscopy, and subsequently, EM. Batch crystals of NYNNYQ were prepared by dissolving the peptide in water, resulting in a 24.5 mM solution. An equal volume of the crystallization reagent (20% MPD, buffered by 0.1 M of sodium acetate to pH 4.5) was added to the peptide solution. The solution was then seeded with crushed crystals grown in hanging drop experiments described in the microfocus X-ray data collection section above.

Collection of MicroED Data

For GSTVYAPFT and QYNNENNFV, 2 μL of crystal slurry was applied to each side of a glow-discharged holey carbon grid (Quantifoil, R 1/4 300 mesh Cu, Electron Microscopy Sciences) followed by plunging into liquid ethane using an FEI Vitrobot Mark IV set to a blot time of 22 and a blot force of 22 for QYNNENNFV and 24 for GSTVYAPFT. For QIGLAQTQ, 1.8 μL of crystal slurry was applied to each side of a glow-discharged holey carbon grid (Quantifoil, R 2/1 200 mesh Cu, Electron Microscopy Sciences) and plunge frozen into liquid ethane using a FEI Vitrobot Mark IV using a blot time of 25−30 s and a blot force of 22. For NYNNYQ, 2 μL of crystal slurry was applied to each side of a glow-discharged holey carbon grid (Quantifoil, R 2/1 200 mesh Cu, Electron Microscopy Sciences) and plunge frozen into liquid ethane using a FEI Vitrobot Mark IV, blotting with the force set at 22 for 20−30 s.

For GSTVYAPFT, QYNNENNFV, and FRNWQAYMQ, diffraction patterns and crystal images were collected under cryogenic conditions using a FEI Tecnai F20 operated at 200 keV in diffraction mode. Diffraction patterns were recorded while continuously rotating at 0.3 deg/s (GSTVYAPFT) or 0.25 deg/s (QYNNENNFV, FRNWQAYMQ) using a bottom mount TemCam-F416 CMOS camera (TVIPS). Individual image frames were acquired with 2 s exposures per image for all peptides and 5 s exposures for some FRNWQAYMQ data sets to increase signal. A selected area aperture corresponding to approximately 4 or 6 μm at the sample plane was selected depending on the crystal. For QIGLAQTQ, diffraction patterns were collected under cryogenic conditions using a Thermo-Fisher Talos Arctica electron microscope operating at 200 keV and a Thermo-Fisher CetaD CMOS detector in rolling shutter mode. Individual frames were acquired with 3 s exposures rotating at 0.3 deg/s using selected area apertures of 100, 150, or 200 μm, as needed to match the size of the crystal. A total of 28 movies were collected from QIGLAQTQ crystals of two distinct morphologies, 15 from crystals of needle morphology and 13 from crystals of plate morphology.

MicroED Data Processing

The collected TVIPS movies were converted to individual images in Super Marty View (SMV) format, which are compatible with X-ray data processing software. The diffraction images were indexed and integrated with XDS. The indexing raster size and scan pattern as well as integration in XDS were optimized to minimize contributions by background and intensities from secondary crystal lattices. The reflection outputs from XDS were sorted and merged in XSCALE. For the linear $P2_1$ QIGLAQTQ structure, three partial data sets, containing 252 diffraction images, were merged to produce a final data set with acceptable completeness (~80%) up to 1.00 Å, which was truncated to 1.2 Å for phasing and refinement with ARCIMBOLDO. For the kinked $P2_12_12_1$ QIGLAQTQ structure, one data set, consisting of 82 diffraction images, was sufficient to produce a final merged data set with high completeness up to 1.2 Å for phasing and refinement by ARCIMBOLDO. For the NYNNYQ structure, four partial data sets, composed of 297 diffraction images, were merged to produce a final data set with high completeness up to 1.10 Å, which was truncated to 1.3 Å for phasing and refinement with ARCIMBOLDO. For the GSTVYAPFT structure, three partial data sets, comprised of 327 diffraction images, were merged to produce a final data set with high completeness up to 1.3 Å. For the QYNNENNFV structure, six partial data sets, containing 931 diffraction images, were merged to produce a final data set with high completeness up to 0.9 Å, which was truncated to 1.5 Å for phasing and refinement with ARCIMBOLDO. For FRNWQAYMQ structure, four partial data sets composed of 224 diffraction images were merged to produce a final data set with acceptable completeness out to 1.5 Å. The statistics for each merge are presented in Table 1.

Structure Determination by Direct Methods and Refinement

Electron diffraction data for NYNNYQ, QIGLAQTQ (plate), and QYNNENNFV were of high enough resolution to yield direct methods solutions. SHELXD was able to reach *ab initio* solutions with all three data sets. The atomic coordinates from SHELXD and corresponding reflection files were used as inputs for SHELXL[34] to generate calculated density maps for each solution. Atomic models consistent with the generated density maps were built in Coot and refined in PHENIX against measured data, using electron scattering form factors. The refinement statistics of the final structures are listed in Table S1.

Generation and Use of Peptide Fragment Libraries for Phasing MicroED Peptide Structures Using ARCIMBOLDO-BORGES

In ARCIMBOLDO, fragments are identified by likelihood-based molecular replacement[15] and expanded through density modification and map interpretation.[16] A library of amyloid peptide structures determined by both X-ray and electron diffraction was assembled to provide a diverse collection of backbone conformations with potential in phasing novel structures. To take advantage of these probes, a high-throughput, fragment-based phasing methodology in the form of the ARCIMBOLDO software was used. The ARCIMBOLDO suite of programs uses secondary structure fragments as initial probes for molecular replacement carried out by Phaser.[35] These fragments undergo rotation and translation analysis and are scored based on log-likelihood gain (LLG) and an initial correlation coefficient (InitCC) to identify potentially accurate starting models (Figure 1C). Rather than making an arbitrary choice on how to direct the superposition, determining the best average for the whole model or of a core, or tolerating outlier atoms to be excluded, an empirical answer was drawn simulating data from a template and refining the location of all other models against the calculated data.

Following this, initial maps are calculated and improved by density modification using the sphere-of-influence algorithm in SHELXE.[16] Finally, main chain autotracing is performed and solutions are scored by correlation coefficient.[36] Generally, a final CC greater than 25% is indicative of a correct solution. The ability to expand partial solutions in SHELXE permits the use of smaller, potentially more accurate, molecular replacement probes that are identified by Phaser. All runs were carried out using ARCIMBOLDO_BORGES version 2020 Phaser version 2.8.3, and the distributed SHELXE version 2019. Electron scattering factors were used for Phaser analysis, and

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

information including molecular weight and predicted solvent content are available upon request to the authors.

Generation of Custom and Rosetta Fragment Libraries

The fragment libraries used as probes were generated in multiple ways. Polyglycine libraries of varying sizes, four to six residues in length, were generated by extracting fragments from a collection of previously solved amyloid peptide structures (supplementary citations). These fragments were separated into custom libraries only containing fragments of the same length, aligned in PyMOL, and used as probes in ARCIMBOLDO_BORGES. The peptides QIGLAQTQ (plate) and NYNNYQ were phased using a polyglycine library of amino acid pentapeptides containing 89 fragments. The peptide QIGLAQTQ (needle) was phased using a polyglycine library of amino acid tetrapeptides containing 270 fragments. The peptide QYNNENNFV was phased using a polyglycine library of amino acid hexapeptides containing 249 fragments.

Additionally, two libraries of poly alanine fragments were generated using the program Rosetta[37] by modeling the packing of the peptide sequences threaded over the steric zipper structure of a previously determined amyloid peptide, NNQQNY.[27] After simple threading over the backbone, the side chains were packed, and the chains were allowed to relax to a calculated energy minimum over iterative cycles. These models were isolated as individual fragments six amino acids long, stripped to polyalanine side chains, and aligned in PyMOL to complete the Rosetta libraries. The Rosetta library for GSTVYAPFT was generated by threading all possible six amino acid segments of the LECT2 sequence against each other and then extracting only those which modeled the peptide sequence of interest. This library contained 111 fragments. The Rosetta library for FRNWQAYMQ was generated by threading all six amino acid long permutations of the peptide sequence against each other pairwise over the NNQQNY backbone. The threading, packing, and relaxation steps were done 20 separate times for each pair of sequences and the resulting library contained 640 single chain fragments.

General Strand Libraries Generated with ALEPH

The general libraries with strands generated by ALEPH[23] for use in ARCIMBOLDO_BORGES and distributed in CCP4 are all larger in size than any of the peptides investigated, as they contained a minimum of three strands forming a sheet in parallel or antiparallel arrangement. To obtain libraries of one or two strands, we started from the distributed three antiparallel strand (udu) library. Using a template of only two strands, we extracted all compatible models (around 24000) using ALEPH. Using the template model, we then generated an artificial data set in space group P1 to a resolution of 2.0 Å. As the models come from a standard library already, they were superposed originally based on their three strands and also clustered geometrically. To achieve that, we performed a Phaser rigid body refinement against the simulated data, using an rmsd error tolerance of 1.0 Å. We then selected the top LLG-scoring fragments for each of the geometric clusters and took them as our representatives for the two-strand library (246 models). Of those, outliers that did not superimpose well were removed. Only 108 models remained, each a pair of antiparallel beta strands. Then we generated two libraries of a single strand by extracting into separate PDB files from the parent library. This procedure yielded two general libraries to use in our experiments.

Structure Refinement

All structures were refined using Phenix version 1.16−3549−000[38,39] and Coot version 0.8.9.[40] All refinements used standard settings and the built-in electron scattering tables in Phenix. For each peptide, the final model output, called the best.pdb, from the ARCIMBOLDO-BORGES run was used directly as the model for the first round of refinement in Phenix. One exception was the QIGLAQTQ plate polymorph structure. In this case, the model from ARCIMBOLDO-BORGES was run through an additional round of Phaser molecular replacement to correct for what appeared to be a translational shift in the chain. This could be due to the extremely tight packing of the chains and the lack of any solvent, ordered or disordered, making initial chain placement difficult.

■ ASSOCIATED CONTENT

Data Availability Statement

All structures determined in this work have been deposited to the PDB. The deposited ARCIMBOLDO structures of peptides and their associated PDB IDs are QYNNENNFV (7N2J), NYNNYQ (7N2K), QIGLAQTQ_plate (7N2F), QIGLAQTQ_needle (7N2G), GSTVYAPFT (7N2I), and FRNWQAYMQ (7N2D). The deposited direct methods structures of peptides are QYNNENNFV (7N2L), NYNNYQ (7N2H), and QIGLAQTQ_plate (7N2E). The truncated .mtz file for the MicroED structure of NYNNYQ is available from the authors upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsbiomedchemau.2c00082.

Data showing statistics and structures of peptides phased *ab initio*, summaries of fragment-based phasing statistics, and details for peptide fragment libraires utilized (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Isabel Usón − *Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB−CSIC), 08028 Barcelona, Spain; ICREA, Institució Catalana de Recerca i Estudis Avançats, 08003 Barcelona, Spain*; Email: iufcri@ibmb.csic.es

Jose A. Rodriguez − *Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles (UCLA), Los Angeles, California 90095, United States*; orcid.org/0000-0002-0248-4964; Email: jrodriguez@mbi.ucla.edu

Authors

Logan S. Richards − *Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles (UCLA), Los Angeles, California 90095, United States*

Maria D. Flores − *Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles (UCLA), Los Angeles, California 90095, United States*

Claudia Millán − *Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB−CSIC), 08028 Barcelona, Spain*

Calina Glynn − *Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles (UCLA), Los Angeles, California 90095, United States*

Chih-Te Zee − *Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles (UCLA), Los Angeles, California 90095, United States*

Michael R. Sawaya − *Department of Biological Chemistry and Department of Chemistry and Biochemistry, University of California Los Angeles (UCLA), Howard Hughes Medical Institute (HHMI), UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California 90095, United States;* orcid.org/0000-0003-0874-9043

Marcus Gallagher-Jones − *Correlated Imaging, The Rosalind Franklin Institute, Didcot OX11 0GD, United Kingdom*

Rafael J. Borges − *Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB−CSIC), 08028 Barcelona, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsbiomedchemau.2c00082

Author Contributions

#L.S.R., M.D.F., and C.M. contributed equally to this work. J.A.R. and I.U. directed the research. L.S.R., M.D.F., and C.M. generated fragment libraries and performed phasing with ARCIMBOLDO. L.S.R, M.D.F, C.G., and J.A.R. crystallized peptides and prepared samples. C.G., M.G.J., and J.A.R. collected diffraction data. M.D.F., L.S.R., C.Z., C.G., M.R.S., and M.G.J processed diffraction data. C.M., I.U., and R.J.B. developed and evaluated MicroED specific features in the ARCIMBOLDO framework. All authors helped to write and provided critical feedback on the article. CRediT: **Logan S Richards** data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (lead); **Maria D Flores** data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (lead); **Claudia Millan** data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), resources (equal), software (lead), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (lead); **Calina Glynn** data curation (equal), formal analysis (equal), investigation (equal), validation (equal), writing-review & editing (equal); **Marcus Gallagher-Jones** data curation (equal), formal analysis (equal), investigation (equal), validation (equal), writing-review & editing (equal); **Rafael J. Borges** data curation (equal), formal analysis (equal), investigation (equal), methodology (equal), software (equal), writing-review & editing (equal); **Isabel Usón** conceptualization (lead), data curation (lead), formal analysis (lead), funding acquisition (lead), investigation (lead), methodology (lead), project administration (lead), resources (lead), software (lead), supervision (lead), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (lead).

Notes

The authors declare the following competing financial interest(s): J.A.R. is an equity stake holder of Medstruc, Inc.

ACKNOWLEDGMENTS

We thank Duilio Cascio (UCLA) for discussions and helpful analysis and the David Eisenberg laboratory (UCLA) for generous access to all of their amyloid peptide structures. This work was performed as part of STROBE, an NSF Science and Technology Center through Grant DMR-1548924. This work is also supported by DOE Grant DE-FC02-02ER63421 and NIH-NIGMS Grant R35 GM128867 and P41GM136508. L.S.R. is supported by the USPHS National Research Service Award 5T32GM008496. M.D.F. was funded by Eugene V. Cota-Robles Fellowship and Ruth L. Kirschstein NRSA GM007185 and is currently funded by a Whitcome Pre-Doctoral Fellowship and a National Science Foundation Graduate Research Fellowship. C.G. was funded by Ruth L. Kirschstein NRSA GM007185 and is currently funded by Ruth L. Kirschstein Predoctoral Individual NRSA, 1F31 AI143368. R.J.B. received fellowship from FAPESP (16/24191-8 and 17/13485-3). C.M. is grateful to MICINN for her BES-2015-071397 scholarship associated with the Structural Biology Maria de Maeztu Unit of Excellence. This work was supported by grants PGC2018-101370-B-I00 and PID2021-128751NB-I00 (MICINN/AEI/FEDER/UE) and Generalitat de Catalunya (2017SGR-1192) to I.U. J.A.R. is supported as a Pew Scholar, a Beckman Young Investigator, and a Packard Fellow. The Northeastern Collaborative Access Team beamline is funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

REFERENCES

(1) Brink, C.; Hodgkin, D. C.; Lindsey, J.; Pickworth, J.; Robertson, J. H.; White, J. G. Structure of Vitamin B12: X-Ray Crystallographic Evidence on the Structure of Vitamin B12. *Nature* **1954**, *174* (4443), 1169−1171.

(2) Sawaya, M. R.; Rodriguez, J.; Cascio, D.; Collazo, M. J.; Shi, D.; Reyes, F. E.; Hattne, J.; Gonen, T.; Eisenberg, D. S. Ab Initio Structure Determination from Prion Nanocrystals at Atomic Resolution by MicroED. *Proc. Natl. Acad. Sci.* **2016**, *113* (40), 11232−11236.

(3) 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.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the Toxic Core of α-Synuclein from Invisible Crystals. *Nature* **2015**, *525* (7570), 486−490.

(4) Gallagher-Jones, M.; Glynn, C.; Boyer, D. R.; Martynowycz, M. W.; Hernandez, E.; Miao, J.; Zee, C.-T.; Novikova, I. V.; Goldschmidt, L.; McFarlane, H. T.; Helguera, G. F.; Evans, J. E.; Sawaya, M. R.; Cascio, D.; Eisenberg, D. S.; Gonen, T.; Rodriguez, J. A. Sub-Ångström Cryo-EM Structure of a Prion Protofibril Reveals a Polar Clasp. *Nat. Struct. Mol. Biol.* **2018**, *25* (2), 131−134.

(5) Henderson, R. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Q. Rev. Biophys.* **1995**, *28* (2), 171−193.

(6) Warmack, R. A.; Boyer, D. R.; Zee, C.-T.; Richards, L. S.; Sawaya, M. R.; Cascio, D.; Gonen, T.; Eisenberg, D. S.; Clarke, S. G. Structure of Amyloid-β (20−34) with Alzheimer's-Associated Isomerization at Asp23 Reveals a Distinct Protofilament Interface. *Nat. Commun.* **2019**, *10* (1), 1−12.

(7) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **2018**, *4* (11), 1587−1592.

(8) Ting, C. P.; Funk, M. A.; Halaby, S. L.; Zhang, Z.; Gonen, T.; van der Donk, W. A. Use of a Scaffold Peptide in the Biosynthesis of Amino Acid Derived Natural Products. *Science* **2019**, *365* (6450), 280−284.

(9) Sheldrick, G. M.; Gilmore, C. J.; Hauptman, H. A.; Weeks, C. M.; Miller, R.; Usón, I. Ab Initio Phasing. In *International Tables for Crystallography*; Brock, C. P., Hahn, T., Wondratschek, U., Müller, U.,

Shmueli, U., Prince, E., Auhier, A., Kopský, V., Litvin, D. B., Arnold, E. et al., Eds.; John Wiley & Sons, Ltd., 2012; pp 413−432.
(10) Rossmann, M. G. The Molecular Replacement Method. *Acta Crystallogr. A* **1990**, *46*, 73−82.
(11) Shi, D.; Nannenga, B. L.; Iadanza, M. G.; Gonen, T. Three-Dimensional Electron Crystallography of Protein Microcrystals. *eLife* **2013**, *2*, No. e01345.
(12) Krotee, P.; Rodriguez, J. A.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Hattne, J.; Nannenga, B. L.; Oskarsson, M. E.; Philipp, S.; Griner, S.; Jiang, L.; Glabe, C. G.; Westermark, G. T.; Gonen, T.; Eisenberg, D. S. Atomic Structures of Fibrillar Segments of HIAPP Suggest Tightly Mated β-Sheets Are Important for Cytotoxicity. *eLife* **2017**, *6*, No. e19273.
(13) de la Cruz, M. J.; Hattne, J.; Shi, D.; Seidler, P.; Rodriguez, J.; Reyes, F. E.; Sawaya, M. R.; Cascio, D.; Weiss, S. C.; Kim, S. K.; Hinck, C. S.; Hinck, A. P.; Calero, G.; Eisenberg, D.; Gonen, T. Atomic Resolution Structures from Fragmented Protein Crystals by the CryoEM Method MicroED. *Nat. Methods* **2017**, *14* (4), 399−402.
(14) Rodríguez, D. D.; Grosse, C.; Himmel, S.; González, C.; de Ilarduya, I. M.; Becker, S.; Sheldrick, G. M.; Usón, I. Crystallographic Ab Initio Protein Structure Solution below Atomic Resolution. *Nat. Methods* **2009**, *6* (9), 651−653.
(15) Read, R. J.; McCoy, A. J. A Log-Likelihood-Gain Intensity Target for Crystallographic Phasing That Accounts for Experimental Error. *Acta Crystallogr., Sect. D: Struct. Biol.* **2016**, *72*, 375−387.
(16) Usón, I.; Sheldrick, G. M. An Introduction to Experimental Phasing of Macromolecules Illustrated by SHELX; New Autotracing Features. *Acta Crystallogr. Sect. Struct. Biol.* **2018**, *74* (2), 106−116.
(17) Millán, C.; Sammito, M.; Usón, I. Macromolecular Ab Initio Phasing Enforcing Secondary and Tertiary Structure. *IUCrJ* **2015**, *2*, 95−105.
(18) Sammito, M.; Millán, C.; Frieske, D.; Rodríguez-Freire, E.; Borges, R. J.; Usón, I. ARCIMBOLDO_LITE: Single-Workstation Implementation and Use. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 1921−1930.
(19) Sammito, M.; Millán, C.; Rodríguez, D. D.; de Ilarduya, I. M.; Meindl, K.; De Marino, I.; Petrillo, G.; Buey, R. M.; de Pereda, J. M.; Zeth, K.; Sheldrick, G. M.; Usón, I. Exploiting Tertiary Structure through Local Folds for Crystallographic Phasing. *Nat. Methods* **2013**, *10* (11), 1099−1101.
(20) Millán, C.; Sammito, M. D.; McCoy, A. J.; Nascimento, A. F. Z.; Petrillo, G.; Oeffner, R. D.; Domínguez-Gil, T.; Hermoso, J. A.; Read, R. J.; Usón, I. Exploiting Distant Homologues for Phasing through the Generation of Compact Fragments, Local Fold Refinement and Partial Solution Combination. *Acta Crystallogr. Sect. Struct. Biol.* **2018**, *74* (4), 290−304.
(21) Richards, L. S.; Millán, C.; Miao, J.; Martynowycz, M. W.; Sawaya, M. R.; Gonen, T.; Borges, R. J.; Usón, I.; Rodriguez, J. A. Fragment-Based Determination of a Proteinase K Structure from MicroED Data Using ARCIMBOLDO_SHREDDER. *Acta Crystallogr. Sect. Struct. Biol.* **2020**, *76* (8), 703−712.
(22) Thorn, A.; Sheldrick, G. M. Extending Molecular-Replacement Solutions with *SHELXE*. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (11), 2251−2256.
(23) Medina, A.; Triviño, J.; Borges, R. J.; Millán, C.; Usón, I.; Sammito, M. D. ALEPH: A Network-Oriented Approach for the Generation of Fragment-Based Libraries and for Structure Interpretation. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2020**, *76*, 193−208.
(24) Millán, C.; Jiménez, E.; Schuster, A.; Diederichs, K.; Usón, I. ALIXE: A Phase-Combination Tool for Fragment-Based Molecular Replacement. *Acta Crystallogr. Sect. Struct. Biol.* **2020**, *76* (3), 209−220.
(25) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−3048.
(26) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: A Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinforma. Oxf. Engl.* **2010**, *26* (5), 689−691.
(27) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. Identifying the Amylome, Proteins Capable of Forming Amyloid-like Fibrils. *Proc. Natl. Acad. Sci.* **2010**, *107* (8), 3487−3492.
(28) Caballero, I.; Sammito, M.; Millán, C.; Lebedev, A.; Soler, N.; Usón, I. ARCIMBOLDO on Coiled Coils. *Acta Crystallogr. Sect. Struct. Biol.* **2018**, *74* (3), 194−204.
(29) Tayeb-Fligelman, E.; Tabachnikov, O.; Moshe, A.; Goldshmidt-Tran, O.; Sawaya, M. R.; Coquelle, N.; Colletier, J.-P.; Landau, M. The Cytotoxic Staphylococcus Aureus PSMα3 Reveals a Cross-α Amyloid-like Fibril. *Science* **2017**, *355* (6327), 831−833.
(30) Salinas, N.; Tayeb-Fligelman, E.; Sammito, M. D.; Bloch, D.; Jelinek, R.; Noy, D.; Usón, I.; Landau, M. The Amphibian Antimicrobial Peptide Uperin 3.5 Is a Cross-α/Cross-β Chameleon Functional Amyloid. *Proc. Natl. Acad. Sci.* **2021**, *118* (3), No. e2014442118, DOI: 10.1073/pnas.2014442118.
(31) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706−710.
(32) Kabsch, W. Integration, Scaling, Space-Group Assignment and Post-Refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 133−144.
(33) Usón, I.; Sheldrick, G. M. Advances in Direct Methods for Protein Crystallography. *Curr. Opin. Struct. Biol.* **1999**, *9* (5), 643−648.
(34) Sheldrick, G. M. Crystal Structure Refinement with SHELXL. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (1), 3−8.
(35) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(36) Fujinaga, M.; Read, R. J. Experiences with a New Translation-Function Program. *J. Appl. Crystallogr.* **1987**, *20* (6), 517−521.
(37) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol.* **2011**, *487*, 545−574.
(38) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; et al. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* **2019**, *75*, 861−877.
(39) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (4), 352−367.
(40) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.

https://doi.org/10.1021/acsbiomedchemau.2c00082
ACS Bio Med Chem Au XXXX, XXX, XXX−XXX

Chapter 3

This chapter is based on the preprint article titled “Structural insights into functional regulation of the human CPEB3 prion by a labile-amyloid-forming segment”

Summary

The work described in this article begins with *in vitro* experiments focusing on CPEB3’s PRD1. As mentioned in Chapter 1, previous findings from Eric Kandel’s lab had characterized the functional relevance of the first 217 residues of CPEB3 in primary neurons and knock-out mouse lines. Here, I show that recombinantly expressed PRD1 is capable of forming reversible amyloid fibrils and utilize single particle cryo-EM and helical reconstruction to reveal the ordered amyloid core of CPEB3’s PRD1. To further investigate the significance of our amyloid core in neurons, I performed immunofluorescence experiments and translational assays. I report that the 42-residue segment identified in the cryo-EM structure is required for proper cellular storage in processing bodies and for translational regulation. In search of the structures of CPEB3 within its native environment, I apply the use of correlative cryo-FIB milling and cryo-ET to select and target U87 cells expressing CPEB3 and displaying traditional puncta aggregates. In cells exogenously expressing CPEB3, I find evidence of ultra-structures that may be related to cellular stress. I further probe these results with a combination of viability assays and transcriptional profiling. My data show that persistent, high level expression of CPEB3 is toxic to cells and elicits immune and inflammatory responses. Future insights into the prion-like role of CPEB3 in neurons and its regulated assembly and disassembly in cells will continue to expand our fundamental understanding of functional prion proteins and how they are regulated and tolerated in the brain.

Ultimately, when the technology is ready, the structure of CPEB3 assemblies and their interacting partners in the brain will be important targets in the quest to understand the molecular basis for cellular memory.

Flores, M. D., Sawaya, M. R., Boyer, D. R., Zink, S., Tovmasyan, S., Saucedo, A., Zee, C.-T., Cardenas, J., Fioriti, L., & Rodriguez, J. A. (2022). Structure of a reversible amyloid fibril formed by the CPEB3 prion-like domain reveals a core sequence involved in translational regulation. https://doi.org/10.1101/2022.12.07.519389

Title: Structural insights into functional regulation of the human CPEB3 prion by a labile-amyloid-forming segment

Maria D. Flores*[1], Michael R. Sawaya[2], David R. Boyer[2], Samantha Zink[1], Susanna Tovmasyan[1], Adrian Saucedo[1], Chih-Te Zee[1], Jorge Cardenas[1], Luana Fioriti*[3,4] and Jose A. Rodriguez*[1]

Affiliations

[1]Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center; University of California, Los Angeles (UCLA); Los Angeles, CA 90095, USA.

[2]Department of Biological Chemistry and Department of Chemistry and Biochemistry, University of California Los Angeles (UCLA), Howard Hughes Medical Institute (HHMI), UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA 90095, USA.

[3]Department of Neuroscience, Istituto di Ricerche Farmacologiche Mario Negri, Milano MI, 20156, Italy.

[4]Department of Neuroscience, Columbia University, New York, NY 10027, USA.

* Correspondence to: Maria D. Flores (mflores@mbi.ucla.edu), Luana Fioriti (luana.fioriti@marionegri.it), Jose A. Rodriguez (jrodriguez@mbi.ucla.edu)

Abstract

The cytoplasmic polyadenylation element-binding protein 3 (CPEB3) is a functional prion. CPEB3 is thought to modulate protein synthesis at synapses and enable consolidation of long-

term memory in neurons. Here, we report that prion-like domain 1 of CPEB3 self-assembles into labile amyloid fibrils *in vitro*. A cryo-EM structure of these fibrils reveals an ordered 48-residue core, spanning L103 to F151. CPEB3 lacking this amyloidogenic segment assembles into abnormal puncta in cells when compared to wild type CPEB3, and concomitantly delocalizes away from dormant p-bodies and toward stress granules. Removal of the amyloid core segment in CPEB3 also abolishes its ability to regulate protein synthesis in neurons. By applying fluorescence-guided cryo-FIB milling and cryo-ET, we gather ultrastructural snapshots of neuronal cells expressing CPEB3 and find CPEB3-GFP signal associated with lamellae enriched in multivesicular bodies (MVBs), multilamellar cavernous compartments, and bundled filaments. These features suggest a state of induced subcellular stress driven by expression of CPEB3 constructs equipped with an amyloid core. Accordingly, cells expressing wild type CPEB3 show lower viability than those expressing CPEB3 without an amyloid core. This suggests that regulation of CPEB3 function mediated by an amyloidogenic sequence in its first prion-like domain requires cellular effort to overcome the potential liability associated with its formation of prion assemblies in cells.

Introduction

Prions, or proteinaceous infectious particles[1], are stable, long-lived, pathogenic or functional assemblies present in species across the tree of life[2]. Many prions rely on an amyloid architecture to recruit naïve proteins and template their conversion into ordered fibrillar β-sheet-rich assemblies [2,3]. Because of their ability to adopt amyloid folds, prion-like proteins present a challenge to the proteostatic framework in neurons[4]. However, since prions can endure protein turnover, they also present an avenue to influence memory consolidation[5].

Prion-based mechanisms have been described to influence learning and memory in *Aplysia*[6], *Drosophila*[7,8] and most recently, mammals[9,10]. Neuronal isoforms of the cytoplasmic polyadenylation element binding protein (CPEB), a functional prion, are involved in regulating localized translation at synapses[4]. Prion-like fibrillar aggregates formed by the *Drosophila* CPEB ortholog, Orb2, isolated from fly brains, have been shown to adopt steric zipper motifs. Their dry amyloid core is formed by tightly mated sheets that meet at interdigitated glutamine side chains. The discovery of this structure supports the theory that CPEB orthologs can form persistent or stable amyloid assemblies, but leaves unknown the origin of these assemblies, their stability in cells, and the mechanism by which they avoid pathology[14]. Additionally, while these studies present remarkable views of functional prion structures, they offer limited insights into the prion-like function of mammalian CPEBs[6,7,10].

CPEB3 influences memory consolidation in humans. It is a 716-residue protein with two N-terminal prion-like domains linked by a short, low-complexity regulatory motif (Figure 1a). CPEB3 is unique among CPEB orthologs in that it contains an N-terminal low complexity domain[8], akin to those found in functional yeast prions[2]. The CPEB3 prion domains influences the activity of its two C-terminal RNA-binding domains and zinc finger domain[8] (Figure 1a). As an RNA binding protein, CPEB3 regulates translation of target mRNAs that are required for synaptic plasticity and the growth of new synaptic spines[9,11,12]. The N-terminal, prion-like domain 1 (PRD1) of CPEB3 rescues aggregation phenotypes in cells and mediates long-term potentiation (LTP) as well as the persistence of memory in mice[9]. CPEB3 aggregates are regulated in part through mono-ubiquitination[11] and SUMOylation[13], which allow for its proper activation and repression, respectively. In its basal state, CPEB3 is SUMOylated and is postulated to undergo phase separation when it colocalizes to p-bodies where it inhibits translation of target mRNAs[12]. Upon proper signal induced activity it is then ubiquitinated and shuttled to polysomes where it increases translation of its targets[12]. These observations support

the notion that the structure, decoration and localization of prion-like CPEB3 assemblies are linked to its function.

The prion-like abilities of CPEB3 are showcased by its formation of heritable aggregates in yeast[10]. Similar aggregates appear to be required for its function in neurons[10]. Our current view of mammalian CPEB3 prion structures is informed by sequence analyses[15], which have revealed segments in the PRD1 of CPEB3 that may form higher-order coiled-coiled α-helical assemblies. In addition, NMR spectroscopy studies predict the presence of a β-rich ordered core from mouse CPEB3 spanning T101-S194 and show their propensity to form amyloid-like fibrils.[16] However, these studies present a limited view of the physical state of CPEB in cells and of the atomic structures adopted by assemblies of the CPEB3 prion-like domain.

With the goal of improving our understanding of CPEB3 functional prions mechanisms, we investigated the CPEB3 PRD1 and found it capable of self-assembling into reversible amyloid fibrils in solution. The structure of these fibrils, determined by single particle cryoEM, showed an ordered core composed of residues 103 to 151 of CPEB3 arranged in a serpentine fold. The sequence composition of this amyloid core (AC) offered clues to its lability and possible sensitivity to environmental changes in pH. CPEB3 constructs lacking this AC sequence showed altered subcellular distribution, indicating a role for this sequence in CPEB3 activity. This hypothesis was confirmed by the altered translational regulation of AC-deficient CPEB3 in primary neurons. To improve our understanding of CPEB3 assemblies in cells, we sought molecular snapshots of CPEB3-GFP inclusions in unstimulated U87 cells using cryo-focused ion beam (cryo-FIB) milling and correlative cryo-ET. We found CPEB3-GFP signal in lamellae that also contained an abundance of multivesicular bodies (MVBs) and dense fibrillar structures, vesicles and phase-separated droplets. Collectively, our observations support a model in which

CPEB3 function is regulated in part by its AC sequence in PRD1, and offer clues into the storage and regulation of prion-like inclusions in neuronal cells.

Results

CPEB3 PRD1 forms reversible amyloid fibrils *in vitro*.

Hypothesizing that CPEB3 PRD1 may form functional amyloid assemblies, we expressed and isolated a construct of human CPEB3 encoding its first 217 residues that included the entirety of its conserved PRD1 (Panel 1, Supplementary Figure 1). This region of CPEB3 has been previously identified as essential for prion-seeding, heritability[10] and is essential for LTP in mice[9]. We assessed the ability of this construct ($CPEB3_{1\text{-}217}$) to form fibrils *in vitro* by light scattering, enhanced fluorescence of the amyloid-binding dye ThT, and negative stain electron microscopy (Supplementary Figure 2). $CPEB3_{1\text{-}217}$ tended to precipitate in denaturing solutions as amorphous aggregates, twisting fibrils, or flat, ribbon like sheets. In solutions containing 125 mM NaCl, 50 mM Tris-Base, 10mM K_2HPO_4 and 5 mM glutamic acid at pH between 4 and 5 (assembly buffer), $CPEB3_{1\text{-}217}$ produced what appeared by negative stain electron microscopy to be fibrils with a homogenous twist (Supplementary Figure 2b). Solutions containing $CPEB3_{1\text{-}217}$ fibrils exhibited enhanced ThT fluorescence (Supplementary Figure 2a). These fibrils were also labile, disassembling when diluted into water or exposed to low ionic strength solutions (Supplementary Figure 2c). To pursue structure determination in the face of this lability, fibrils had to be vitrified within an hour of formation in assembly buffer.

Structure of the CPEB3 amyloid fibril core

Solutions containing $CPEB3_{1\text{-}217}$ fibrils were interrogated by single particle cryoEM, with the goal of *de novo* structure determination (Supplementary Figure 4). CryoEM images confirmed that $CPEB3_{1\text{-}217}$ forms rapidly twisting amyloid fibrils in assembly buffer, with a single polymorph

representing over 90% of imaged particles (Supplementary Figure 5a). The remaining 10% of species exhibited variable twist or no twist, and therefore eluded further classification and structure determination (Supplementary Figure 5b). Two-dimensional class averages confirmed canonical amyloid features corresponding to the stacking of β-sheets separated by ~4.8 Å inter-strand distances, as well as a disordered flanking region, or fuzzy coat, surrounding the fibril core (Supplementary Figure 5b). Three dimensional reconstructions obtained from 40,329 boxed fibril segments yielded a 3.0 Å map that showed density for a single asymmetric protofilament that fit residues L103-F151 of CPEB3 (Figure 1b). We refer to this structure and its corresponding core sequence as the CPEB3 amyloid core ($CPEB3_{AC}$).

$CPEB3_{AC}$ fibrils adopt a helical pitch of 536 Å and a twist of -3.32° (Figure 1c). In $CPEB3_{AC}$ fibrils, molecules stack as strands, parallel and in register as part of intermolecular β-sheets. The tight serpentine fold adopted by the $CPEB3_{AC}$ core is enabled by five sharp kinks interspersed between its seven β-strands (Figure 1c). The core contains two hydrophobic pockets and encloses three internal solvent channels. It displays a total solvent accessible surface area per chain of 1653 $Å^2$. At its N-terminus, L103 contacts a surface-facing segment of a beta turn in the fibril core, making a hydrophobic pocket composed of residues P105, F107, G126 and I127 (Supplementary Figure 7b). An additional hydrophobic pocket is formed by residues F124, V130, M134 and F136 (Supplementary Figure 7b). The core is further stabilized by stacking of aromatic residues, although some appear solvent facing, including W111, F139, and F151. Polar residues T110, S112 and T116 face the largest solvent channel buried within the fibril core. A second, smaller solvent channel is lined by a group of largely hydrophobic residues: V119, P139, V143, P145; E120 stands as a lone hydrophilic residue in this group. This same glutamic acid forms part of a minute channel within the core, which is also lined by

S122 and N138. The significance of E120 bridging these two solvent channels is unclear, but may indicate a role for it in the lability of the $CPEB3_{AC}$ core.

Structural features of labile $CPEB_{AC}$ fibrils

Energy values calculated using coordinates of the $CPEB3_{AC}$ fibril core show that, per layer, this core is less stable than those of pathogenic amyloids (Supplementary Figure 7a)[17]. Estimates of its free energy are more in line with those of fibril structures formed by reversible, functional amyloids[18–20]. Solvent facing hydrophobic residues and buried solvent channels play a role in its predicted lability, but the $CPEB3_{AC}$ is also strained by a series of kinks. Some of these kinks tilt strands out of plane, warping the layers of the fibril to create a height difference of 14.5 Å across each of its layers (Figure 1d, Supplementary Figure 7c). While the role of warping in amyloid fibril cores remains unclear, it may help relieve the strain associated with the large twist angle observed in $CPEB3_{AC}$ fibrils. In addition, some kinks appear to be stabilized by local interactions. For example, a kink near G115 facilitates hydrogen-bonding between side chains of S112 and T116. Four of the 28 proline residues in $CPEB3_{1\text{-}217}$ are resolved in the core. The remaining 24 proline residues are unresolved in the fuzzy coat of $CPEB3_{AC}$ and may function to limit the size and stability of the ordered core. They appear as smeared density in 2D averages of $CPEB3_{AC}$ fibrils (Supplementary Figure 4b). While some of these regions are hypothesized to form coiled-coils[15], that region of the fibril remains insufficiently resolved to assign secondary structure.

The role of negatively charged residues in $CPEB_{AC}$

We observed four potentially charged residues in $CPEB3_{AC}$: two histidine residues (H141 and H142) are solvent exposed, and two acidic residues (E120 and D121) are buried within the fibril core (Supplementary Figure 7a). The two histidine residues are located too far from the acidic residues for charge complementarity to stabilize the fold. E120 is located in the second solvent

channel and is otherwise surrounded by hydrophobic residues, while D121 faces the opposing solvent channel (Supplementary Figure 7b). As $CPEB3_{AC}$ aggregates near a pH of 5, E120 and D121, are anticipated to remain protonated within the fibril core and are capable of forming stabilizing hydrogen bonds. These same residues would render the core susceptible to pH triggered changes, as observed with other functional amyloids [21,22]such as pmel17[21,22] and β-endorphin[23]. To evaluate this hypothesis, we monitored the stability of $CPEB3_{AC}$ fibrils in response to changing buffer conditions, using a ThT fluorescence assay. We found $CPEB3_{AC}$ fibrils diminish in abundance as pH increases or as the ionic strength of the buffer decreases, denoted by a decrease of ThT signal and the absence of fibrils in negative stain images compared to fibrils that remain in their assembly buffer (Supplementary Figure 2c). This dissolution in response to changing environmental conditions is recapitulated by simulations tracking fibril stability over a period of 200 nanoseconds (Supplementary Figure 8a). Interlayer distances in a 5-layer fibrillar $CPEB3_{AC}$ structure are disrupted as a pH increase from 3 to 7 is simulated; structural deviations are pronounced in charged residues E120 and D121 (Supplementary Figure 8b).

To assess whether E120 and D121 influence the morphology, distribution, and abundance of CPEB3 aggregates in cells, we expressed GFP-tagged full-length CPEB3 constructs encoding wild type and E120/D121 variations, in HT22 and U87 neuronal cells (Figure 2a,b). In both cell types, wild type, GFP-fused CPEB3 formed fine, granular fluorescent puncta, as previously described (Figure 2c, Supplementary Figure 9a)[10,12]. Forty-eight hours after transduction, the fine puncta were replaced by larger puncta (Supplementary Figure 9a). To probe the influence of acidic side chains at residues ^{120}ED^{121} we expressed a CPEB3 variant encoding ^{120}QN^{121}, which retains the polarity and size of the wild type residues but ablates their negative charge. We also expressed a variant encoding ^{120}AA^{121} to assess the effect of ablating polarity and reducing the size of wild type residues. Puncta formed by both constructs differed

morphologically from those formed by wild type CPEB3-GFP under similar conditions (Figure 2c, Supplementary Figure 11). CPEB3-^{120}QN^{121} generally formed irregularly sized and shaped puncta, while CPEB3-^{120}AA^{121} formed fewer puncta that were intermediate in size and remained more circular after 24hrs (Figure 2d, Supplementary Figure 10a, Supplementary Figure 11). The effects of these residue substitutions on the size and distribution of CPEB3$_{AC}$ puncta implicate a functional role for E120 and D121.

CPEB3$_{AC}$ is required for proper aggregation and localization in cells.

To more generally assess the functional relevance of CPEB3$_{AC}$ in cells, we asked whether the presence of the AC sequence affected CPEB3 aggregation. While CPEB3 is known to form aggregates of various sizes when overexpressed in cells[9,10,12], a construct of CPEB3 lacking its amyloid core, CPEB3$_{\Delta AC}$ displayed a reduced capacity to form small aggregates (Supplementary Figure 9). Instead, this construct showed a delayed onset of larger aggregate formation as compared to wild type CPEB3 in both HT22 and U87 cells (Supplementary Figure 9). This prompted us to determine whether the localization of CPEB3$_{\Delta AC}$ differed from that of wild type CPEB3 in both cell lines. Under normal conditions, CPEB3 is expected to localize in p-bodies[12] that contain mRNAs and associated regulatory proteins. There, mRNAs are stored in a dormant state while they are trafficked to their destination to be translated, or can be degraded. Alternatively, CPEB3 could reside in stress granules (SG), induced to promote cell survival by condensing translationally stalled mRNAs.[23] We observed that wild type CPEB3-GFP aggregates primarily colocalized with the de-capping coactivator and p-body marker, Dcp1a, in unstimulated cells (Figure 3b). In contrast, in mouse HT22 hippocampal neurons, CPEB3$_{\Delta AC}$ aggregates primarily colocalized with the SG assembly endoribonuclease, G3BP, and minimally colocalized with the p-body marker, Dcp1a (Figure 3b). Similar localization patterns were observed in human U87 glioblastoma cells (Figure 3c).

CPEB3$_{AC}$ impacts translational regulation by CPEB3

In its basal state, CPEB3 represses translation of target mRNAs, is SUMOylated, and colocalizes to p-bodies. After neuronal stimulation CPEB3 undergoes de-SUMOylation, leaves p-bodies and is then shuttled to polysomes to promotes translation of targets[12] . The increase in translational activity of CPEB3 is thought to be facilitated by monoubiquitination of PRD1 by Neuralized-1[11]. The ubiquitination site of CPEB3 is currently unknown and efforts to identify and characterize all SUMOylation sites[13] have not yielded a complete understanding of their role in CPEB3 activity. Since CPEB3$_{AC}$ is situated within PRD1, we postulated it may influence translational regulation by CPEB3. To test this hypothesis, we assessed translation of CPEB3 targets, comparing the impact of wild type protein to CPEB3$_{\Delta AC}$. Translation of Renilla luciferase fused to the 3'UTR of two established targets of CPEB3 targets, GluA2 and SUMO2 3' UTR, in HEK-293 cells, was strongly inhibited by both wild type CPEB3 and CPEB3 lacking PRD1 (ΔN) (Figure 3d). In contrast, translation of the same targets was unaffected by CPEB3$_{\Delta AC}$; a marked change in translational regulation in the absence of the amyloid core segment. Co-expression of wild type and CPEB3$_{\Delta AC}$ in cells re-established the translational repression observed with the wild type protein alone, indicating CPEB3$_{\Delta AC}$ does not act in dominant negative fashion (Figure 3d).

Expression of wild type CPEB3 in primary neurons repressed translation of its targets in unstimulated cells (Figure 3e). To stimulate neurons, we used a protocol consisting of bath application of bicuculline and glycine, a chemical treatment that induces long-term potentiation (LTP) of synapses (cLTP)[24]. As expected, glycine-mediated cLTP induced the translation of CPEB3 targets (Figure 3e). In contrast, expression of CPEB3$_{\Delta AC}$ under the same conditions ablated all translational regulation, rendering the protein essentially inactive in this capacity (Figure 3e). In fact, luciferase activity in the presence of CPEB3$_{\Delta AC}$ was most similar to that of

cells expressing GFP alone, indicating that $CPEB3_{AC}$ is required for regulated translation of CPEB3 targets.

Subcellular structure of U87 cells expressing CPEB3-GFP

Hypothesizing that the subcellular structure of U87 cells expressing CPEB3-GFP might offer clues into its structure and function *in situ*, we targeted fluorescent puncta in these cells for focused ion beam (FIB) milling and cryo electron tomography (cryoET), reconstructing tomograms from cellular lamellae with associated CPEB3-GFP fluorescence. U87 cells grown directly on EM grids and expressing CPEB3-GFP formed fluorescent inclusions; after 24 hours of expression, grids were flash frozen and loaded onto a cryo-FIB-SEM instrument equipped with an integrated fluorescence objective that allowed targeting of sufficiently large cytoplasmic inclusions in cells (Supplementary Figure 11). We found that GFP signal was associated with lamellae containing several cellular structures, including: an enrichment of multivesicular bodies (MVBs), cavernous compartments, aligned cytoskeletal filaments, autophagic bodies and extensive networks of vesicles (Figure 4, Figure 5). Some of the observed MVBs were found to contain densely packed structures, occasionally appearing to harbor fibrillar assemblies ~10 nm in diameter, packed in tight parallel fashion and resembling yarn balls (Figure 4a, 4c). Neighboring these structures were also intralumenal vesicles and droplets (Figure 4d). Interestingly, MVBs were often found in clusters, in the vicinity of elongated membranes and droplets. Notably absent from all surveyed tomograms of CPEB3-GFP expressing cells were the large fibrillar inclusion bodies observed in cells expressing pathogenic amyloids[25].

In addition to the yarn ball structures observed in MVBs, our iterative sampling of lamellae from CPEB3 expressing cells revealed uniquely shaped, cavernous compartments (Figure 5a). These compartments were layered, but distinct from onion-like concentric lipid droplets (LDs); their layers were nearly twice the diameter of other reported striations[26]. The layered structures

could be categorized by size. A small subset of compartments were less than ~ 0.5 μM in diameter and displayed ordered spiraled shells that entwined toward their interior (Figure 5a). The outer layers of the compartments displayed regular periodic structures, ~7 nm in cross-section. The components of these regularly spaced arrays were seen to twist toward the center of each compartment, blending into less distinguishable features (Figure 5b). Notably, the smaller compartments often contained cavernous spaces encasing poorly delineated densities (Figure 5c). We also observed ordered assemblies that formed large cavernous structures measuring up to 1 μM in diameter in surveyed lamellae (Figure 5d). Line profiles across layers of the cavernous compartment borders showed regular spacings of ~3-4 nm (Figure 5e, 5f). These features share characteristics with LDs composed of cholesteryl-esters (CEs) seen in mitotically arrested HeLa cells[27]. However, the architecture of these structures does not resemble canonical subcellular organelles.

Persistent elevated expression of CPEB3 induces apoptosis and triggers stress response

Features observed in cells expressing CPEB3-GFP suggested a change in the physiological state of the cell. We now asked if persistent exogenous expression of CPEB3-GFP could induce distress in U87 cells. We performed viability assays simultaneously staining for the apoptotic marker, Annexin and 7-amino-actinomycin D, to account for necrotic cells. Using flow cytometry, we found that after 24 hours of expression, approximately 13% of WT expressing cells are apoptotic, where cells expressing CPEB3$^{\Delta AC}$-GFP were closer to GFP control wells with 7% and 5% of cells staining positive for apoptotic marker, respectively (Supplementary Figure 12a). To evaluate the molecular sequelae associated with this cellular response, we analyzed transcriptional profiles of HT22 hippocampal neurons expressing WT CPEB3-GFP, CPEB3$^{\Delta AC}$-GFP, GFP and untreated cells. Based on normalized fold changes compared against both control samples, cells expressing CPEB3-GFP contained elevated transcripts of a variety of

developmental and disease associated gene clusters (Supplementary Figure 12b). To specifically probe the role of the CPEB3 AC, we performed a cross comparison of normalized fold changes in WT and ΔAC cells (Supplementary Figure 12c), identifying genes that were unique to WT cells). Here, we found that half of identified gene targets were involved in immune and inflammatory responses (Supplementary Figure 12d), including a handful of guanylate-binding proteins (GBPs) that are functionally involved in pathogen protection responses[28]. These transcript changes and associated pathways indicate a cellular state of alert that is accentuated in wild type vs AC-deficient CPEB3-GFP expression.

Discussion

CPEB3, a member of the cytoplasmic polyadenylation family of proteins, binds mRNA transcripts to activate or repress protein synthesis[29,30]. This process was hypothesized to be regulated by its prion-like self-assembly[4], although the formation of prion-like structures by CPEB3 remained unwitnessed. We now lend direct evidence to this hypothesis by finding that a segment in the first CPEB3 prion-like domain forms labile amyloid fibrils. The ordered core of fibrils formed by this segment, $CPEB3_{AC}$, is composed of a single asymmetric protofilament with parallel, in register strands. However, while the overall architecture of $CPEB3_{AC}$ resembles that of a canonical amyloid structure, it contains features that may decrease its stability, such as buried polar residues and solvent channels. The ready assembly and disassembly of these fibrils via this amyloid core proving an important feature for the role of CPEB3 as a translational regulator.

The amino terminal domains of many CPEB proteins are identified as prion-like in part due to their low-complexity sequences and their high number of uncharged polar residues[4,6,31]. In agreement with our data, previous experimental predictions of the mouse PRD1 core note the

most rigid fragment E124-H145 divided into three main β-strands[16]. This is in support of $CPEB3_{AC}$, which contains the largest hydrophobic pockets split between four β-sheets (Figure 1, Supplementary Figure 7b). Specifically, we identify a short 48-residue segment within PRD1 forms the core of reversible fibrils. However, the reported proteinase K resistance of the PRD1 predicted core fragment[16] is in stark contrast to our data. This could be in part due to our construct containing the entire PRD1, including the proline and glutamine rich segments which are likely key contributors to the labile nature of the fold. Other features that may contribute to the lability of $CPEB3_{AC}$ include its buried charged residues (E120, D121), its four prolines, buried solvent channels, and its highly kinked and warped layers. The buried charged residues could allow $CPEB3_{AC}$ to respond to pH changes [32], enabling environmental queues to possibly regulate assembly of amyloid-like $CPEB3_{AC}$ structures in neurons alongside other post-translational modifications (PTMs) involved in CPEB3 regulation in cells[11–13]. In fact, ablation of these sites altered spatial distribution of CPEB3 in cells (Fig 2).

The physiological relevance of $CPEB3_{AC}$ is underscored by the fact that translational regulation by CPEB3 is ablated in cells when it lacks its AC. Although the structures of amyloid fibrils reconstituted *in vitro* are known to be distinct from those isolated from disease tissue[33–37], the relevance of $CPEB3_{AC}$ to physiological CPEB3 function remains supported by several lines of evidence. 1) the role of its AC sequence in localization to puncta in cells; 2) the fact that point mutations ablating charges in $CPEB3_{AC}$ alter CPEB3 distribution in cells; 3) cellular stress and loss of viability induced by wild type CPEB3 overproduction in cells is absent in those lacking $CPEB3_{AC}$; and most importantly, 4) that deletion of $CPEB3_{AC}$ eliminates its ability to repress or activate mRNA targets, highlighting its role in translational activation by CPEB[9,10] (Figure 4).

Finally, by visualizing neuronal subregions harboring CPEB3, we begin to define the ultra-structures associated its prion-like state. These structures include MVBs with yarn-like fibrillar

inclusions and a variety of intralumenal vesicles, similar to early-stage melanosomes containing Pmel17[22], a functional amyloid glycoprotein involved in melanin pigmentation. Lamellae in CPEB3 expressing cells also showed distinctive compartments varying in size yet all containing cavernous interiors occasionally trapping unidentified densities. These compartments were often large enough that our sampling within lamellae did not allow the boundary of caverns and cavities to be fully visualized, making it difficult to discern whether these compartments were open to the cytoplasmic environment. These compartments were present in a majority of U87 cells expressing CPEB3 and might be associated with a stress response, akin to the lipidic structures reported in stressed yeast[26] and HeLa cells[27]. This is consistent with our finding that prolonged exogenous expression of CPEB3 reduces cell viability and alters transcriptional profiles to elicit immune and inflammatory responses. Importantly, these deleterious effects of CPEB3 overexpression were less evident in cells expressing CPEB3$^{\Delta AC}$-GFP, indicating that CPEB3$_{AC}$ is a defining feature of CPEB3 associated with cellular disruption.

The formation of ordered prion-like aggregates by the CPEB3 PRD1 is facilitated by CPEB3$_{AC}$, a segment that appears to influence both its translational regulation in cells and its detrimental impact on cell viability. This paints CPEB3$_{AC}$ as a double edged sword that requires close regulation in cells to avoid potential pathology. In fact, our collective observations of CPEB3 support a model by which the amyloid core of CPEB3 is an essential liability, a segment required for CPEB3 function but also associated with CPEB3-induced cellular stress. The localization of CPEB3 to p-bodies is sensitive to perturbations of CPEB3$_{AC}$ and in its absence, CPEB3 localizes to stress granules. Based on these observations, CPEB3$_{AC}$ may act as a regulatory element or may itself be subject to post-translational modifications such as SUMOylation[13] and ubiquitination[11]. Perhaps regulation of CPEB3$_{AC}$ in turn dictates the transition of CPEB3 into and out of its prion-like state to limit its formation of large fibrillar inclusion bodies, as observed for pathogenic protein aggregates[25,38]. Our studies illuminate the

capacity of CPEB3 to form highly labile amyloid fibrils *in vitro*, and importantly, reveals structural features associated with *in situ* regulation of functional prion-like proteins in neuronal cells.

Limitations of this study

The structural data we present are drawn primarily from constructs focusing on CPEB3$^{1-217}$, its first prion-like domain. In contrast, cellular experiments rely on exogenous expression of CPEB3-GFP, rather than endogenously tagged CPEB3 expressed at baseline levels. Further work is needed to fully understand the role of native CPEB3 in primary hippocampal neurons, and to further characterize structural changes occurring as it transitions from basal to its stimulated states. While we found that over time, cells expressing CPEB3-GFP grew large puncta, these same cells suffered in viability at later timepoints, indicating a degree of toxicity elicited by prolonged overexpression of the protein. Thus, our experiments highlight the need for close regulation of CPEB expression levels, and its aggregation and localization in cells.

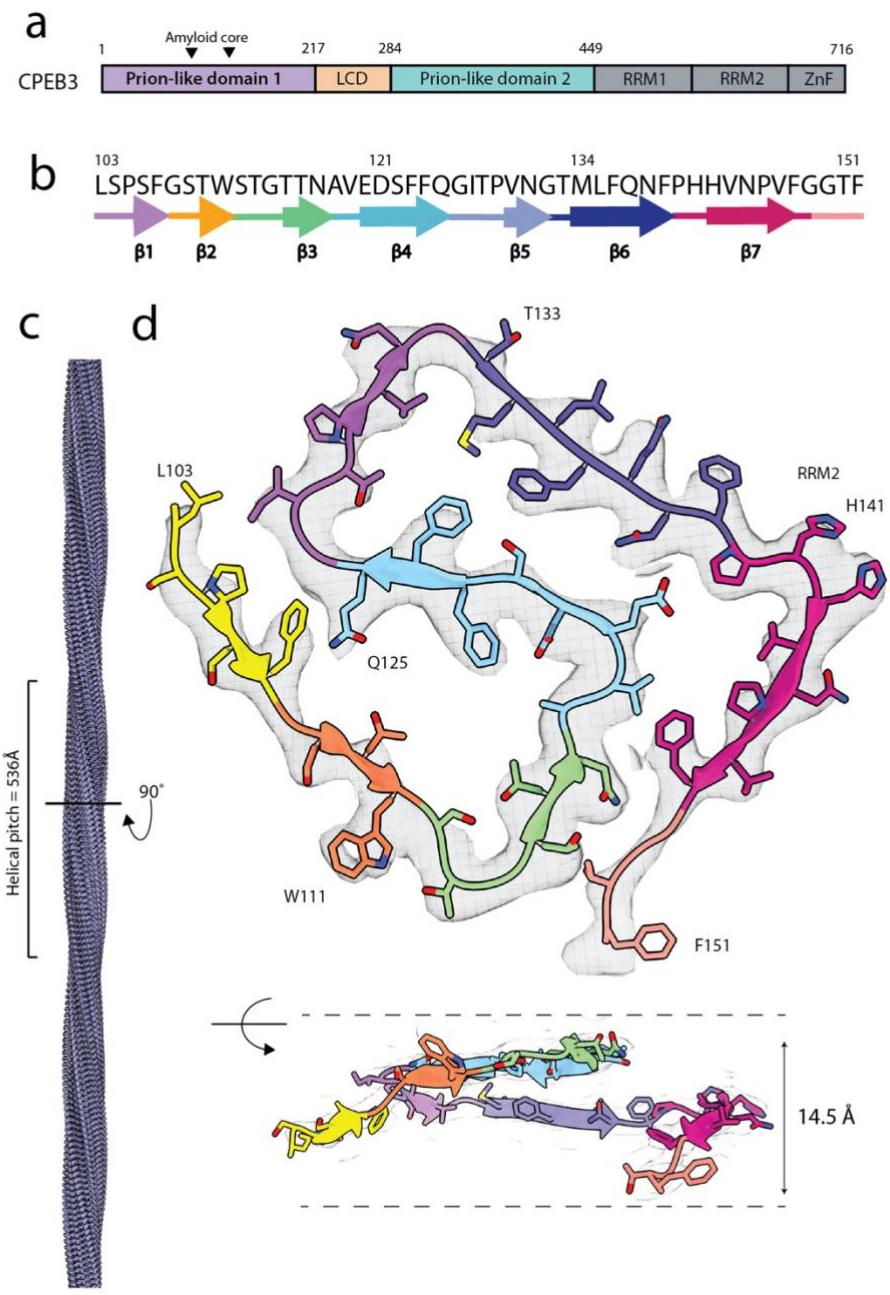


Figure 1. *CryoEM structure of fibrils formed by CPEB3 prion-like domain 1.* **a,** Schematic summary of CPEB3 domains. **b,** Amino acid sequence alignment of fibril core corresponding to secondary structure β-strands. **c,** Model of CPEB3AC filaments with a corresponding helical pitch of 536 Å. **d,** Top: cryoEM map of cross section perpendicular to the helical axis with model rendering colored to match secondary structure schematic. Bottom: side view of map and model measuring height different within one helical layer.

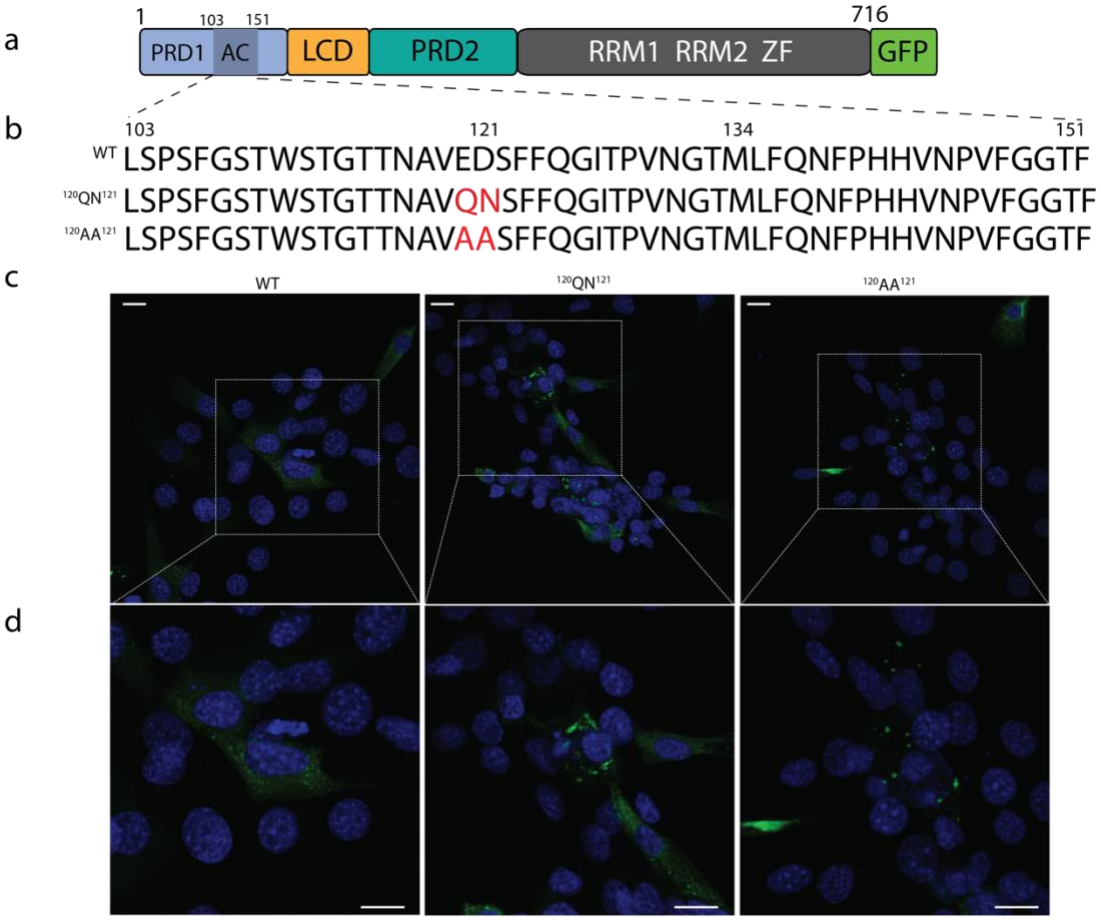


Figure 2. *The role of E120 and D121 in CPEB3 puncta formation in cells***.** Wild type CPEB3-GFP shows fine, granular like aggregates in HT22 mouse hippocampal neurons after 24 hours of expression. **a,** Schematic of full-length WT construct expressed in cells. **b,** Sequence alignment of AC region in FL constructs with corresponding mutated residues in red. **c,** CPEB3-GFP WT and mutants (green) and nuclear staining (blue) in HT22 neurons. **d**, insets of panel c. Scale bar = 10 µM.

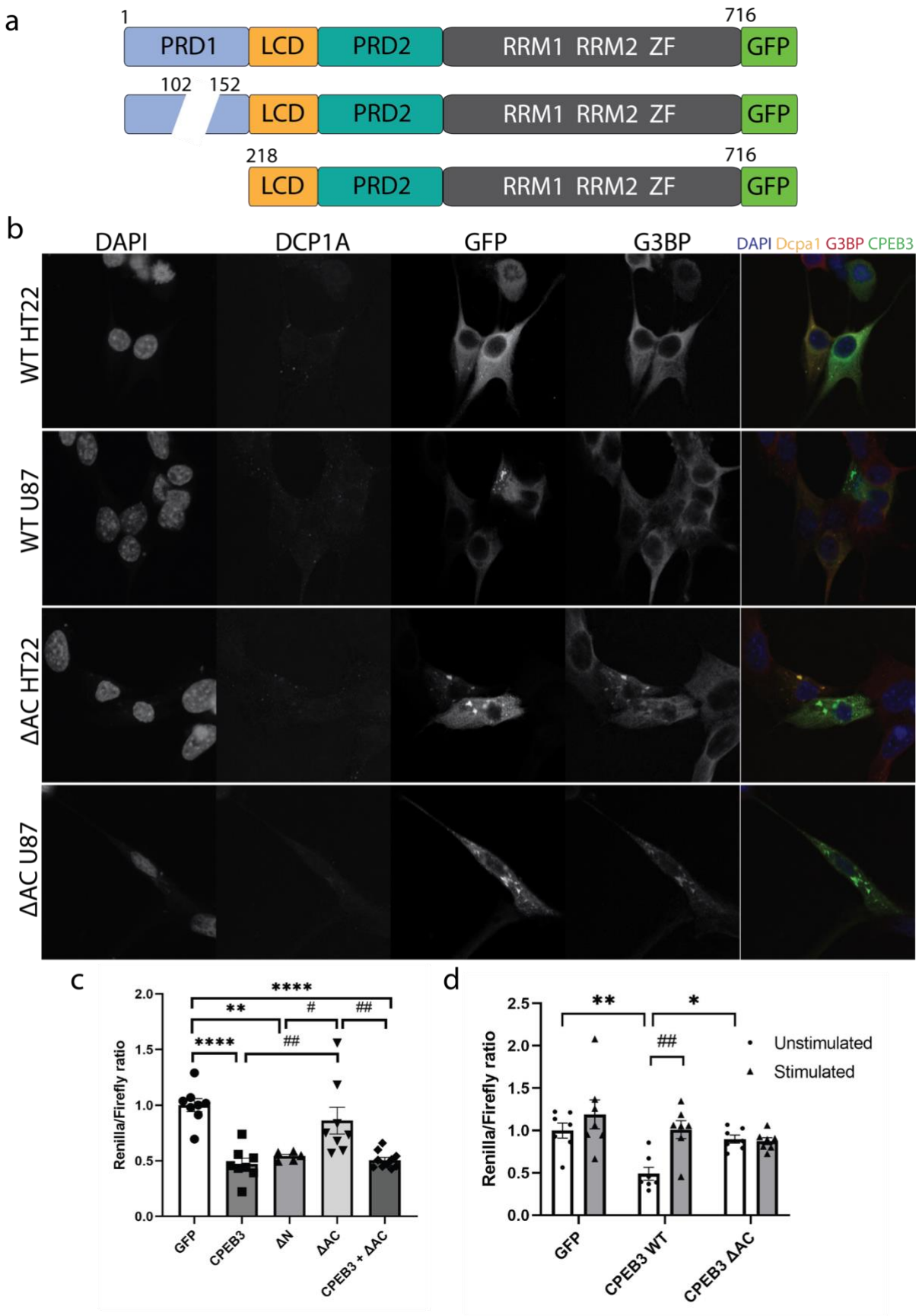

a
1
716
PRD1
LCD
PRD2
RRM1 RRM2 ZF
GFP
102
152
218
b
DAPI
DCP1A
GFP
G3BP
DAPI Dcpa1 G3BP CPEB3
WT HT22
WT U87
ΔAC HT22
ΔAC U87
c
Renilla/Firefly ratio
GFP
CPEB3
ΔN
ΔAC
CPEB3 + ΔAC
d
Renilla/Firefly ratio
Unstimulated
Stimulated
GFP
CPEB3 WT
CPEB3 ΔAC

Figure 3. *$CPEB3_{AC}$ regulates subcellular location of GFP-tagged CPEB3 in cells and affects its translational regulation.* **a,** Schematic representation of full-length constructs used in localization and translational assays. **b,** HT22 and U87 cells expressing WT (top) and AC (bottom) (green) and subsequent co-localization with antiDcp1a (yellow) and antiG3BP (red) with nuclear stain (blue). **c,** HEK 293T cells inhibitory renilla luciferase assays (n =5-11 replicates per construct).) Graphs show the quantification of renilla/Firefly ratio, mean ± s.e.m. One-way ANOVA, followed by Tukey's multiple comparisons post hoc test, ** p<0.01, ****p<0.0001 vs GFP; # p< 0.05, ## p<0.01, vs $CPEB3^{\Delta AC}$. **d,** Renilla luciferase assays performed in primary neurons. (n=7 per condition). Graphs show the quantification of renilla/Firefly ratio, mean ± s.e.m. Two-way ANOVA, followed by Tukey's multiple comparisons post hoc test, * p< 0.05, ** p<0.01, vs CPEB3 unstimulated; ##p< 0.01 vs CPEB3 stimulated.

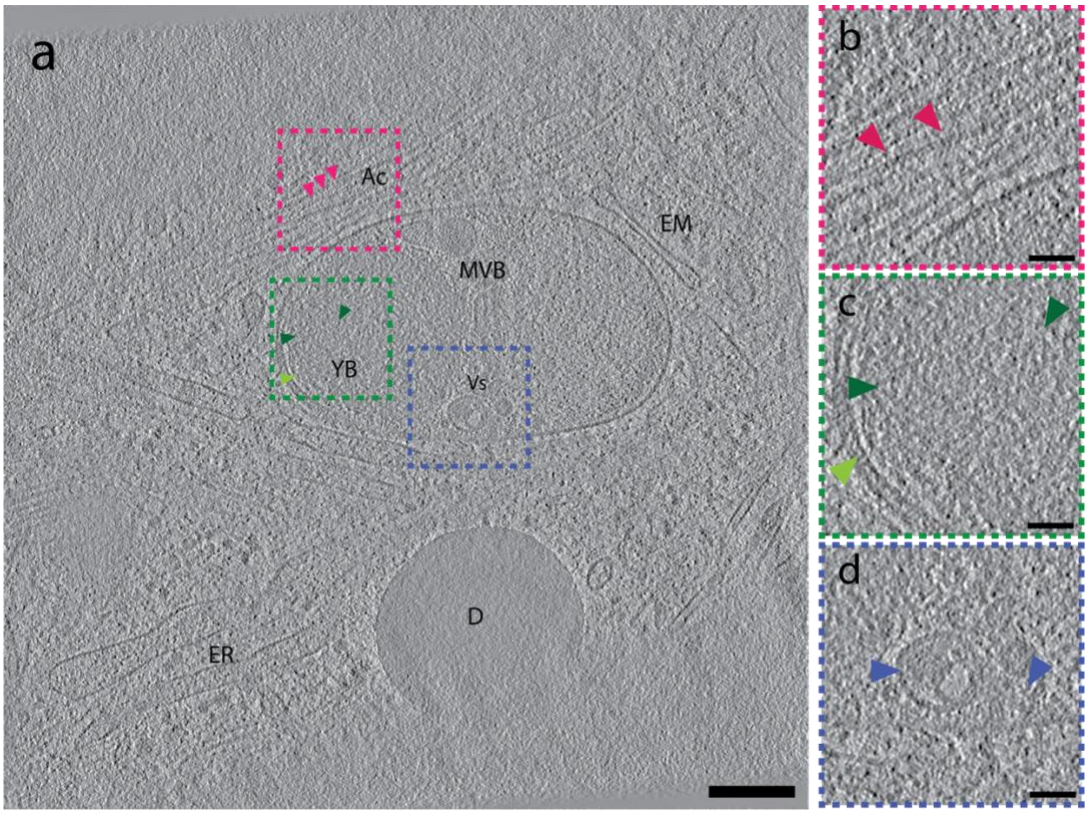


Figure 4. *Cytoplasmic regions containing MVBs harbor filament-like clumps.* **a,** Tomographic slice from U87 cells expressing CPEB3-GFP. Where ER= endoplasmic reticulum, Ac= actin, YB= yarn ball, D= droplet, Vs= vesicle and EM= elongated membrane. Color dashed squares highlighted in b, c, and d. Scale bar = 200 nm. **b,** Zoomed inset from pink box in a, highlighting bundled actin filaments. **c,** Zoomed inset from green box in a, highlighting bundled filament-like yarn balls. **d,** Zoomed inset from blue box in a, highlighting small intralumenal vesicles in MVBs. Scale bar from b, c and d = 50 nm.

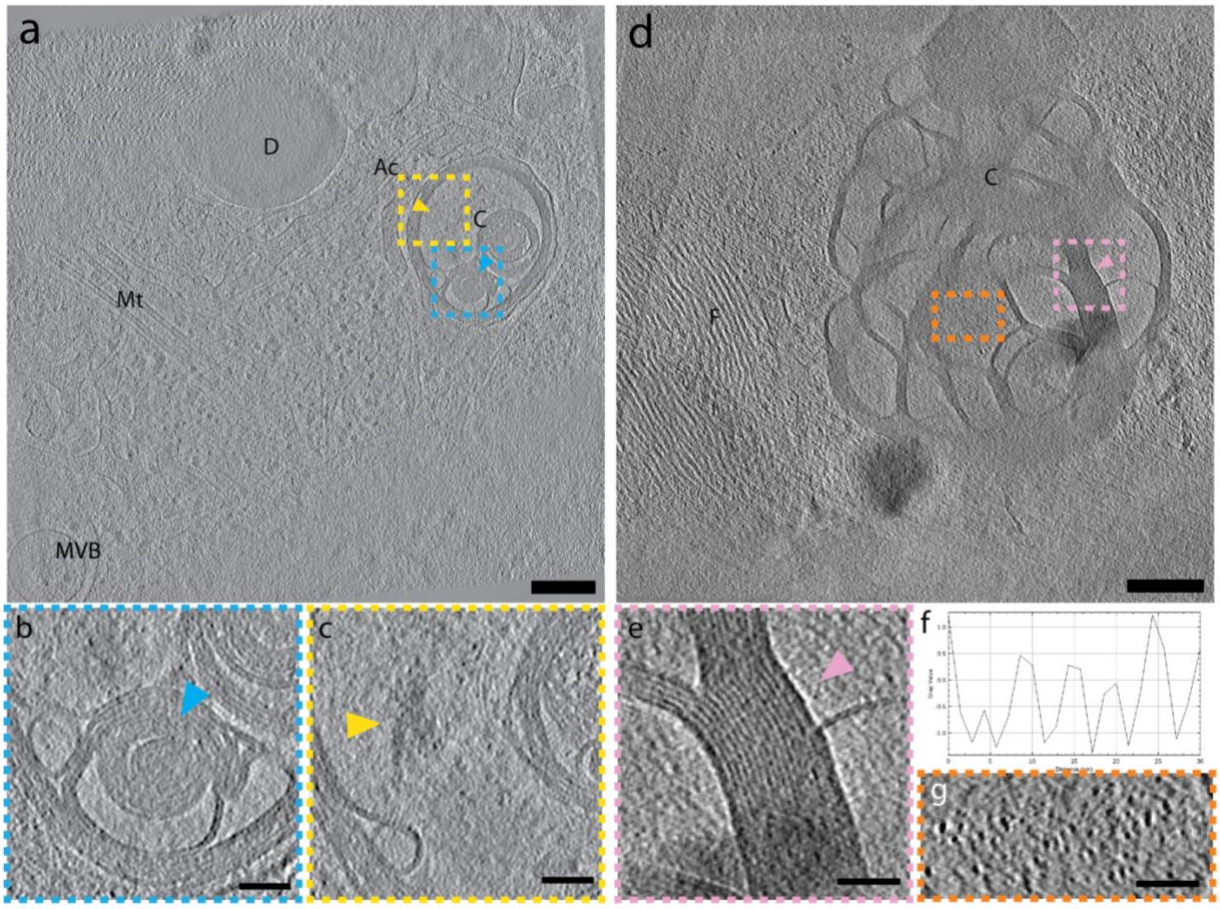


Figure 5. *Ordered layers and cavities make up structured sub-compartments.* **a,** Tomographic slice of cytoplasmic region containing smaller compartments. Where MVB= multivesicular body, Mt= microtubule, D= droplet, Ac= actin, and C= compartment. Scale bar = 200nm. **b,** Zoomed inset from blue box in a, highlighting spiral entwining layers and blending of features. **c,** Zoomed inset from yellow box in a, highlighting unknown densities. **d,** Tomographic slice of cytoplasmic region containing larger, cavernous compartments. Scale bar= 200 nm. **e,** Zoomed inset from pink box in d, highlighting ordered peripheries. Scale bar= 50nm. **f,** Line layer profile of features in e, denoting ~4nm spacing. **g,** Zoomed inset from orange box in d, highlighting unknown densities within caverns. Scale bar= 50nm.

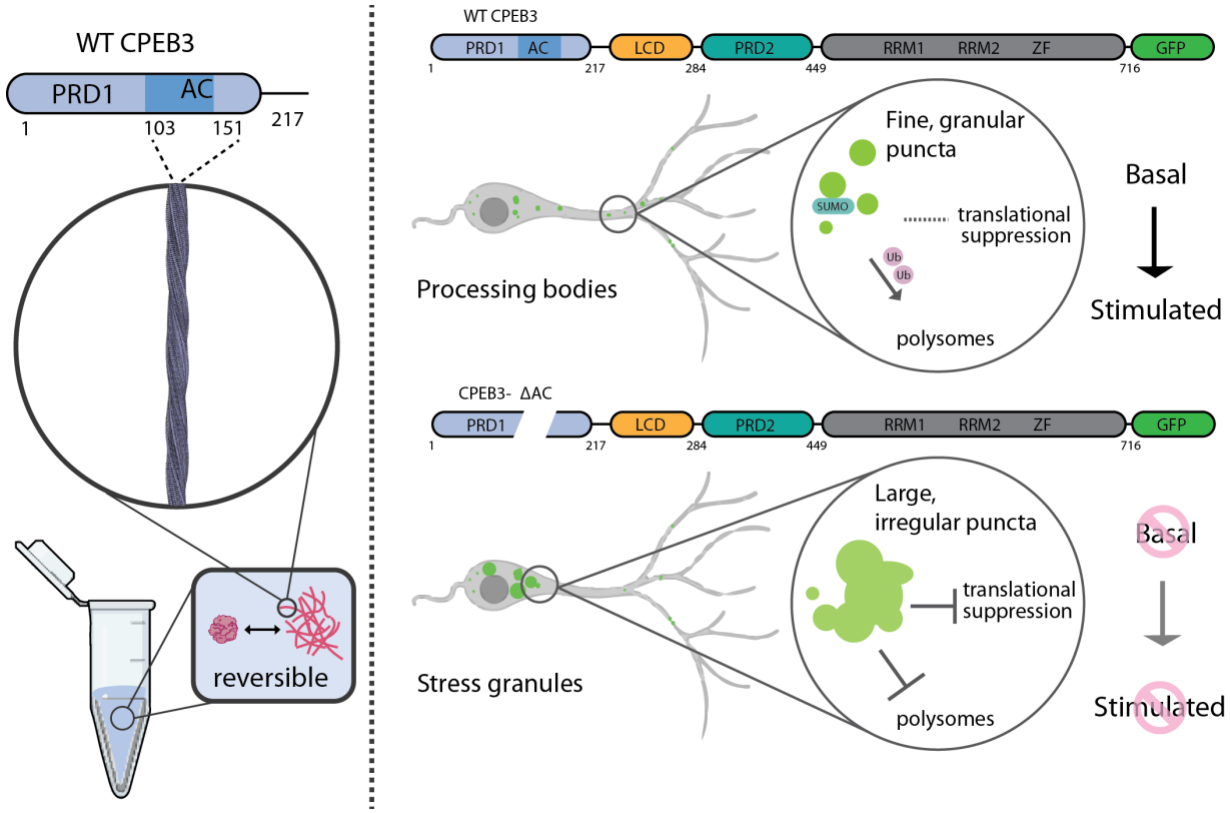
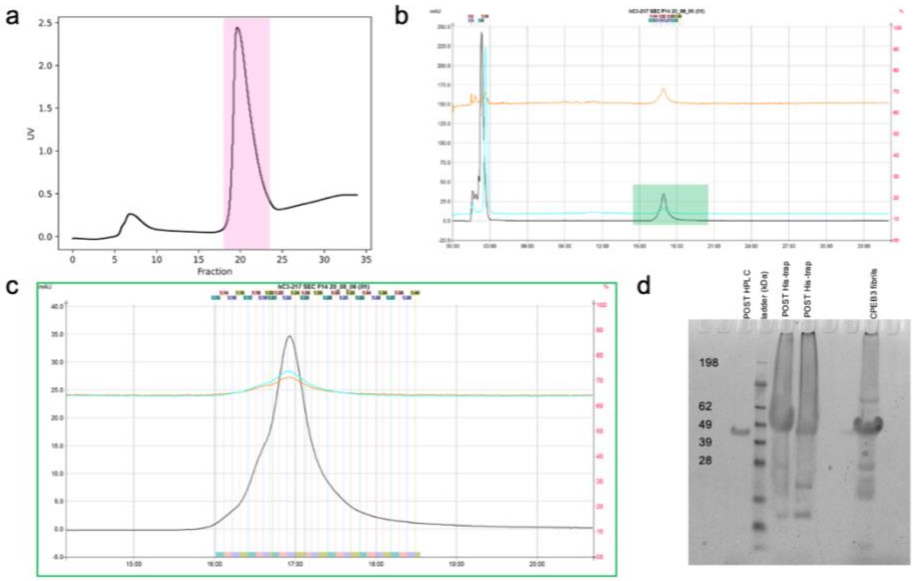


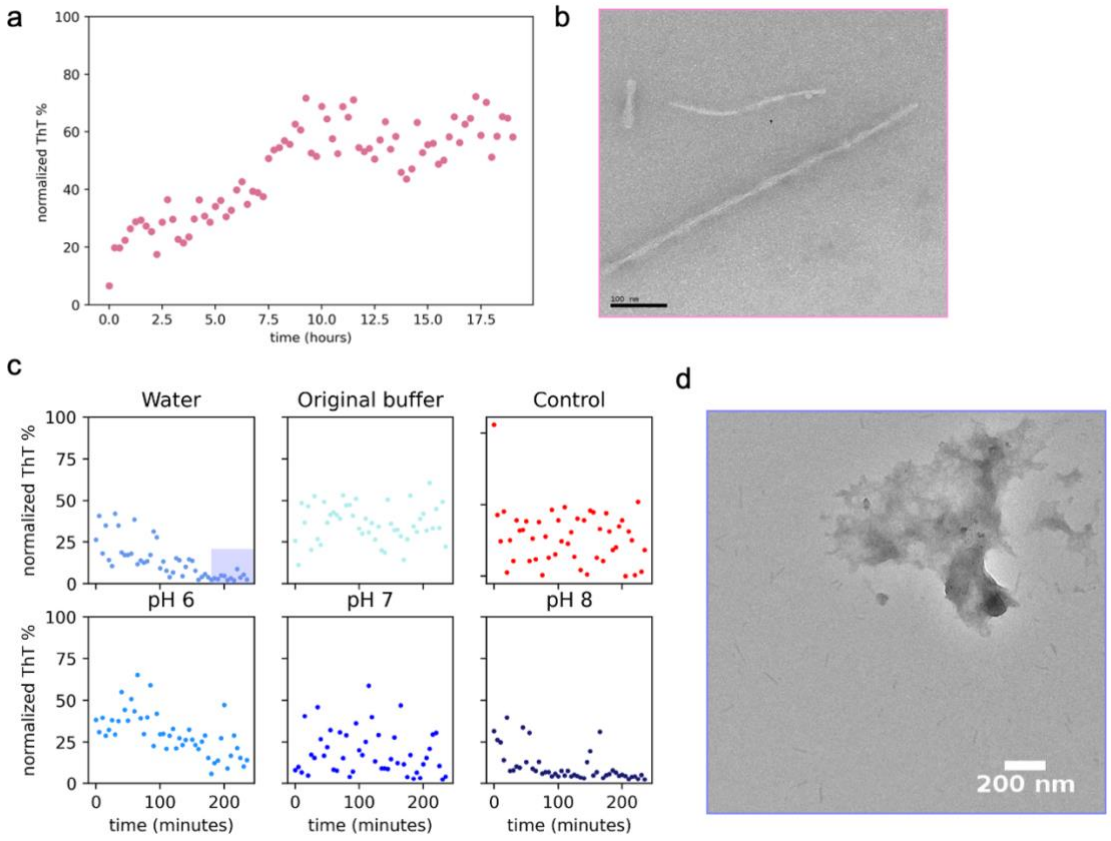
Figure 6. *Model of CPEB3 AC and its role in CPEB3 aggregation and regulation.* (Left panel) Amyloid fibrils formed in-vitro are reversible. Cryo-EM identifies a core sequence within the CPEB3 prion-like domain. (Right panel) WT CPEB3 displays fine, granular puncta that co-localize in p-bodies in the basal state and are moved to polysomes upon stimulation. $CPEB3_{\Delta AC}$ aggregates are larger and co-localize with SG markers, displaying no functional activity upon stimulation.

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

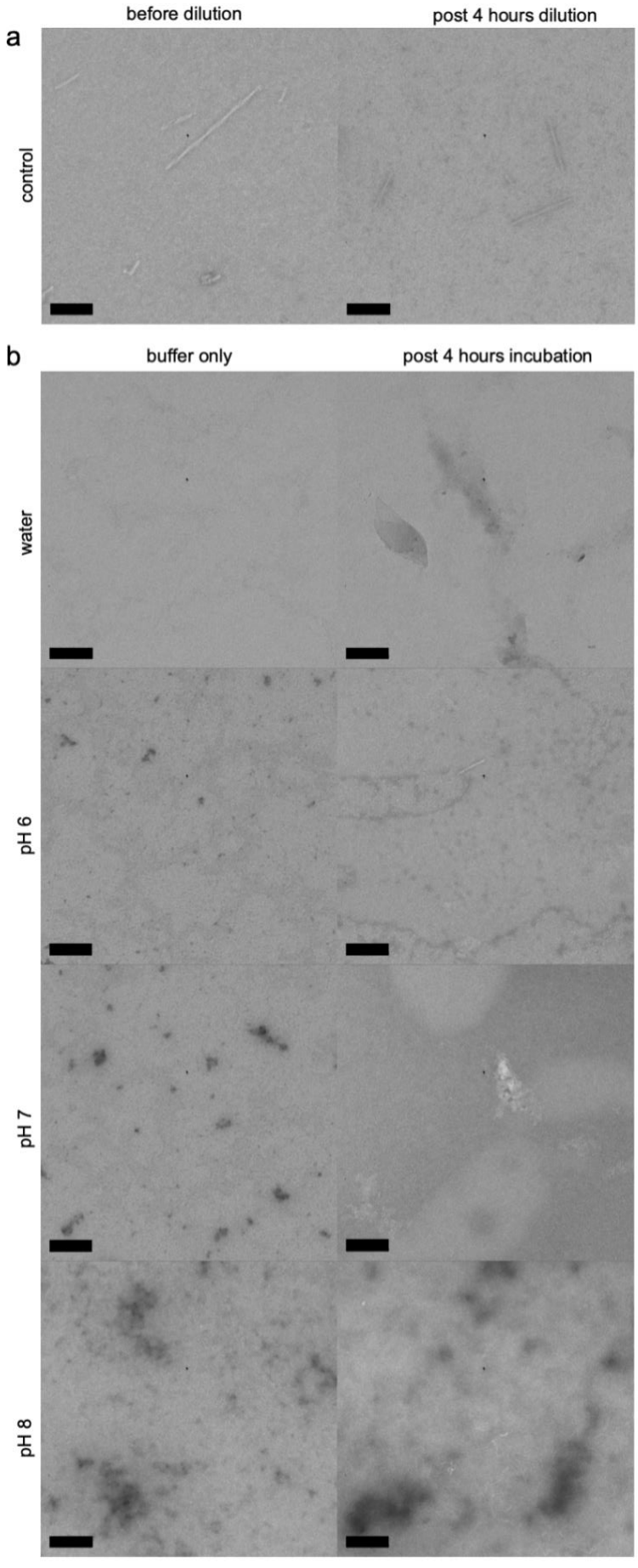
	$CPEB3_{AC}$
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	52
Defocus range (μm)	1-3
Pixel size (Å)	0.86
Symmetry imposed	C1
Helical rise (Å)	4.8
Helical twist (°)	-3.32
Initial particle images (no.)	2,410,032
Final particle images (no.)	40,329
Map resolution (Å)	3.0
FSC threshold	0.143
Map resolution range (Å)	200-3.0
Refinement	
Initial model used (PDB code)	De novo
Model resolution (Å) FSC threshold	0.143
Model resolution range (Å)	200-3.0
Map sharpening *B* factor (Å^2)	- 100.35
Model composition	
Non-hydrogen atoms	374
Protein residues	49
Ligands	0
B factors (Å^2)	
Protein	
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.753
Validation	
MolProbity score	1.79
Clashscore	9.26
Ramachandran plot	
Favored (%)	95.74
Allowed (%)	4.26



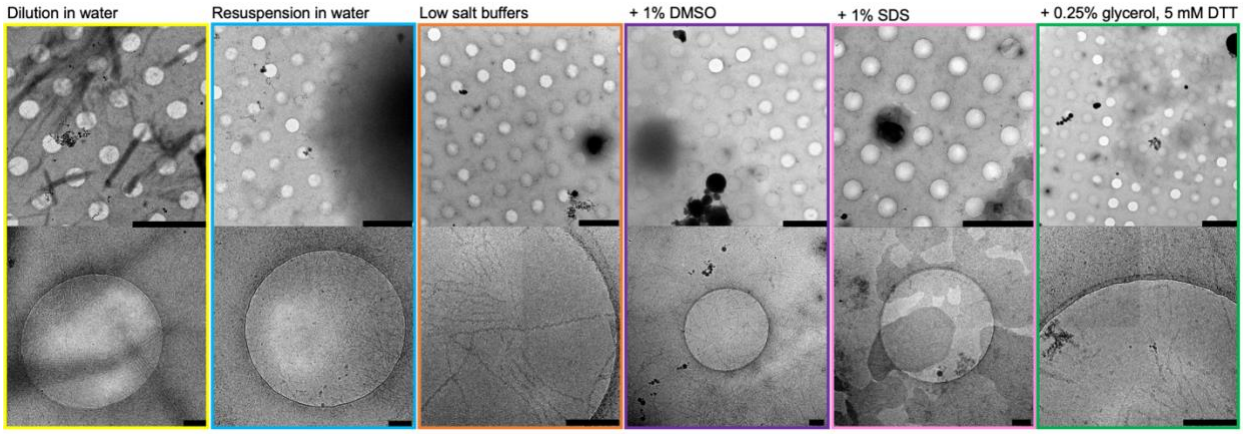
Supplemental Figure 1. Recombinant protein purification. **a,** Example His-trap elution profile. CPEB3$^{1-217}$ eluted in fractions 18-23 (pink subset). **b,** Full view of purification chromatogram. Three wavelengths were collected to monitor the purification: 214nm (black), 280 nm (orange), and 254 nm (light blue). **c,** Zoomed purification chromatogram (green) where the target protein started eluting. Three wavelengths were collected to monitor the purification: 214nm (black), 280 nm (orange), and 254 nm (light blue). **d,** SDS-PAGE shows harvested fractions after His-trap and HPLC purification. Freshly harvested fibrils also present.



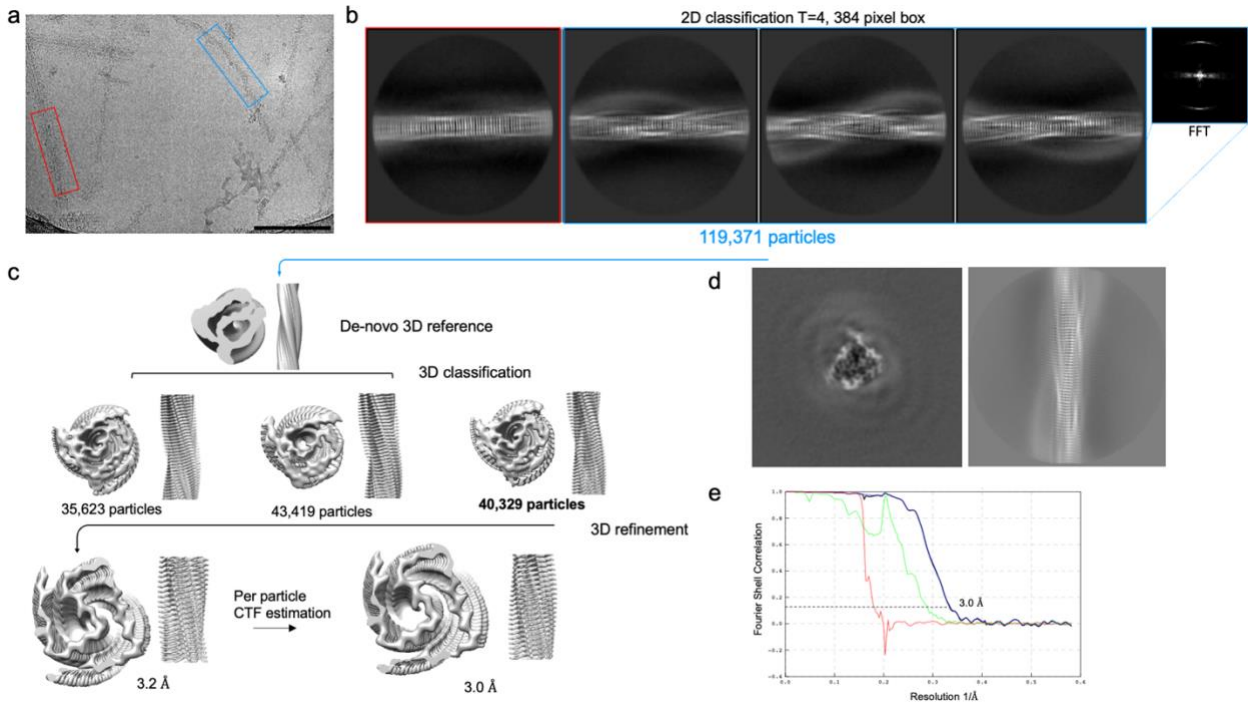
Supplemental Figure 2. Biochemical characterization of CPEB3$^{1\text{-}217}$ segment. **a**, ThT curves of CPEB3$^{1\text{-}217}$. Data shown are normalized triplicate conditions, n=3. **b,** Negative stain TEM images of fibrils formed by CPEB3$^{1\text{-}217}$. **c,** Dissolution assays of pre-formed fibrils. Samples were harvested after formation from SF 2a**,** and diluted 2-fold in corresponding buffers for approximately 4 hours. Data shown n= 4 wells per condition. **d,** Negative stain TEM image of fibril sample after dilution in water for 4 hours (purple subset in SF 2c).


a
before dilution
post 4 hours dilution
control
b
buffer only
post 4 hours incubation
water
pH 6
pH 7
pH 8

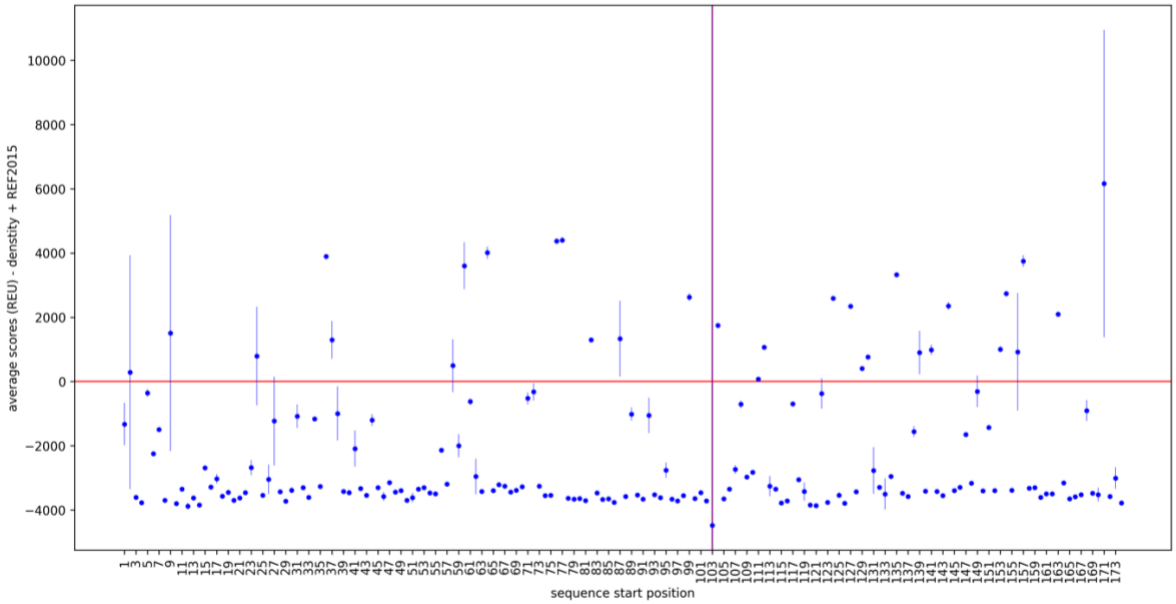
Supplemental Figure 3. Negative stain TEM micrographs of CPEB3 fibrils after dissolution assays. Samples harvested from dissolution assays in Supplemental Figure 1d were prepped for TEM analysis. **a,** Control wells were diluted in original fibrilization buffer. **b,** Representative micrographs of blanks corresponding to blank, dilution buffer (left column), and protein samples after a 4-hour incubation in designated buffer (right column). Scale bar = 200 nm.



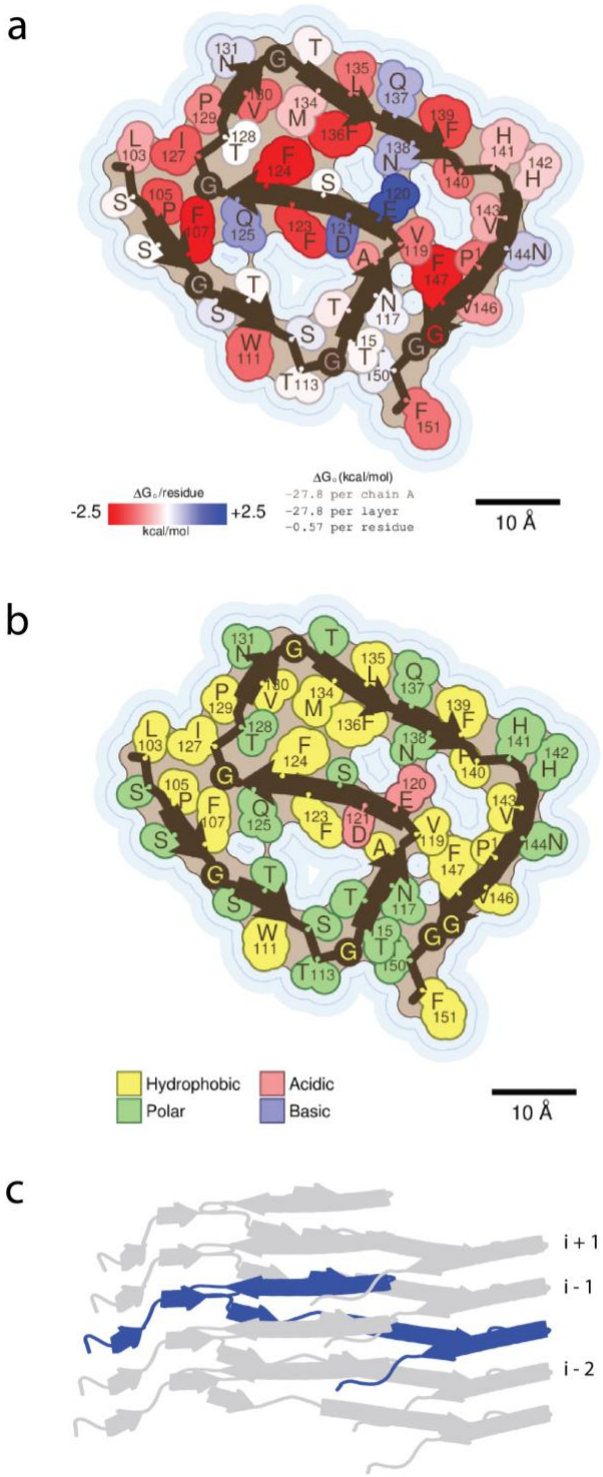
Supplemental Figure 4. CryoEM grid optimization. Reversible fibrils were plunge frozen in various buffers for optimal particle dispersion and minimum interactions at the air-water interface (AWI), representative micrographs shown. Yellow and blue: dilution and resuspension in water caused significant clumping and dissociation. Orange: low salt buffers < 100 mM NaCl cause denaturation at the air water interface. Purple and pink: addition of DMSO and SDS, respectively, reduced clumping but impacted ice and fibril integrity. Green: final freezing condition 0.25% glycerol and 5 mM DTT resulted in acceptable clumping and reduced AWI denaturation. Scale bar top row = 2 μM, scale bar bottom row = 200 nm.



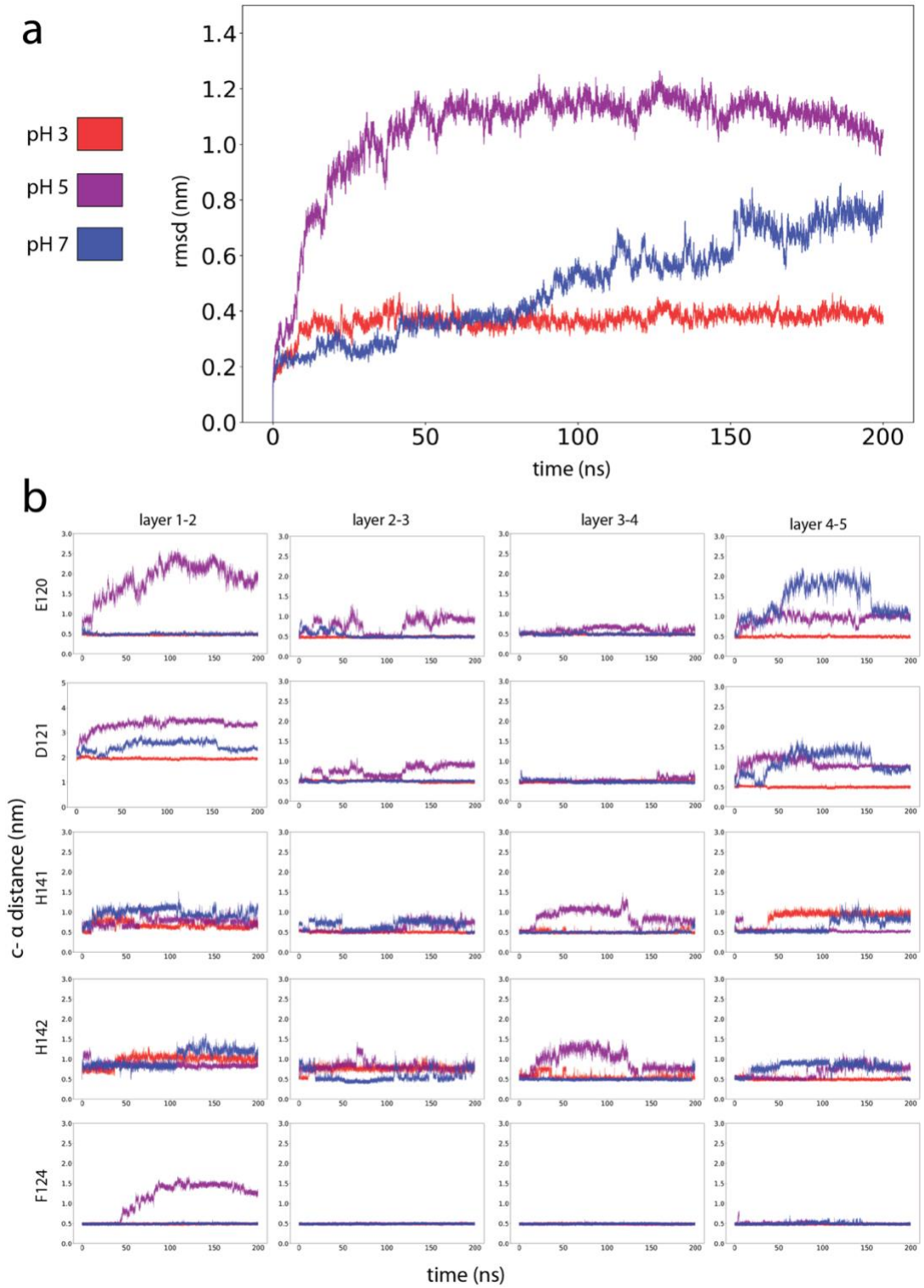
Supplemental Figure 5. CryoEM data processing and model comparisons. **a,** CryoEM micrograph of twisting fibril polymorphs (blue) and non-twisting (red) Scale bar = 50 nm. **b,** Representative 2D class averages of polymorphs pictured in SF1a (left) and 4.8 Å reflection pattern in generated Fourier transform of 2D class (right). **c,** Schematic representation of helical reconstruction of major twisting species performed in RELION. **d,** Central slice from final reconstruction (left) and projection of 3D model (right). **e,** FSC curves between independently refined half-maps masked and corrected (black), unmasked and corrected (green), masked (navy blue) and phase randomized (red). Black dotted line indicates the FSC=0.143 resolution of 3.0 Å.



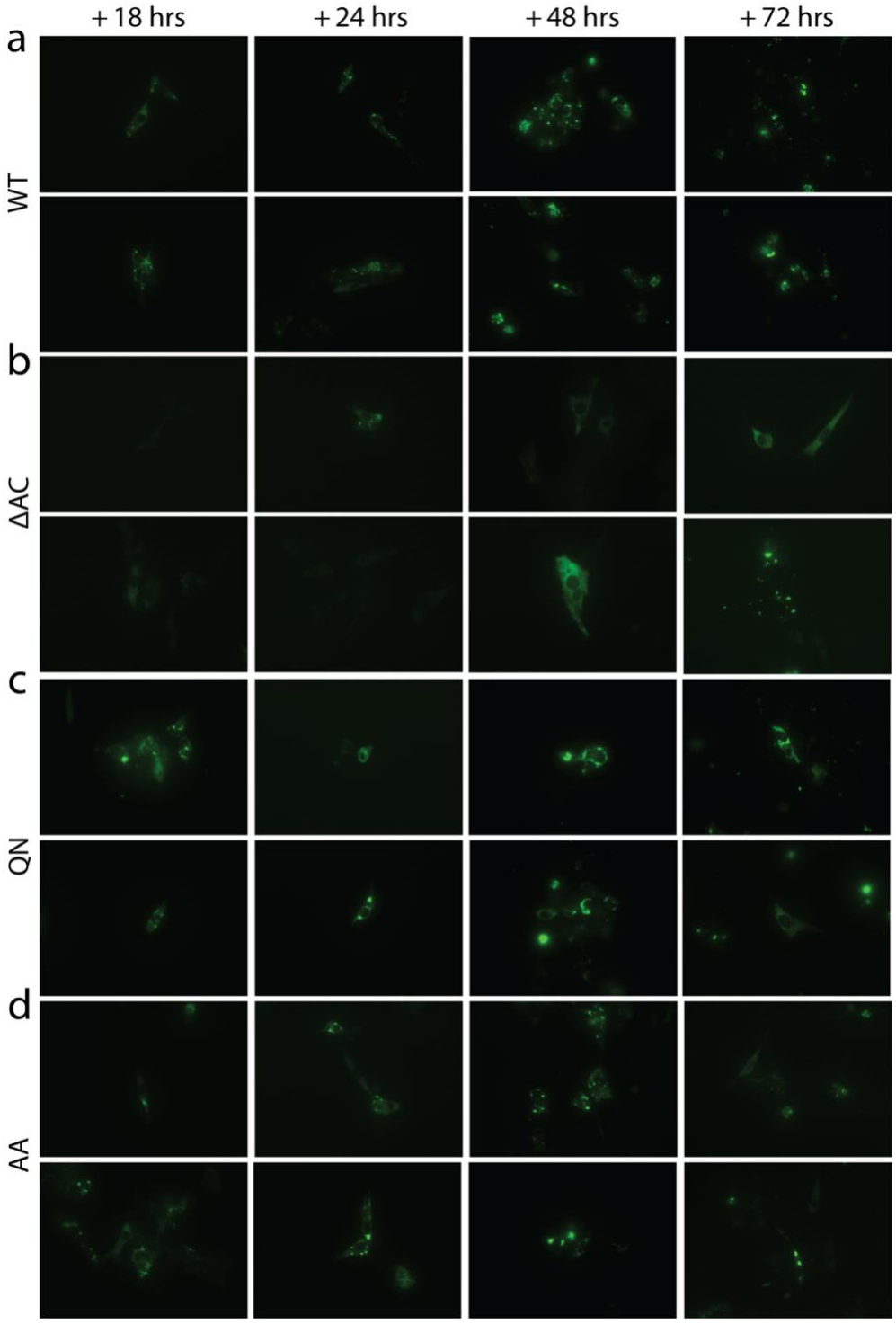
Supplemental Figure 6. Average scores of threading experiment plotted against sequence start position. Average scores were calculated with Rosetta's REF2015 energy function combined with a cryo-EM density score term. Each point represents the average of three trials. Sequence start position 103 shows the most favorable sequence window and is in agreement with the modeled sequence.



Supplemental Figure 7. *Structural features of CPEB3$_{AC}$ that may influence its stability* **a,** Energy per residue stabilization, key lower left. Thin blue lines represent the closest approach that the center of a water molecule can achieve. b, **c,** Cartoon rendering secondary structure elements and highlighting the height differences within a single chain that interact with 3 consecutive layers.

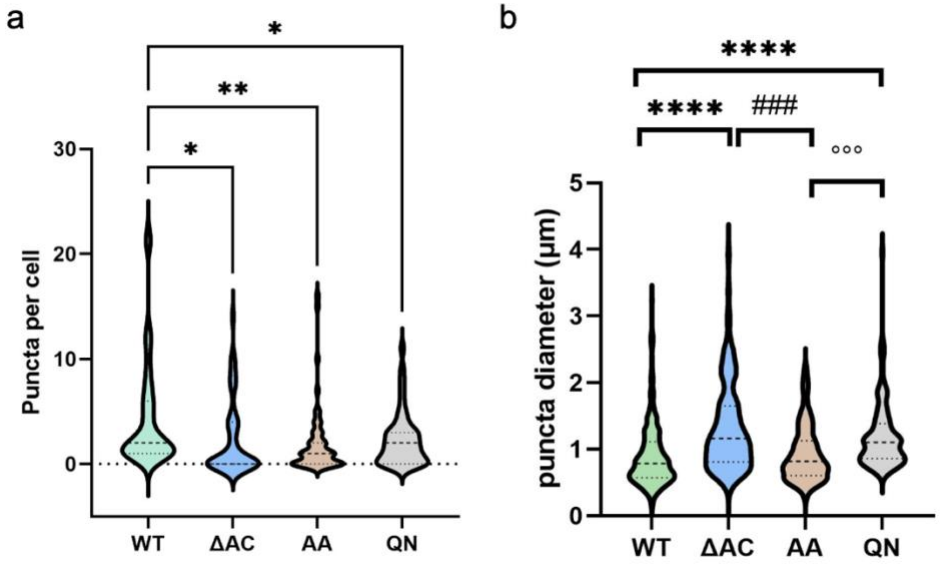


Supplemental Figure 8. Molecular dynamics simulations of pH triggered instabilities in fibril core. pH color legend where pH 3 (red), pH 5 (purple) and pH 7 (blue). **a,** Averaged RMSD for all layers across 200 ns **b,** Selected charged residues from top to bottom: E120, D121, H141 and H142. Control residue F124. From left to right: neighboring layer c-α distances across 200 ns simulations.

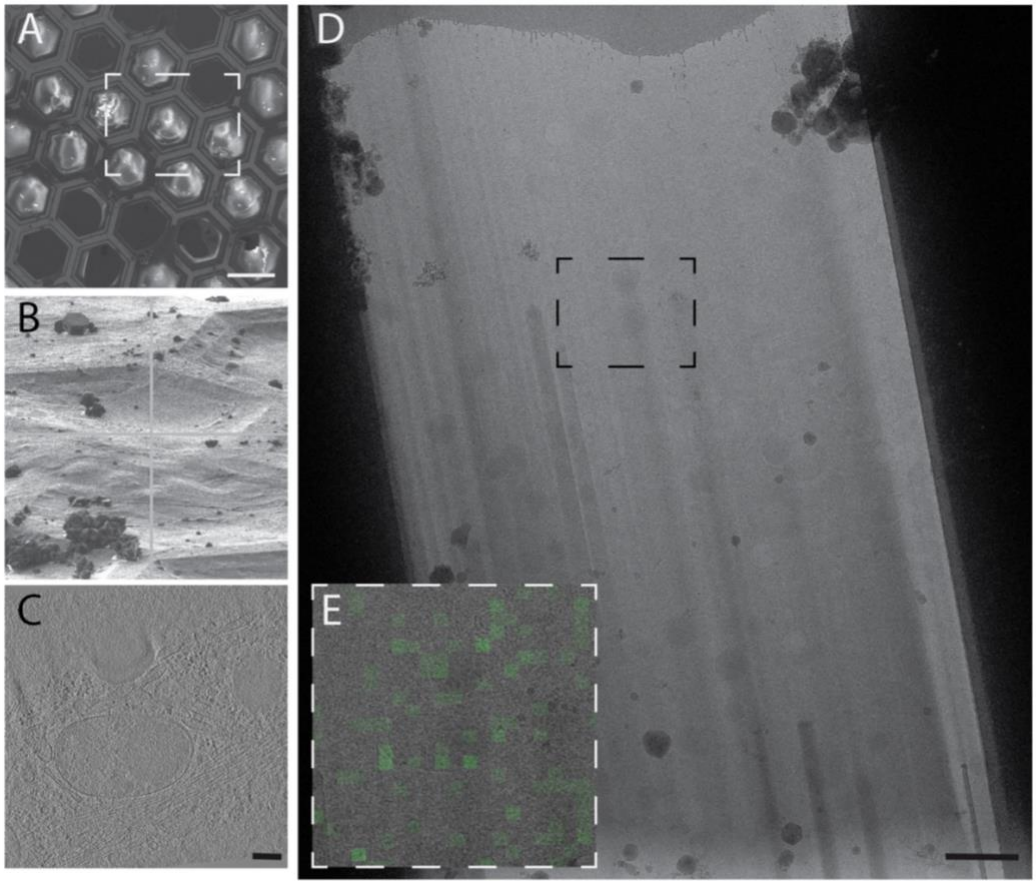


Supplemental Figure 9. Aggregation phenotypes of CPEB3 WT and ΔAC constructs in HT22 in GFP channel (green). **a,** Representative images of WT CPEB3-GFP expression across four time points 18, 24, 48 and, 72 hrs. **b,** Representative images of CPEB3$^{\Delta AC}$-GFP expression across four time points 18, 24, 48 and, 72 hrs with earlier time points displaying significantly diffuse phenotypes. **c,** Representative images of CPEB3-^{120QN121}-GFP across four time points

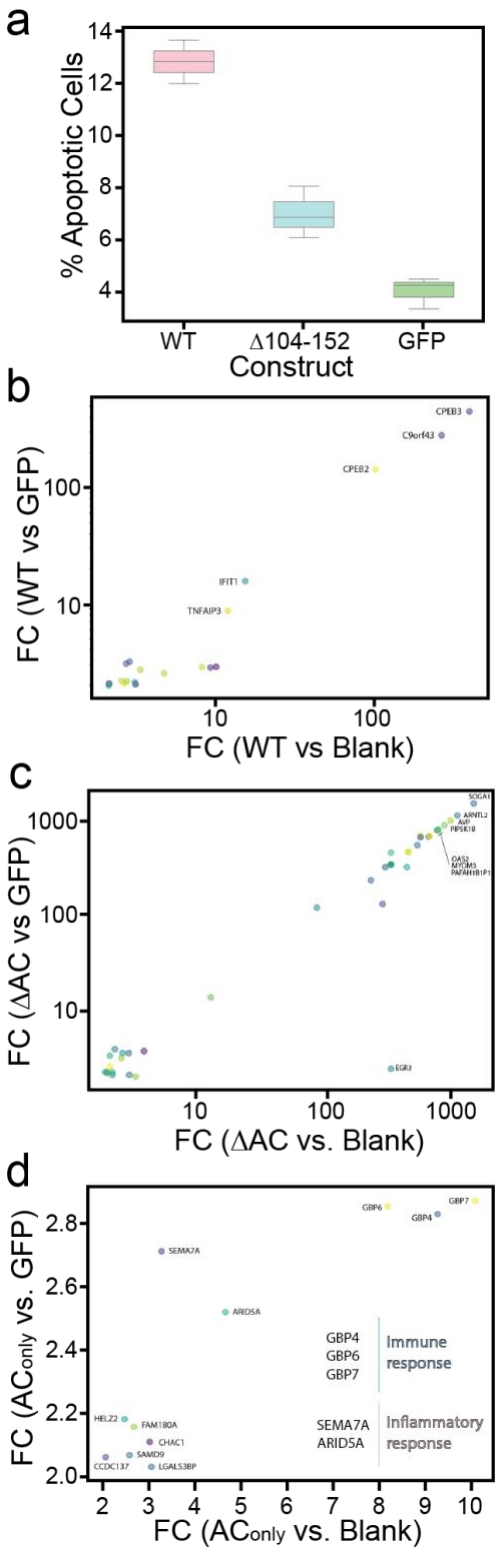
18, 24, 48 and, 72 hrs. **d,** Representative images of CPEB3^{120AA121}-GFP expression across four time points 18, 24, 48 and, 72 hrs. All time points indicate time after incubation with lipofectamine and corresponding vector.



Supplemental Figure 10. CPEB3 mutant phenotypes. **a,** Quantification of puncta per cell after 24 hours of exogenous expression. Where WT-GFP n=60 cells, CPEB3$^{\Delta AC}$-GFP n=52 cells, CPEB3-^{120QN121}-GFP n=56 cells and CPEB3^{120AA121}-GFP n=50 cells. One-way ANOVA, followed by Tukey's multiple comparisons post hoc test, * p<0.05, ** p<0.01, vs WT. **b,** Quantification of average puncta diameter from cells surveyed in **a.** One-way ANOVA, followed by Tukey's multiple comparisons post hoc test, **** p<0.0001, vs WT, ### p<0.001, vs dAC, °°° p<0.001 vs QN.



Supplementary Figure 11. Cryo-FIB, Correlative Light and Electron Microscopy (CLEM) and Cryo-ET workflow. **a,** SEM micrograph of cell targeted for FIB-milling. Scale bar = **b,** Side view of same area boxed in a, from view of the ion beam. **c,** Scale bar = 50nm. Slice from reconstructed tomogram acquired from lamella pictured in **d,** and depicted in **e,** overlayed with maximum intensity projection. Scale bar = 1uM.



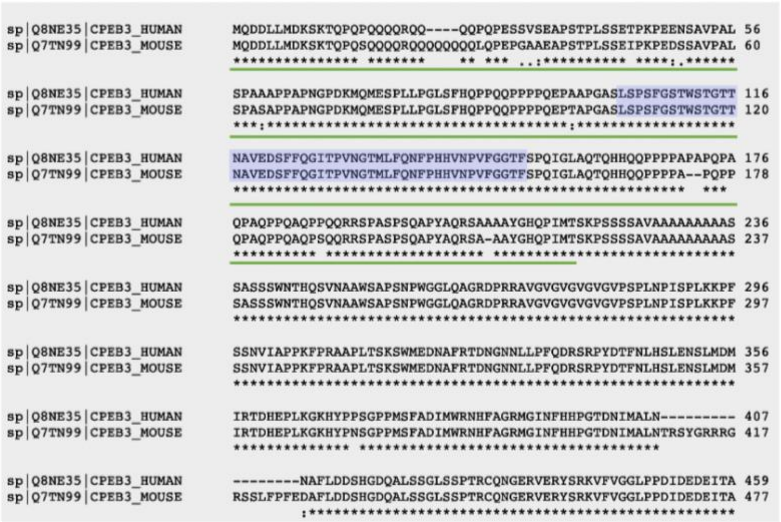
Supplementary Figure 12. *CPEB3 viability assays and RNA sequencing.* **a,** Box plot representing n=3 wells per sample condition with y-axis representing % apoptotic rate. **b,** Fold change (FC) of significant gene transcripts identified in HT22 cells expressing CPEB3-GFP. Plot contains normalized values against both GFP and untreated (blank) controls. **c,** Identical to plot b, but for HT22 cells expressing CPEB3$^{\Delta AC}$-GFP. **d,** FC of gene transcripts found

unique to WT expressing cells from cross comparisons of b, and c, plotted relative to each control. Top five hits highlighted by immune and inflammatory categories.

Materials and Methods

Construct design

Design of recombinant CPEB3 segment included previously reported residues critical for aggregate formation and prion-like behavior in cells[10]. CPEB3 residues 1-217 were inserted into pET—24(+) vector from TWIST biosciences. Sequences were verified via DNA sequencing (Genewiz). CPEB3-EGFP constructs were a kind gift from Luana Fioriti.



Panel 1: Sequence alignment of CPEB3 human and mouse sequences. Green light denotes residues within PRD1 and lilac highlight corresponds to residues placed in amyloid core structure.

Protein expression and purification

Recombinant CPEB3$^{1-217}$ was expressed in *Escherichia coli* BL21 GOLD cells. Briefly, cells were grown in LB media with 50 ug/ml kanamycin at 37° C to an OD600 of 0.6-0.8. 1 mM isopropyl B-D-1-thiogalactopyranoside (IPTG) was added for protein expression and cells were cultured an additional 6 hrs at 37° C. Cells were harvested and stored at -20° C until purification. Cell lysates were resuspended in lysis buffer containing 10 mM Tris, 1 mM EDTA at pH 8.2 and supplemented with 1% halt protease inhibitor cocktail (Thermo Scientific), 1% homemade Benzonase and subsequently lysed using 6 consecutive freeze-thaw cycles in liquid nitrogen. Cell lysates were centrifuged (12,000gs for 30min) and pellet was resuspended in 6M Gdn-HCl, 50mM Tris-Base, 25mM imidazole at pH 8 and left gently shaking on ice overnight. Lysates were centrifuged (3,000gs for 30 min) and supernatant was filtered and loaded onto an equilibrated HisTrap HP column (GE) using NGC (Biorad). The HisTrap was pre-equilibrated with 6M Gdn-HCl, 50mM Tris-Base and 25 mM imidazole. Protein was eluted with gradient protocol containing 6M Gdn-HCl, 50mM Tris-Base and 500 mM imidazole. Eluted protein was flash frozen and stored at -80° C until further purification. Further purification was executed with RP-HPLC, using an Interchim puriFlash® 4125 Preparative Liquid Chromatography System equipped with an Interchim PT-15C8-F0040 cartridge (C8,15 µm particle size). The eluted protein fractions were defrosted from -80°C and pooled. The pooled sample was then diluted to 1.8 mL with a dilution buffer consisting of 6 M Gdn-HCl, 50 µM NaH 2 PO 4, 10 µM Tris, and 20 µM imidazole. Before injection the 2-mL sample loop was washed with the above dilution buffer

to maintain high concentrations of guanidine to minimize protein aggregation. [A] consisted of a 99 to 1 water to acetonitrile solution with 15 mM NH 4 OH added as a buffer. [B] consisted of a 40 to 60 water to acetonitrile solution with 15 mM NH 4 OH. Flow rate was set at 40 mL/min. Before injection, the system was equilibrated to 100% [A]. The protein sample was loaded onto the loop and injected. Three wavelengths were used to monitor the purification: 214nm (black), 280 nm (orange), and 254 nm (light blue). The gradient was held at 100% [A] for 5 minutes then a gradual increase to 100% [B] was executed over a span of 25 minutes (Δ5% per minute). The target protein started eluting at 16 minutes.

Fibril preparation and negative stain TEM

Lyophilized protein powder was resuspended in H2O at 1.2 mg/mL concentration, aliquoted and immediately flash frozen and stored at -20°C until further use. Various buffers containing urea and guanidine were initially screened for fibril formation to attempt to control clumping and aggregation. These aggregates were heterogenous and difficult to reproduce. 96-well screen ranging in salt concentrations and pH were used for subsequent attempts without denaturants. The optimized fibril growth condition was incubation in 125mM NaCl, 50mM Tris-Base, 10mM K2PO4 and 5mM glutamic acid ~pH 4.5 and left shaking overnight on the acoustic shaker. Fibrils were collected by scraping and pooling solution from wells and immediately prepared for subsequent experiments. Negative-stain transmission EM samples were prepped by applying 3 µl of fibril solution to glow-discharged 300 mesh carbon-coated formvar support films mounted on Cu grids (Ted Pella). The samples were wicked, washed briefly with H2O, stained with 2% uranyl acetate for 90 seconds and wicked again, allowing to air dry for 3 minutes. Grids were imaged on a T12 (FEI) electron microscope.

Single particle cryo-EM sample preparation and data collection

Freshly harvested fibril solution was diluted 5 fold with buffer containing 100mM NaCl, 10mM Tris-Base, 5mM DTT and 0.25% glycerol. 1.6 µL of diluted fibril solution was applied to each side of a glow-discharged Ultrathin carbon on Quantifoil 1.2/1.3 300 mesh on Au grid (Ted Pella) and plunge frozen into liquid ethane using a Vitrobot Mark IV (FEI) set at 100% humidity, 4°C with a blot force of -1 and blot time of 1.5 seconds. Data were collected on a Titan Krios G3i microscope with a K3 direct detection camera (Gatan) operated at an accelerating voltage of 300 keV and a 20eV slit width (BioQuantum). Automated data acquisition were performed using EPU 2.8 software (Thermo Scientific). Images were acquired with a nominal magnified pixel size of 0.86 Å/pixel, a total of 40 frames with a dose of 1.3 electrons per $Å^2$ per frame resulting in a total dose of 52 electrons per $Å^2$.

Data processing and helical reconstruction

CTF estimation were performed with CTFFIND 4.1.8[39]. Drift-correction and dose-weighting were performed using MotionCor2 implemented in RELION 3.1[40]. Particles were picked with the automatic particle picking software crYOLO[41]. All subsequent image processing and helical reconstruction were performed in RELION as previously described.[33,42] Briefly, fibril segments were initially extracted with a 90% overlap into 512 pixel boxes binned by 2 (1.72 A/pix) and subjected to reference-free 2D classification using a T=2 regularization parameter. Segments belonging to suboptimal 2D averages were discarded and homogenous subsets were selected for further processing. Upon identification of a rapidly twisting species, a smaller box size was chosen. All segments were then extracted with a 90% overlap into 384 pixel boxes for a total of 2,410,032 segments. Segments were split into 10 groups and several rounds of reference free 2D classification were performed to weed out suboptimal segments. Homogenous subsets from each were regrouped and reference free 2D classification were iteratively run decreasing psi_step and offset_step from 8 to 1. After the majority of suboptimal particles had been

discarded an additional round of classification were performed with the tau regularization parameter set to T= 8. Only segments contributing to averages showing clear 4.8 Å signal in corresponding 2D FFTs were selected for subsequent 3D processing, resulting in 119,372 segments. Using both the 512 pixel box and 384 pixel box an initial helical pitch of 536 Å and helical twist of -3.22° were estimated. Starting with these estimates a 3D reference was reconstructed de-novo with 384 pixel 2D averages showing clear β-strand separation along the helical axis. 3D manual refinements were performed using a 30 Å low-pass filtered initial model, K=3 and with manual control of tau_fudge and healpix to reach a resolution of ~ 5Å. Segments contributing to a homogeneous class as defined by stable helicity and separation of β-strands resulted in a final subset of 40,329 segments. Again, a manual 3D refinement was performed on selected segments leading to a reconstruction of 3.5 Å. Several 3D auto-refinements using a 7Å low-pass filtered map from the final manual refinement were run with optimization of helical parameters leading to an estimated helical rise of 4.91 Å and a helical twist of -3.32°. This map was then used for per-particle CTF estimation and a 3D auto-refinement was repeated. The nominal pixel size was adjusted from 0.86 Å to 0.879 Å, resulting in an expected helical rise of 4.8 Å. Post-processing of the map with an extended initial mask of 3 pixels and soft edge of 10 pixels was followed by clipping and centering the map using clip.com. Auto-sharpening using phenix.auto_sharpen at the resolution cutoff indicated by the half-map GSFSC led to a final overall resolution of 3.0 Å. The atomic model was built into the refined map using COOT[43]. We performed automated structure refinement using phenix.real_space_refine (Phenix 1.2)[44].

Threading

Threading was performed with a custom python script utilizing the PyRosetta[45] software package and a sliding window approach. For each window, the experimentally expressed sequence was threaded onto a poly-alanine backbone then energy minimized with the PyRosetta FastRelax with cryo-EM density function. FastRelax used the Ref2015[46] energy function with a modified fa_elec weight to 1.5 and elec_dens_fast weight of 25. Each sequence window was energy minimized in triplicate and the resulting scores were averaged. Symmetry was applied to pose objects to increase speed of computation.

Energetic calculations

Solvation free energy calculations were performed as recently described[17]. Briefly, the solvent accessible surface area (SASA) for each atom in a residue was determined (folded state). Next, the SASA for each atom of a residue was determined in the absence of the other residues (reference state), and the difference between both states was calculated ($SASA_{Ref} – SASA_{Fold}$). This value was then multiplied by the Atomic Solvation Parameter (ASP) specific to each atom, as determined previously by Eisenberg et al[47]. An entropic term is also included to take into consideration the degrees of freedom lost in going from a disordered to ordered state[48]. The energies of all atoms were then summed to generate the solvation energy for each structure. Difference energy maps were generated by subtracting solvation free energies pairwise for each atom in the two structures being compared.

Molecular dynamics simulations

Molecular dynamics simulations were performed with the GROMACS[49] software package (version 2022.2) using the CHARMM27[50] all-atom forcefield. Simulations were carried out with a 5-layer CPEB fibril under varying charge states to simulate fibril dynamics at pHs 3, 5, and 7. In each case, the fibril was placed in a cubic box, solvated with SPC/E three-point water molecules

and added counter ions. The system was energy minimized, then temperature and pressure equilibrated for 100 ps. Final simulations were executed for 200 ns.

Thioflavin T-binding and dissolution assays

Protein was resuspended in fibril formation buffer as described previous at 25 μM concentration with equal parts ThT. Solution was pipetted into a 96-well well plate with optical bottom (Sigma-Aldrich) and incubated at 37°C with continuous shaking. ThT fluorescence was measure with an excitation filter of 440 nm and emission filter of 480nm using a Varioskan plate reader (ThermoFischer). Aggregation curves were generated from triplicate, n=3 wells. Dissolution assays were started after 18 hours of fibril formation. Because fibrils were not able to be resuspended in different buffers due to their liability, a 96-well plate was prepared by diluting each fibril condition 3-fold in corresponding buffers. Normalization of ThT intensity was performed to account for variations across experiments and reduction of signal when diluting samples for dissolution assays.

Cell culture

HT22 hippocampal neurons and U87 glioblastomas were cultured in Dulbecco's modified Eagle's medium with 10% Fetal Bovine Serum (ThermoFisher) and 1% penicillin/streptomycin (ThermoFisher) in a tissue-culture incubator at 37°C and 5% CO2. Cells were cultured as described above and plated on 12-mm glass coverslips coated with poly–L-lysine (Sigma) and Laminin 50 uG (Sigma) in 6-well cell-culture plates. Cells were plated at 70% confluency overnight and transfected using Lipofectamine 3000 kit (ThermoFisher) for an 8 hr incubation. Cells were harvested and/or imaged at 24 hours post transfection for all experiments. Live fluorescent images were acquired on an EVOS M7000 (ThermoFisher) on a 40X objective.

Immunofluorescence and puncta analysis

Cells were fixed in 2% paraformaldehyde at 4°C for 10 minutes and permeabilized using 0.5% PBS-Triton X-100 for 10 minutes. Cells were subsequently washed with PBS-T and blocked with 2% BSA, 0.5%FSG, and PBS-T solution for 1hr at RT. Cells were incubated with 1° antibodies diluted in blocking solution for approximately 2 hours. Cells were washed with PBS-T three times before beginning incubation with 2° antibodies and washed with PBS-T. Coverslips were mounted using Prolong Glass Antifade Mountant (ThermoFisher) and allowed to dry for 72 hours. Fluorescent images were acquired on a Leica confocal SP8-STED/FLIM/FCS on a 60X water immersion objective. Puncta were manually counted and distances were manually calculated by averaging X and Y for each puncta in GFP channels. Image analyses were performed in FIJI (ImageJ).

Luciferase assays

HEK293 cells were plated at 50% confluency overnight and transfected using Lipofectamine 3000 kit (ThermoFisher) with ~0.5 μg of CPEB3 DNA, 0.5 μg of Renilla luciferase appended with GluA2 3′ UTR or SUMO2 3'UTR, and 0.5 μg of Firefly luciferase[51].
18/24 hr after transfection, cells were lysed with 100 μl of buffer for dual luciferase assay (Promega). 20 μl of the lysates were used for the quantification of Firefly and Renilla Luciferase activity using a Luminometer (GloMax, Promega).

Hippocampal neurons cultured for 9–10 days in Neurobasal medium with B27 supplement (Invitrogen) at a cell density of 30 000–40 000/cm2 were cotransfected (Lipofectamine LTX,

LifeScience) for 3 h with ~0.5 µg of CPEB3 DNA, 0.5 µg of Renilla luciferase appended with GluA2 3′ UTR, and 0.5 µg of Firefly luciferase. We used glycine to stimulate the NMDA receptors selectively, a protocol that mimics LTP induction in cultured neurons. Chem-LTP was induced as described previously[52]. Briefly, neuronal cultures were transferred from Neurobasalgrowth medium to extracellular solution (ECS) containing: 150 mM NaCl, 2 mM CaCl 2 , 5 mM KCl, 10 mM HEPES(pH 7.4), 30 mM glucose, 0.5mM TTX, 20mM bicuculline methiodide. The transfected neurons were stimulated with 200µM glycine for 3 min and then washed for 30 minutes before lysis in 100 µl of buffer for dual luciferase assay (Promega). 20 µl of the lysates were used for the quantification of Firefly and Renilla Luciferase activity using a Luminometer (GloMax, Promega). The ratio between Renilla and Firefly is calculated for each experimental condition.

Statistical analyses

Number of repeats and (Avg) ± SE are stated in figure legends. Statistical significance was measured by one or two-way ANOVA followed by the Tukey's multiple comparisons test.

Cryo-ET sample preparation and cryo-FIB milling

Ultrathin carbon (2 nm) Quantifoil 2/1 300 mesh on Au grids or Quantifoil 2/2 300 mesh on silicon dioxide grids were glow discharged for 30 seconds and subsequently sterilized with 70% ethanol and coated with poly-L-lysine (Sigma) and Laminin 50 uG (Sigma). U87 cells were seeded and cultured as described previously in cell culture section. Neurons were transfected as previously described in cell culture section and samples were selected for vitrification by screening on an EVOS M7000 (ThermoFisher) with a 20X objective to check for CPEB3-GFP expression and distribution on grids. Selected samples were plunge frozen into liquid ethane using a Vitrobot Mark IV (FEI) set at 100% humidity, 37°C with a blot force of 10 and blot time of 15 seconds after 24 hours of expression. Grids were mounted on cryo-FIB autogrids and were loaded onto an Aquilos II cryogenic FIBSEM instrument for cryo-FIB milling (ThermoFisher). MAPS software (ThermoFischer) was used to aid in correlation using a combination of room temperature fluorescence images and images acquired using the integrated fluorescence light microscope (iFLM) in the FIBSEM and for acquisition of a SEM medium magnification montage. Fluorescent and SEM atlases were overlaid for correlation and selection of CPEB3-GFP expressing cells. Cells were selected for milling based on puncta phenotypes and overall placement on grid squares. Grids were coated using a gas injection system with an organometallic platinum layer of ~7-10 nm and then sputter coated with platinum for ~30 seconds. Briefly, cryo-FIB milling was performed using a stepwise decreasing current (500 pA – 30 pA), manually as described previously[53], and automatically using AutoTEM (ThermoFischer) with milling angles ranging from 8° – 15°. SEM imaging was used to monitor milling progress throughout. Final lamellae thickness ranged from 150 nm – 250 nm. Final polished lamellae were sputter coated for 15 seconds and the iFLM was used to acquire GFP fluorescence z-stacks to aid in correlation during data collection.

Cryo-ET data collection and tomographic reconstruction

Data were collected on a Titan Krios G3i microscope with a K3 direct detection camera (Gatan) operated at an accelerating voltage of 300 keV and a 20eV slit width (BioQuantum) or with a Titan Krios G4 with a Falcon4EC (ThermoFisher) and a 10eV slit width. Data were also collected on a Talos Arctica with a K2 Summit direct detection camera (Gatan) operating at an accelerating voltage of 200 keV and 20eV slit width (BioQuantum). For lamellae with fluorescent data, medium magnification montages were acquired using SerialEM and overlaid with GFP maximum intensity projections acquired on iFLM using MAPS. Tilt series acquisition was selected for based on signal presence and intensity on lamellae. Automated data acquisition were performed using SerialEM software at a nominal magnified pixel size of 3.64 Å/pix or 4.27

Å/pix. Dose symmetric tilt series were acquired between -60° to +60° with either 2° or 3° steps. On average, 7 frames were recorded for each image for a total electron dose of ~ 100 - 120 $e^-/Å^2$ and a defocus target of -5 - -7 µM. Frames were aligned using MotionCor2[54] and automatically reconstructed in EMAN2[55]. Tomograms were screened and selected for based on their contents and quality and subsequently reprocessed as follows. Tilt series were aligned using fiducial-less patch tracking and reconstructed by weighted back projection in IMOD[56] binned by two, with frames containing obstructions such as ice contaminations manually removed before alignment . A CTF-based deconvolution filter was applied to each tomogram for visualization and segmentation purposes[57].

Tomographic segmentation

Tomograms containing features of interest were further anisotropically binned by 2 in X and Y and denoised using a 3D general model in Topaz[58].

Acknowledgments:

We are grateful to Wong Hoi Hui of the Electron Imaging Center for NanoMachines (EICN) for help and advice during cryoEM screening and sample preparation. We thank Peng Ge (UCLA) for helpful discussions during cryoEM data processing, assistance during data acquisition and Duilio Cascio (UCLA) for assistance during model building and refinement. We thank Corey Hecksel (SLAC) for assistance during data acquisition. Part of this work was performed at the Stanford-SLAC Cryo-EM Center (S2C2), which is supported by the National Institutes of Health Common Fund Transformative High-Resolution Cryo-Electron Microscopy program (U24GM1295410). We thank Chensong Zhang and Jae Yang for assistance during cryo-ET data collection. Part of this work was performed at the Midwest Center for Cryo-ET (MCCET) and the Cryo- EM Research Center located in the Department of Biochemistry at the University of Wisconsin-Madison, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM139168). M.D.F is supported by Eugene V. Cota-Robles Fellowship, Ruth L. Kirschstein NRSA GM007185, Whitcome Pre-Doctoral Fellowship, Audree V. Fowler Fellowship and a National Science Foundation Graduate Research Fellowship. J.A.R. is supported as a Pew Scholar, a Beckman Young Investigator and a Packard Fellow. This work was performed as part of STROBE, an NSF Science and Technology Center through Grant DMR-1548924, and by DOE Grant DE-FC02-02ER63421. The work was also supported by NIH-NIGMS Grant R35 GM128867.

Author Contributions:

M.D.F designed experiments, purified recombinant constructs, performed negative stain EM and ThT assays, prepared cryo-EM samples, performed cryoEM data collection and processing helical reconstruction, cultured mammalian cells, performed live fluorescence imaging, immunofluorescence experiments, performed cryo-FIB milling and cryo-ET data collection, reconstruction, post processing and segmentation. M.R.S performed solvation energy calculations and assisted in model building and refinement. S.Z. performed computational threading and MD simulations. D.R.B assisted in cryo-EM data processing and helical reconstruction. S.T and A.S analyzed and quantified fluorescent puncta. T.Z. and J.C. assisted in recombinant protein expression and purification. L.F. cultured primary neurons, performed luciferase assays and analysis and guided the project. J.A.R supervised and guided the project. M.D.F and J.A.R wrote the manuscript with input from all authors.

Competing Financial Interests:
JAR is an equity stake holder of MedStruc Inc.

References

1. Prusiner, S. B. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science (80-.).* **216**, 136–144 (1982).
2. Shorter, J. & Lindquist, S. Prions as adaptive conduits of memory and inheritance. *Nature Reviews Genetics* vol. 6 435–450 (2005).
3. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
4. Rayman, J. B. & Kandel, E. R. Functional Prions in the Brain. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
5. McGaugh, J. L. Memory - A century of consolidation. *Science (80-.).* **287**, 248–251 (2000).
6. Si, K., Lindquist, S. & Kandel, E. R. A Neuronal Isoform of the Aplysia CPEB Has Prion-Like Properties. *Cell* **115**, 879–891 (2003).
7. Khan, M. R. *et al.* Amyloidogenic Oligomerization Transforms Drosophila Orb2 from a Translation Repressor to an Activator. *Cell* **163**, 1468–1483 (2015).
8. Keleman, K., Krüttner, S., Alenius, M. & Dickson, B. J. Function of the Drosophila CPEB protein Orb2 in long-term courtship memory. (2007) doi:10.1038/nn1996.
9. Fioriti, L. *et al.* The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. *Neuron* **86**, 1433–1448 (2015).
10. Stephan, J. S. *et al.* The CPEB3 Protein Is a Functional Prion that Interacts with the Actin Cytoskeleton. *Cell Rep.* **11**, 1772–1785 (2015).
11. Pavlopoulos, E. *et al.* Neuralized1 activates CPEB3: A function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell* **147**, 1369–1383 (2011).

12. Ford, L., Ling, E., Kandel, E. R. & Fioriti, L. CPEB3 inhibits translation of mRNA targets by localizing them to P bodies. **116**, 18078–18087 (2019).
13. Drisaldi, B. *et al.* SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep.* **11**, 1694–1702 (2015).
14. Hervas, R. *et al.* Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science (80-.).* **367**, 1230–1234 (2020).
15. Fiumara, F., Fioriti, L., Kandel, E. R. & Hendrickson, W. A. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143**, 1121–1135 (2010).
16. Reselammal, D. S. *et al.* Mapping the Fibril Core of the Prion Subdomain of the Mammalian CPEB3 that is Involved in Long Term Memory Retention. *J. Mol. Biol.* **433**, (2021).
17. Sawaya, M. R., Hughes, M. P., Rodriguez, J. A., Riek, R. & Eisenberg, D. S. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**, 4857–4873 (2021).
18. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627.e16 (2017).
19. Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218-289) prion form a β solenoid with a triangular hydrophobic core. *Science (80-.).* **319**, 1523–1526 (2008).
20. Lu, J. *et al.* CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. doi:10.1038/s41467-020-17905-y.
21. McGlinchey, R. P. & Lee, J. C. Reversing the Amyloid Trend: Mechanism of Fibril Assembly and Dissolution of the Repeat Domain from a Human Functional Amyloid. *Isr. J. Chem.* **57**, 613–621 (2017).
22. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science (80-.).* **325**, 328–332 (2009).

23. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26**, 668–679 (2016).

24. Jaafari, N. *et al.* SUMOylation Is Required for Glycine-Induced Increases in AMPA Receptor Surface Expression (ChemLTP) in Hippocampal Neurons. *PLoS One* **8**, 52345 (2013).

25. Bäuerlein, F. J. B. *et al.* In Situ Architecture and Cellular Interactions of PolyQ Inclusions. *Cell* **171**, 179-187.e10 (2017).

26. Rogers, S. *et al.* Triglyceride lipolysis triggers liquid crystalline phases in lipid droplets and alters the LD proteome. *J. Cell Biol.* **221**, (2022).

27. Mahamid, J. *et al.* Liquid-crystalline phase transitions in lipid droplets are related to cellular states and specific organelle association. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16866–16871 (2019).

28. Vestal, D. J. & Jeyaratnam, J. A. The Guanylate-Binding Proteins: Emerging Insights into the Biochemical Properties and Functions of This Family of Large Interferon-Induced Guanosine Triphosphatase. *J. Interf. Cytokine Res.* **31**, 89 (2011).

29. Richter, J. D. CPEB: a life in translation. *Trends Biochem. Sci.* **32**, 279–285 (2007).

30. Kozlov, E., Shidlovskii, Y. V, Gilmutdinov, R., Schedl, P. & Zhukova, M. The role of CPEB family proteins in the nervous system function in the norm and pathology. *Cell Biosci* **11**, 64 (2021).

31. Raveendra, B. L. *et al.* Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. *Nat. Struct. Mol. Biol.* **20**, (2013).

32. Sinning, A. *et al.* Minireview: pH and synaptic transmission. (2013) doi:10.1016/j.febslet.2013.04.045.

33. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat. Struct. Mol. Biol.* doi:10.1038/s41594-019-0248-4.

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

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

36. Manka, S. W. *et al.* 2.7 Å cryo-EM structure of ex vivo RML prion fibrils. doi:10.1101/2021.12.13.472424.

37. Kraus, A. *et al.* High-resolution structure and strain comparison of infectious mammalian prions. *Mol. Cell* **81**, 4540-4551.e6 (2021).

38. Guo, Q. *et al.* In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. *Cell* **172**, 696-705.e12 (2018).

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

40. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *bioRxiv* 1–9 (2019) doi:10.1101/823310.

41. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 1–13 (2019).

42. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nat. 2017 5477662* **547**, 185–190 (2017).

43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Biological Crystallography Features and development of Coot. doi:10.1107/S0907444910007493.

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

45. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

46. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and

design HHS Public Access. *J Chem Theory Comput* **13**, 3031–3048 (2017).

47. Eisenberg, David, M. D. A. Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).

48. Koehl, P. & Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249–275 (1994).

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

50. Vanommeslaeghe, K. *et al.* CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671 (2010).

51. Cormier, M. J. & Kazuo, H. Studies on the bioluminescence of Renilla renoformis. *Biochim. Biophys. Acta - Spec. Sect. Biophys. Subj.* **88**, 99–104 (1964).

52. Jaafari, N., Konopacki, F. A., Owen, T. F., Kantamneni, S. & Rubin, P. SUMOylation Is Required for Glycine-Induced Increases in AMPA Receptor Surface Expression (ChemLTP) in Hippocampal Neurons. *PLoS One* **8**, 52345 (2013).

53. Schaffer, M. *et al.* Cryo-focused Ion Beam Sample Preparation for Imaging Vitreous Cells by Cryo-electron Tomography. *BIO-PROTOCOL* **5**, (2015).

54. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods 2017 144* **14**, 331–332 (2017).

55. Marrs, V. *et al.* A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat. Methods* doi:10.1038/s41592-019-0591-8.

56. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).

57. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods 2019 1611* **16**, 1146–1152 (2019).

58. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun. 2020 111* **11**, 1–12 (2020).

Chapter 4

CLOSURE

Amyloid and prion proteins have fascinated the realm of pathogenic and structural biology for several decades. The discovery of the functional prion-like protein in the brain, CPEB, demonstrated that the amyloid structural motif is an exceptional, evolutionary conserved assembly, successfully exploited and tolerated in the nervous system. To fully understand its conformational plasticity and mechanisms of assembly, molecular resolution structural information is required. A major emphasis must be placed on the toleration of CPEB in the brain, how its switch-like activation is fine-tuned, and how its rapid assembly and propagation is regulated in the cell. To address these challenges, a combination of cellular and structural biology techniques must be used. The breakthrough of single particle cryo-EM and helical reconstruction software has provided high-resolution snapshots of CPEB from a model organism. However, to truly decipher the complex intricacies of CPEB and draw clear distinctions from its pathogenic counterparts, molecular networks and storage mechanisms need to be directly visualized in unperturbed environments.

My thesis work focused on the mammalian CPEB, CPEB3, and its reliance on the amyloid state to perform its function. Assisted by the use of single particle cryo-EM, I determined the structural identity of the transient CPEB3 amyloid core. Translational assays and immunofluorescent data showed the core sequence is required for translational regulation by CPEB3 and proper cellular colocalization. With the combined application of specimen preparation and correlative microscopy on the forefront of cryo-

ET, I visualized the macromolecular structures present in neurons expressing CPEB3. I described unique ultrastructures present within the cytoplasm and proposed mechanisms for the ailing health of CPEB3 expressing neurons. In conclusion, my work demonstrated that the use of the functional amyloid motif for CPEB is conserved in mammals. I give glimpses of the molecular ensembles and eventual toxicity present when CPEB3 is over expressed in cells, bearing the reality that CPEB assemblies are capable of perpetuating cellular changes found in neurodegenerative disease.

Appendix Chapter 1

This chapter is based on the published article “Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain”.

Ferrero, S., **Flores, M. D**., Short, C., Vazquez, C. A., Clark, L. E., Ziegenbein, J., Zink, S., Fuentes, D., Payes, C., Batto, M. V., Collazo, M., García, C. C., Abraham, J., Cordo, S. M., Rodriguez, J. A., & Helguera, G. (2021). Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain. Journal of Virology, 95(17), 1868–1888. https://doi.org/10.1128/JVI.01868-20/FORMAT/EPUB

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

STRUCTURE AND ASSEMBLY

Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain

Sol Ferrero,[a] Maria D. Flores,[b] Connor Short,[b] Cecilia A. Vazquez,[c,g] Lars E. Clark,[d] James Ziegenbein,[b] Samantha Zink,[b] Daniel Fuentes,[b] Cristian Payes,[a] María V. Batto,[a] Michael Collazo,[e] Cybele C. García,[c,g] Jonathan Abraham,[d,f] Sandra M. Cordo,[c] Jose A. Rodriguez,[b] Gustavo Helguera[a]

[a]Laboratory of Pharmaceutical Biotechnology, Instituto de Biología y Medicina Experimental (IBYME-CONICET), Buenos Aires, Argentina
[b]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, California, USA
[c]Laboratorio de Virología, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), IQUIBICEN, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)-UBA, Ciudad Universitaria, Buenos Aires, Argentina
[d]Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA
[e]Department of Biological Chemistry and Department of Chemistry and Biochemistry, University of California Los Angeles, Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, USA
[f]Department of Medicine, Division of Infectious Diseases, Brigham and Women's Hospital, Boston, Massachusetts, USA
[g]IQUIBICEN, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)-UBA, Ciudad Universitaria, Buenos Aires, Argentina

Sol Ferrero and Maria D. Flores contributed equally to this work and share equal right to be listed as principal author. Author order was determined alphabetically.

ABSTRACT Pathogenic clade B New World mammarenaviruses (NWM) can cause Argentine, Venezuelan, Brazilian, and Bolivian hemorrhagic fevers. Sequence variability among NWM glycoproteins (GP) poses a challenge to the development of broadly neutralizing therapeutics against the entire clade of viruses. However, blockade of their shared binding site on the apical domain of human transferrin receptor 1 (hTfR1/CD71) presents an opportunity for the development of effective and broadly neutralizing therapeutics. Here, we demonstrate that the murine monoclonal antibody OKT9, which targets the apical domain of hTfR1, can sterically block cellular entry by viral particles presenting clade B NWM glycoproteins (GP1-GP2). OKT9 blockade is also effective against viral particles pseudotyped with glycoproteins of a recently identified pathogenic Sabia-like virus. With nanomolar affinity for hTfR1, the OKT9 antigen binding fragment (OKT9-Fab) sterically blocks clade B NWM-GP1s and reduces infectivity of an attenuated strain of Junin virus. Binding of OKT9 to the hTfR1 ectodomain in its soluble, dimeric state produces stable assemblies that are observable by negative-stain electron microscopy. A model of the OKT9-sTfR1 complex, informed by the known crystallographic structure of sTfR1 and a newly determined structure of the OKT9 antigen binding fragment (Fab), suggests that OKT9 and the Machupo virus GP1 share a binding site on the hTfR1 apical domain. The structural basis for this interaction presents a framework for the design and development of high-affinity, broadly acting agents targeting clade B NWMs.

IMPORTANCE Pathogenic clade B NWMs cause grave infectious diseases, the South American hemorrhagic fevers. Their etiological agents are Junin (JUNV), Guanarito (GTOV), Sabiá (SABV), Machupo (MACV), Chapare (CHAV), and a new Sabiá-like (SABV-L) virus recently identified in Brazil. These are priority A pathogens due to their high infectivity and mortality, their potential for person-to-person transmission, and the limited availability of effective therapeutics and vaccines to curb their effects. While low homology between surface glycoproteins of NWMs foils efforts to develop broadly neutralizing therapies targeting NWMs, this work provides structural evidence that OKT9, a monoclonal antibody targeting a single NWM glycoprotein binding site on hTfR1, can efficiently prevent their entry into cells.

Citation Ferrero S, Flores MD, Short C, Vazquez CA, Clark LE, Ziegenbein J, Zink S, Fuentes D, Payes C, Batto MV, Collazo M, García CC, Abraham J, Cordo SM, Rodriguez JA, Helguera G. 2021. Antibody-based inhibition of pathogenic New World hemorrhagic fever mammarenaviruses by steric occlusion of the human transferrin receptor 1 apical domain. J Virol 95:e01868-20. https://doi.org/10.1128/JVI.01868-20.

Editor Rebecca Ellis Dutch, University of Kentucky College of Medicine

Address correspondence to Gustavo Helguera, gustavoh@ibyme.conicet.gov.ar.

Received 24 September 2020
Accepted 22 May 2021
Accepted manuscript posted online 16 June 2021
Published 10 August 2021

KEYWORDS Sabiá-like, X-ray crystallography, antiviral agents, electron microscopy, Junin, Machupo, mammarenavirus, monoclonal antibodies, transferrin receptor

Despite the gravity of South American hemorrhagic fevers (1–3), treatments against pathogenic New World mammarenaviruses (NWMs) remain limited. Low sequence homology among the glycoproteins of this family of viruses hinders the development of broadly neutralizing virus-targeting therapeutics. A vaccine against Argentine hemorrhagic fever (AHF) derived from attenuated JUNV strain CANDID#1 is produced in Argentina and has reduced incidence of the disease, but to date no effective FDA-approved treatment or vaccine exists against all New World hemorrhagic fevers (NWHFs) (4). Small molecules have shown limited efficacy against NWHF viruses; the antiviral favipiravir (T-705) in combination with rivabirin (5) and RNA aptamers are under investigation but have yet to reach clinical trials (44). Recent efforts have found success in the development of neutralizing monoclonal antibodies against NWHF mammarenavirus glycoproteins. However, while virus-neutralizing monoclonal antibodies can mimic important receptor contacts when engaging the glycoprotein of a particular NWM (6), they show limited or no activity against the glycoproteins of other members of the NWM family (4, 6).

Despite the low sequence identity between their glycoprotein sequences (25% to 46%), clade B NWMs share routes of entry into human cells. One such example is the human transferrin receptor (hTfR1), a single-pass transmembrane protein that regulates iron uptake into cells (7). Structural studies have determined the atomic nature of the interaction of Machupo virus (MACV) GP1 with hTfR1 and revealed a conserved binding site for clade B NWM GP1 binding on the apical domain of hTfR1 (8). The putative binding of all members of this clade to the same epitope on hTfR1 makes it an attractive molecular target for the development of broad-spectrum inhibitors of viral entry, including antibodies (9). Recombinant monoclonal antibodies have been identified that exploit this vulnerability by binding hTfR1 and blocking the internalization of pseudotyped viral particles decorated with Junin (JUNV), Guanarito (GTOV), Sabiá (SABV), MACV, Chapare (CHAV), and Sabiá (SABV) virus glycoproteins (GP) but not pseudotyped viral particles expressing glycoproteins from Old World hemorrhagic fever viruses, such as Lassa virus (LASV) (9). More recently, such antibodies have been shown to block entry of the North American mammarenavirus AV96010151 into cells, expanding their potential efficacy against a broader spectrum of NWHF viruses (45). Likewise, arenacept, a recombinant protein consisting of the apical domain of hTfR1 fused to an Fc domain, is capable of binding GP1 of several NWMs and preventing pseudotyped virus internalization into cells (10).

While hTfR1-targeting antibodies have demonstrated potential as broadly neutralizing therapeutics against clade B NWMs, this strategy faces key challenges. A concrete molecular basis for the inhibition of viral entry through hTfR1-targeting antibodies is still lacking. This is due in part to the absence of structures of these targeting agents bound to hTfR1 despite a large number being discovered and explored for other therapeutic applications. The murine monoclonal OKT9 was among the first antibodies shown to recognize hTfR1 in several cell lines and have its activity explored in a variety of contexts (11–13). We now demonstrate that OKT9 is a potential blocker of cell entry by clade B NWMs, as it engages an epitope on the apical domain of hTfR1 that is a shared binding site for clade B NWM glycoproteins. OKT9 prevents cell entry by pseudotyped viral particles with the surface glycoproteins of a newly sequenced and lethal Sabiá-like (SABV-L) NWM (14). We also demonstrate that OKT9 blockade can inhibit the *in vitro* replication of a competent, albeit attenuated, strain of JUNV (15). Knowledge that OKT9 blocks the GP1-TfR1 interaction through steric occlusion presents a mechanistic platform for the development of broadly active competitive inhibitors of NWHF that target viral entry mediated by hTfR1.

RESULTS

The OKT9 variable region binds hTfR1 with nanomolar affinity. Despite the longstanding characterization of OKT9 as a monoclonal antibody that binds hhTfR1, we first set out to biochemically assess this interaction using a recombinantly generated soluble form of hTfR1 with a C-terminal hexahistidine tag that we refer to as hTfR1 (16). Size-exclusion chromatography (SEC) of hTfR1 yields peaks that elute at column fractions consistent with monomer and multimer species in solution; sampling of peak fractions and evaluation by SDS-PAGE shows a single principal band at ~79 kDa, corresponding to a monomer. Traces for OKT9-Fab alone show a principal peak eluting at column fractions consistent with a monomer; a 50-kDa band is observed from these fractions by SDS-PAGE, consistent with the expected molecular weight for a Fab monomer (Fig. 1A). hTfR1 was mixed with OKT9-Fab at a 3:1 molar ratio and incubated for 20 min at room temperature to facilitate the formation of a solution-state complex. Traces of the mixture show a shift of peaks to earlier column fractions compared to the receptor alone. The peak corresponding to OKT9-Fab alone disappears in traces for the mixture, indicating a loss of free Fab in solution. This is confirmed by SDS-PAGE of peak fractions showing a coelution of Fab with TfR1 in early fractions (Fig. 1B).

We evaluated the binding of full-length OKT9 and its proteolytic antigen binding fragment (OKT9-Fab) to hTfR1 immobilized on a solid surface or in solution. Both OKT9 and OKT9-Fab bound surface-immobilized hTfR1 in a concentration-dependent manner with approximate 50% effective concentration (EC_{50}) values of 0.411 nM (OKT9) and 3.659 nM (OKT9-Fab) (Fig. 1C). The affinity of OKT9-Fab for hTfR1 was further assessed by biolayer interferometry (BLI), with various concentrations of full-length OKT9 exposed to a fixed concentration of hTfR1 immobilized on the biolayer sensor via an anti-His antibody and assuming a 1:1 interaction between Fab and receptor. Averaging across four biolayer cycles produced an equilibrium dissociation constant of 4.8 nM for the interaction between OKT9-Fab and hTfR1 (Fig. 1D). Rate constants for each cycle are shown in Table 1. Biolayer interferometry was also used to directly evaluate competition between OKT9-Fab and a MACV GP1-Fc fusion protein for binding to a soluble form of hTfR1 (sTfR1) in solution (Fig. 1E). MACV GP1-Fc bound to a biosensor surface was able to efficiently bind hTfR1, but this binding was abrogated in the presence of OKT9-Fab. By comparison, transferrin, which does not compete with MACV for binding to hTfR1 (7), had no effect on hTfR1 binding to immobilized MACV GP1-Fc. These results suggest that OKT9 and MACV GP1 bind overlapping sites on hTfR1.

OKT9 and OKT9-Fab block cellular entry of pseudotyped viruses presenting clade B NWM glycoproteins and a nonpathogenic strain of JUNV. Entry of pseudotyped viruses presenting clade B NWM glycoproteins into cells expressing hTfR1 was monitored through their expression of eGFP upon cellular entry, serving as a platform to assess the effectiveness of hTfR1 blockade by OKT9 (7). In this context, both OKT9 and OKT9-Fab inhibit the internalization of clade B NWM glycoprotein-presenting pseudovirus into HEK-293T cells. Unperturbed pseudovirus presenting glycoproteins from JUNV, SABV, MACV, GTOV, CHAV, SABV-L, and LASV all entered cells and produced green fluorescence. Cultured cells preincubated with 100 nM OKT9-Fab, OKT9, or a non-TfR1 specific antibody were exposed to JUNV, MACV, SABV-L, and LASV pseudovirus. Wide-field fluorescence images of cells treated with OKT9 or OKT9-Fab showed significantly lower levels of eGFP expression than cells without antibody treatment (Fig. 2A). While treatment with anti-TfR1 antibodies has been associated with downregulation of cell surface TfR1 (17), HEK-293T cells incubated with 100 mM OKT9 or OKT9-Fab for 48 h showed no decrease in cell binding of Tf-TMR. Fluorescence microscopy images showed similar binding of Tf to the cell surface across all conditions, as indicated by colocalization with the membrane-targeting fluorophore DiO (Fig. 2B). Changes in pseudovirus internalization were quantified by flow cytometry (Fig. 2C and D), showing decreased internalization of MACV, JUNV, and SABV-L in the presence of OKT9 and OKT9-Fab. Internalization of LASV pseudovirus was unaffected by OKT9 or OKT9-Fab, while a non-TfR1 specific antibody did not affect pseudovirus internalization (Fig. 2C and D). A 30-min preincubation of cells with OKT9 showed dose-dependent

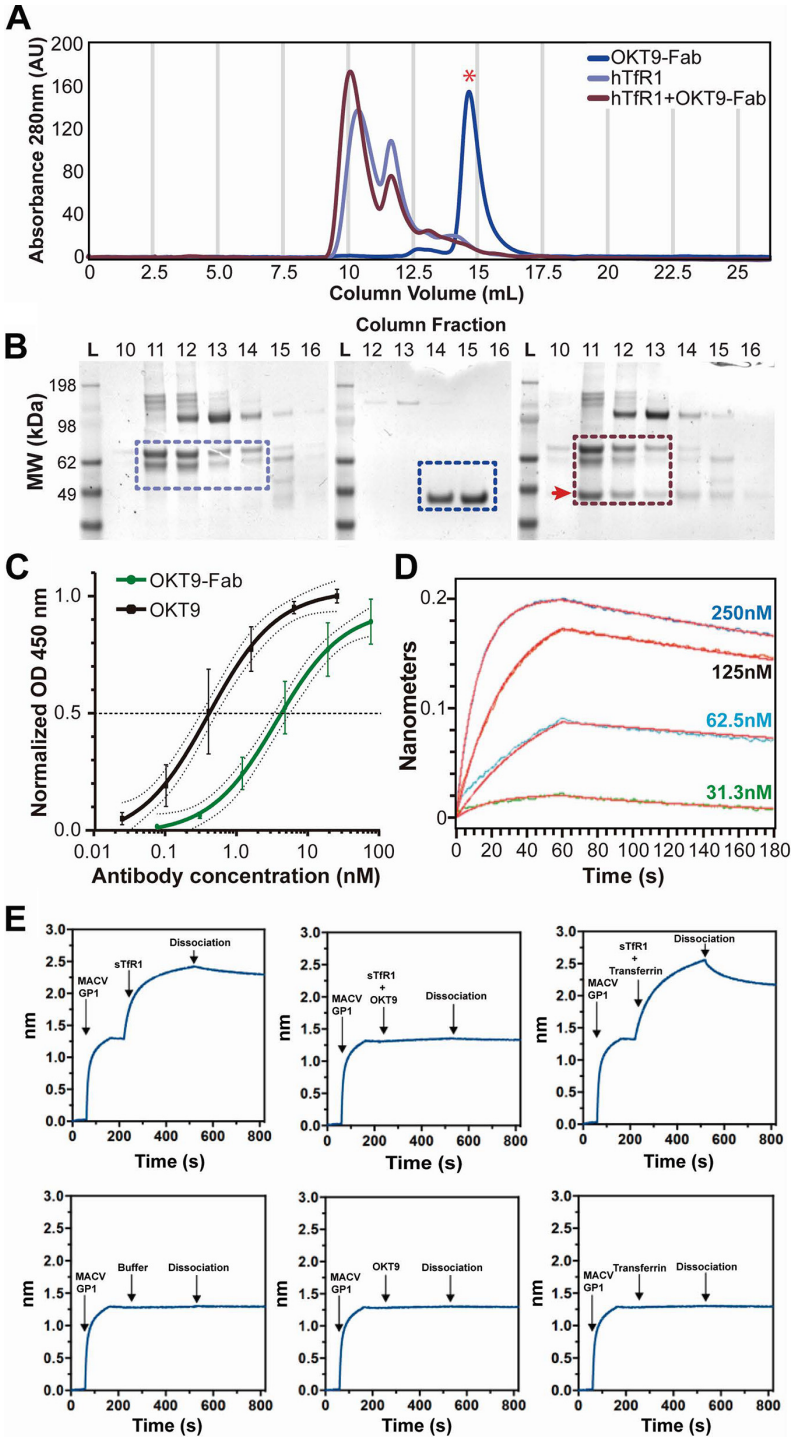


FIG 1 Characterization of OKT9 binding to hTfR1. (A) Formation of an hTfR1-OKT9 complex evaluated by SEC coelution. Samples include hTfR1 (purple), OKT9-Fab (blue), and sTfR1+OKT9-Fab (maroon). An overlay of all traces is shown for comparison. Gray bars indicate collected fractions. The red star denotes the expected peak fraction for OKT9-Fab alone. (B) SDS-PAGE of fractions collected from isolated and complex SEC runs. On the left is shown the SDS-PAGE of the SEC fractions of the hTfR1 run, in the center the fractions of the OKT9-Fab run, and on the right the fractions of the complex. Colored boxes indicate the presence of the protein of interest, coordinated

(Continued on next page)

TABLE 1 Kinetics analysis of OKT9-Fab binding to hTfR1 via bilayer interferometry[a]

Cycle	K_D (M)	K_D error	k_a (1/ms)	k_a error	K_{dis} (1/s)	K_{dis} error	Full X^2	Full R^2
A	4.76e−09	8.32e−11	2.38e+05	1.68e+03	1.14e−03	1.81e−05	0.0728	0.9951
B	4.82e−09	8.00e−11	2.50e+05	1.73e+03	1.20e−03	1.82e−05	0.0730	0.9954
C	4.90e−09	8.15e−11	2.54e+05	1.81e+03	1.25e−03	1.87e−05	0.0767	0.9952
D	4.78e−09	7.59e−11	2.62e+05	1.78e+03	1.25e−03	1.80e−05	0.0695	0.9956

[a]K_D, equilibrium dissociation constant; k_a, association rate constant; K_{dis}, dissociation rate constant.

inhibition of pseudovirus internalization (Fig. 2E), yielding 50% inhibitory concentration (IC_{50}) values of 0.329 nM (JUNV), 0.234 nM (SABV), 1.415 nM (MACV), 0.362 nM (GTOV), 0.274 nM (CHAV), and 0.632 nM (SABV-L) (Table 2).

Both OKT9 and OKT9-Fab blocked entry of a replicating strain of JUNV into A549 cells. Viral replication was measured 24 h postinfection by quantifying viral particles released to the supernatants and by analyzing viral RNA levels in infected cell monolayers. Cultures treated with OKT9 or OKT9-Fab produced smaller amounts of viral particles, as determined by plaque assays (Fig. 2F), and also had lower levels of viral RNA, as measured by reverse transcription-quantitative PCR (RT-qPCR) (Fig. 2G), compared to the nontreated control. These results suggest that both OKT9 and OKT9-Fab blocked viral entry, reducing the cell susceptibility to this virus and, thus, preventing cell infection. The decrease in viral replication is specific, since cultures treated with an IgG that did not target TfR1 showed no significant differences from buffer control.

The crystallographic structure of OKT9-Fab. A crystal structure of OKT9-Fab was determined to 2.0-Å resolution by microfocal X-ray diffraction (Table 3). The molecule crystallized with space group $P2_12_12$ with a single molecule in the asymmetric unit. The structure was determined by molecular replacement using chains from PDB entries 4Q0X and 4OTX. Sequences for the variable light and heavy chains, determined by sequencing of their genomic loci from the OKT9 hybridoma cell line (Fig. 3A), were used as guides during model building and data refinement (Fig. 3B).

The overall structure of OKT9-Fab resembles a canonical murine IgG1 variable region. All three complementary determining regions (CDRs) in the heavy and light chains were resolved (Fig. 3B). Their sequences were identified as CDR_{L1} QSLFNSANQENY (27-38), CDR_{L2} WAS (56-58), CDR_{L3} QNDYSYPLT (95-103), CDR_{H1} GYTFTNYY (26-33), CDR_{H2} IYPGDGNT (51-58), and CDR_{H3} ARDNSGYVGFPY (97-108). The constant heavy-chain region spanning from A136 to T139 showed poor density that precluded building during refinement and remained unbuilt in the final structure. Of greatest relevance was missing density for residue Y103 in CDR 3 of the heavy chain (Fig. 3B). Its side chain could not be placed during refinement and was omitted from our deposited structure but was included in subsequent computational analyses.

OKT9 binds hTfR1 to form closed bivalent complexes identifiable by electron microscopy. To evaluate the basis for inhibition of cellular internalization of NWM by OKT9 targeting of hTfR1, we sought to reveal the mode of interaction between full-length

FIG 1 Legend (Continued)
with the color used in SEC traces. The red arrow denotes the expected band for OKT9-Fab. (C) ELISA binding of OKT9 to hTfR1. An indirect ELISA was performed decorating the plate with sTfR1 and then incubating with different concentrations of OKT9-Fab and OKT9. Anti-mouse IgG conjugated to HRP was used as a secondary antibody. The EC_{50} calculated over the normalized OD_{450} for OKT9 was 0.411 nM with a 95% confidence interval (CI) of 0.213 to 0.633, and for OKT9-Fab the EC_{50} was 3.695 nM with a 95% CI of 2.463 to 6.028. (D) Kinetics of OKT9-Fab interaction with hTfR1 immobilized on a biosensor surface. The receptor was exposed to increasing concentrations of OKT9-Fab as labeled (31.3 nM, 62.5 nM, 125 nM, and 250 nM). One hundred eighty seconds of biolayer recordings show binding and dissociation. (E) Assessment of OKT9-Fab and MACV GP1-Fc for binding to sTfR1 by biolayer interferometry. MACV GP1-Fc immobilized onto anti-human Fc biosensor tips and incubated with sTfR1 (top left) or buffer alone (bottom left), sTfR1 in complex with OKT9-Fab (top center), OKT9-Fab alone (bottom center), sTR1 in complex with transferrin (top right), or transferrin alone (bottom right). The arrows indicate the time points at which the indicated proteins were added and the dissociation step. The data are representative of two replicates for each of the experimental conditions shown.

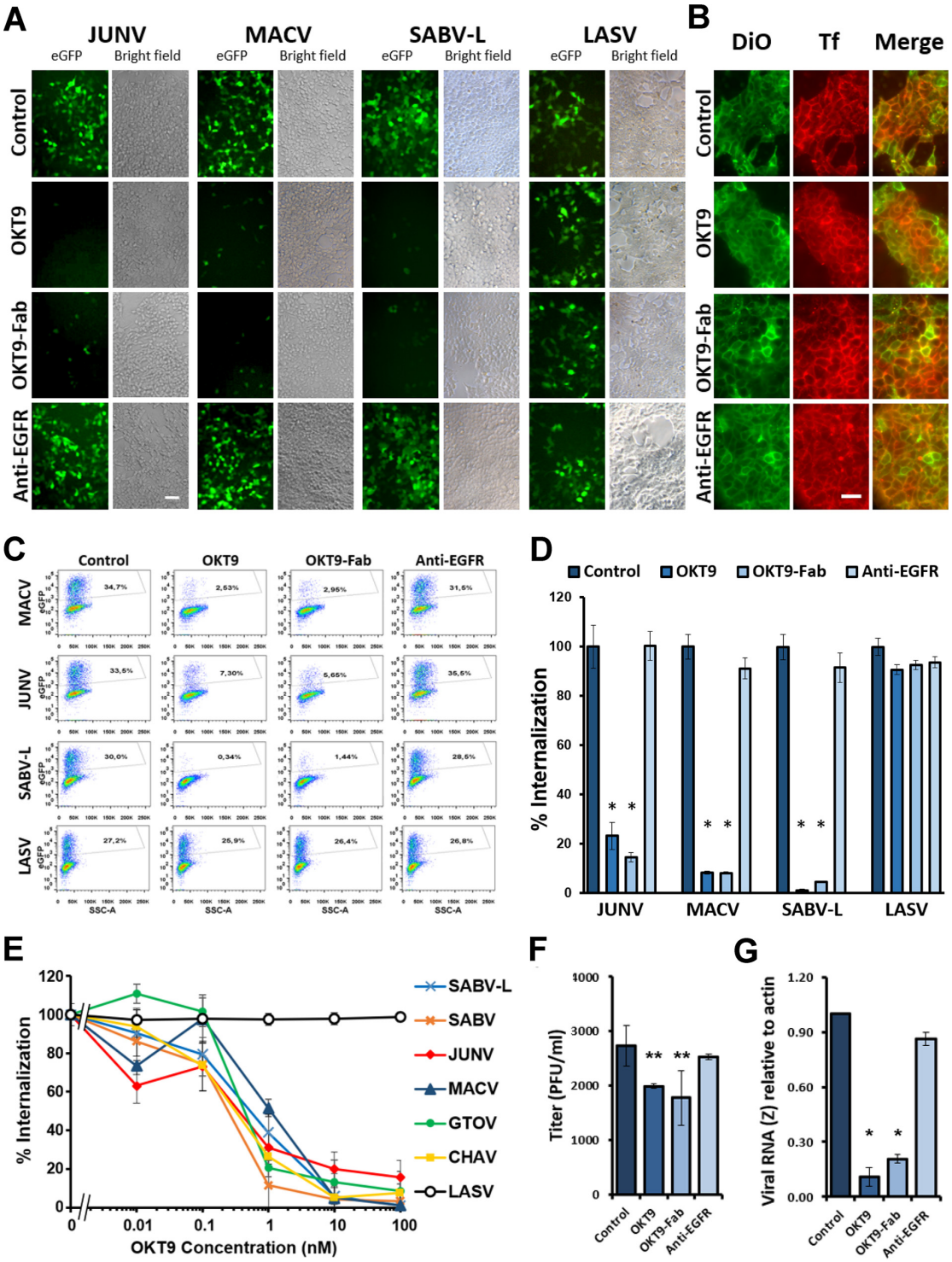


FIG 2 Inhibition of internalization of NWHF pseudovirus by OKT9. (A) Fluorescence microscopy images of HEK-293T cells. The images were taken from the inhibition of JUNV, MACV, SABV-L, and LASV pseudovirus internalization assays mentioned for panel B: pseudovirus + buffer control, pseudovirus + 100 nM OKT9, pseudovirus + 100 nM OKT9-Fab, and pseudovirus + 100 nM anti-EGFR as a nonrelevant antibody control. The reference bar indicates 50 μm. (B) Fluorescence microscopy of HEK-293T cells treated 48 h with buffer control, 100 nM OKT9,

(Continued on next page)

TABLE 2 Analysis of NWM pseudovirus inhibition of internalization by OKT9

Parameter	SABV	JUNV	MACV	GTOV	SABV-L	CHAV
IC_{50}[a] (nM)	0.2339	0.3294	1.4150	0.3622	0.6320	0.2740
95% CI, IC_{50} (nM)	0.1393 to 0.3954	0.1019 to 0.9872	0.7233 to 2.865	0.2056 to 0.6384	0.4270 to 0.9354	0.2213 to 0.3399
R^2	0.9609	0.8658	0.9311	0.9434	0.9803	0.9948

[a]The fitting model used is $Y = (\text{bottom} + \text{top-bottom})/(1 + 10^{(X\text{-LogIC50})})$, where IC_{50} is the concentration of OKT9 that gives a response halfway between bottom and top, and top and bottom are plateaus in the units of the *y* axis.

OKT9 and hTfR1. A solution containing OKT9 and hTfR1 was drop-cast onto electron microscopy grids and negatively stained with uranyl acetate. Use of an equimolar ratio of OKT9 to hTfR1 allowed for the formation of stable, homogenous assemblies; complexes with a prominent diamond-shaped structure were identified (Fig. 3C). Imaging of these grids at various tilt angles, ranging from 0 to 45 degrees, circumvented the preferred orientation of the flat, diamond-like assemblies on the grid. Combining images across all tilts revealed multimeric complexes of receptor and OKT9 that could be subclassified using single-particle methods (Fig. 3C). A low-resolution structure of OKT9 in complex with hTfR1 was determined from these images. Particles contributing to the best classes were used to generate an *ab initio* model used for further refinement (Fig. 3E). Despite low particle count and preferred particle orientations, the refined three-dimensional (3D) density determined from these data revealed the binding interface between hTfR1 and OKT9 (Fig. 4A). The Fc region of OKT9 could not be identified in 2D averages or in the final 3D density, likely due to the flexibility of the hinge region (Fig. 3C). The resulting density showed sufficient resolution to suggest the relative orientation of the OKT9 variable regions with respect to the apical domain of hTfR1.

Computational prediction of OKT9-Fab binding footprint on hTfR1. We validated the configuration of OKT9 relative to hTfR1 in our reconstruction by comparing the reconstructed density to a series of docked models generated using the web-based server ClusPro 2.0 (18). For this, we relied on the known structure of hTfR1 (PDB entry 3KAS) and our crystal structure of OKT9-Fab. We used ClusPro in antibody mode, where CDR residues and the framework surface residues were excluded in the repulsion mask of the Fab. An attraction mask was generated on the surface residues of hTfR1 based on the reconstructed 3D density of the OKT9-TfR1 complex. Lastly, a repulsion mask was generated to exclude other residues on hTfR1 that flanked this area. ClusPro generated 10 docked models (Fig. 4B). OKT9-Fab models docked onto hTfR1 were fit onto the 3D density maps for the OKT9-hTfR1 complex using a rigid-body procedure in Chimera. This procedure yielded two top models that matched the relative orientation

FIG 2 Legend (Continued)
100 nM OKT9-Fab, or 100 nM anti-EGFR, labeled with the DiO membrane dye (green), and incubated for 30 s with transferrin-TMR (red). The reference bar indicates 34 μm. (C) Flow cytometry dot plot analysis of JUNV, MACV, SABV-L, and LASV pseudovirus internalization in HEK-293T cells showing eGFP expression in the absence and presence of 100 nM OKT9, OKT9-Fab, and anti-EGFR. In the case of MACV, JUNV, and SABV-L, a reduction of the percent eGFP-positive events occurs in the presence of OKT9 and OKT9-Fab compared to buffer control and the nonrelevant antibody. In the case of LASV, similar levels of eGFP events are observed independently of the treatment. The percentages of eGFP-positive events are indicated inside the gates. (D) Inhibition of JUNV, MACV, and SABV-L pseudovirus internalization. Relative entry rate of JUNV, MACV, the recently reported SABV-L, and control LASV pseudovirus in HEK-293T cells was quantified in the presence of 100 nM OKT9-Fab, 100 nM OKT9, and 100 nM anti-EGFR. The data were 100% normalized with the cells without treatment, and the significant differences are indicated by comparing Fab OKT9, OKT9, and the negative-control anti-EGFR versus no treatment (*, $P < 0.005$, Student's *t* test for unpaired data of triplicate determinations). (E) Relative entry rate of pseudoviruses decorated with the GP1/GP2 complex of SABV, JUNV, MACV, GTOV, CHAV, SABV-L, and LASV to HEK-293T cells in the presence of OKT9 (0.01 to 100 nM). Pseudoviruses were loaded with an eGFP expression vector to express once internalized. After 48 h, the cells were fixed and the percentage of positive internalization events quantified by flow cytometry. The data are expressed as the means ± standard deviations (SD) from the sample. The data were normalized to 100% with the cells without treatment. (F) OKT9 and OKT9-Fab inhibit the infectivity of JUNV IV4454 virus strain. A549 cells were preincubated for 1 h with 200 nM OKT9, OKT9-Fab, anti-EGFR, or medium alone (control) and then infected with JUNV at an MOI of 0.01. After 1 h of incubation, viral inocula were replaced with the respective antibody-supplemented medium or medium alone, and 24 h postinfection, total JUNV production in A549 cell supernatants was measured using a PFU assay in Vero cells. The graph shows means ± SD from a representative experiment (from four independent experiments). The statistical analysis performed was ANOVA followed by Duncan's test (**, $P < 0.05$). (G) Set of 200 nM OKT9-, OKT9-Fab-, anti-EGFR-, or medium alone (control)-treated A549 cells monolayers was harvested with TRIzol for RNA extraction 18 h postinfection with JUNV IV4454. Viral RNA (z gene) was quantified using RT-PCR, using actin as a housekeeping gene. The graph shows means ± SD from two independent experiments. The statistical analysis performed was ANOVA followed by Dunnett's test (*, $P < 0.0001$).

TABLE 3 Data collection and refinement statistics for OKT9-Fab

Parameter	Value
Data collection	
Space group	P 2_1 2_1 2
Cell constants a, b, c (Å)	94.66, 113.96, 40.84
Cell constants α, β, γ(°)	90, 90, 90
Resolution (Å)	48.82–2.00
% Data completeness	99.2 (48.82–2.00)
Mean I/σ(I)	15.66 (3.12)
Wilson B-factor (Å^2)	34.78
Refinement	
No. of total reflections	154,169
No. of unique reflections	30,478
$CC_{1/2}$	99.9
No. of reflections used in refinement	30,471
No. of reflections used for Rfree	3,048
Rwork	0.2164
Rfree	0.2535
No. of nonhydrogen atoms	3,475
No. of protein residues	433
Wavelength (Å)	0.979
Ramachandran plot (%)	
Residues in favored region	97.65
Residues in allowed region	2.35
Residues in outlier region	0

of 3D density for OKT9-Fab with respect to hTfR1 and best fit the 3D antibody density (Fig. 4C). While the resolution of this 3D model does not allow for the identification of residue-level interactions, the overall binding footprint of OKT9 on hTfR1 is identifiable (Fig. 4D). Analysis of the putative binding interface for the best of these models shows a significant overlap of the binding site for MACV GP1 on hTfR1, with both OKT9 and MACV GP1 competing to bind an exposed loop in the hTfR1 apical domain (Fig. 4E).

DISCUSSION

The treatment of infectious diseases caused by viral agents with high sequence and structural diversity presents a challenge for targeted therapy (19). Broadly neutralizing antiviral antibodies are scarce given the sequence diversity of viral epitopes and their constant evolutionary pressure; this applies to NWMs. In contrast, strategies that target a host receptor, such as hTfR1, are appealing given its cell surface accessibility, low sequence variability across populations, and well-characterized structure and physiology (20). The identification of a single conserved epitope on hTfR1 as a shared binding site for clade B NWMs and the determination of a crystallographic structure of MACV GP1 bound to the conserved epitope have illuminated avenues for NWM neutralization. This knowledge has already enabled the isolation and characterization of various NWM neutralizing antibodies (4, 6, 7, 21).

Antibodies capable of neutralizing or preventing cell entry of more than one NWM have targeted either the site of hTfR1 binding on MACV and JUNV GP1 or the hTfR1 apical domain itself (9, 22). Taking advantage of this opportunity for broad NWM neutralization by hTfR1 targeting (8), we establish the well-known, hTfR1-targeting murine monoclonal antibody OKT9 as a framework for future development of NWM-neutralizing agents that reduce cellular uptake of NWMs. OKT9 competes with only one known natural ligand, heavy-chain ferritin (Fe) (Fig. 4E) (23). However, homeostatic alterations resulting from this competition are unclear given that iron uptake via ferritin is not a principal mechanism of iron uptake into cells. Importantly, OKT9 is not expected to interfere with transferrin (Tf) binding to hTfR1, as that interaction relies entirely on the membrane, proximal helical, and protease-like domains (Fig. 4E) (20). Moreover,

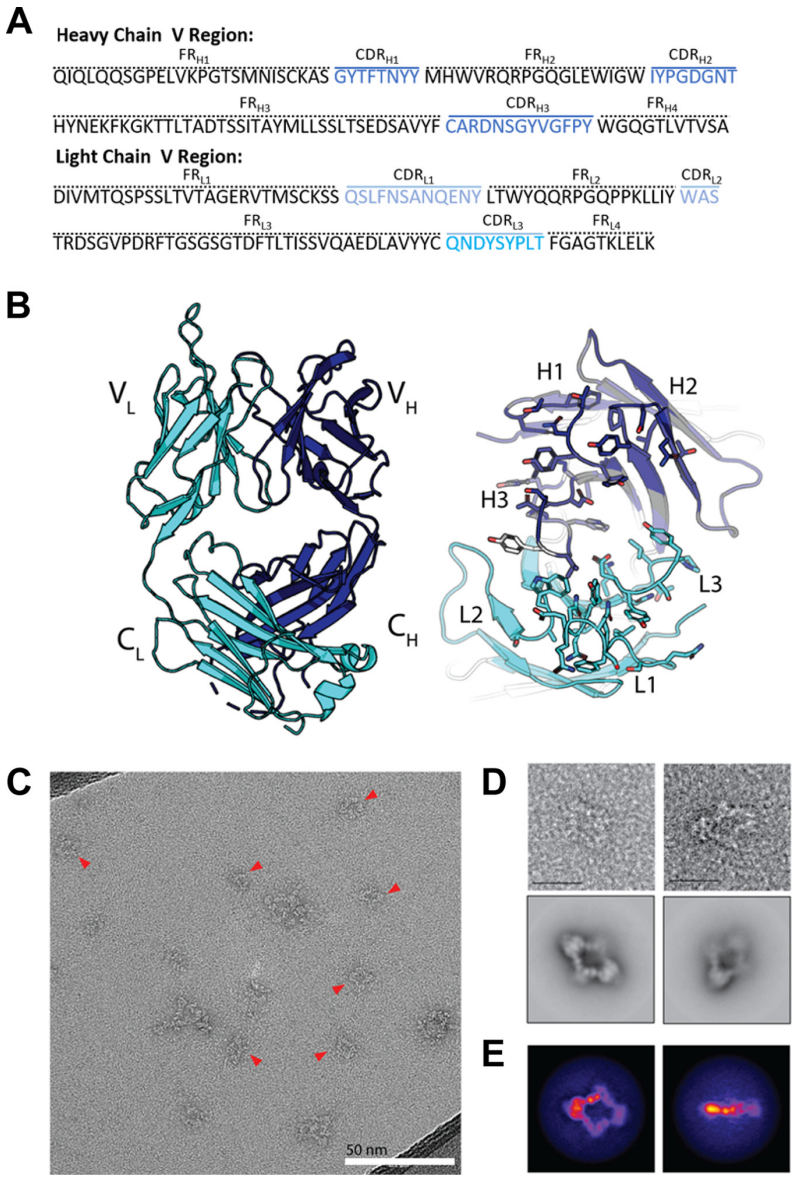


FIG 3 Crystallographic structure of OKT9-Fab and visualization of the hTfR1-OKT9 complex. (A) Amino acid sequence of OKT9 heavy-chain (top) and light-chain (bottom) variable region. In blue are the heavy-chain CDRs H1, H2, and H3 and in light blue the light-chain CDRs L1, L2, and L3. (B) X-ray crystallographic structure of OKT9-Fab with heavy chain colored in slate and light chain in cyan. Loops corresponding to variable light chain CDRs are labeled L1, L2, and L3, and CDRs for the variable heavy chain are labeled H1, H2, and H3, where unmodeled regions are shown as dashed lines. An unmodeled tyrosine residue in CDR_{H3} is shown in white. (C) Micrograph of negatively stained sample containing hTfR1-OKT9 complex. Red triangles indicate particles selected for 2D classification (scale bar, 50 nm). (D) Representative particles and 2D class averages of negatively stained hTfR1-OKT9 complexes on ultrathin carbon (scale bar, 200 Å). (E) Projections of 3D hTfR1-OKT9 complex density along two orthogonal directions. 3D density was obtained by *ab initio* reconstruction from negative-stain images of the complex.

despite previous reports of anti-TfR1 antibodies depleting cell surface TfR1 by altering its recycling and inducing its accumulation and degradation in late endosomes or lysosomes (17), 48-h treatment of HEK-293T cells with OKT9 or OKT9-Fab did not significantly alter cell surface binding of Tf.

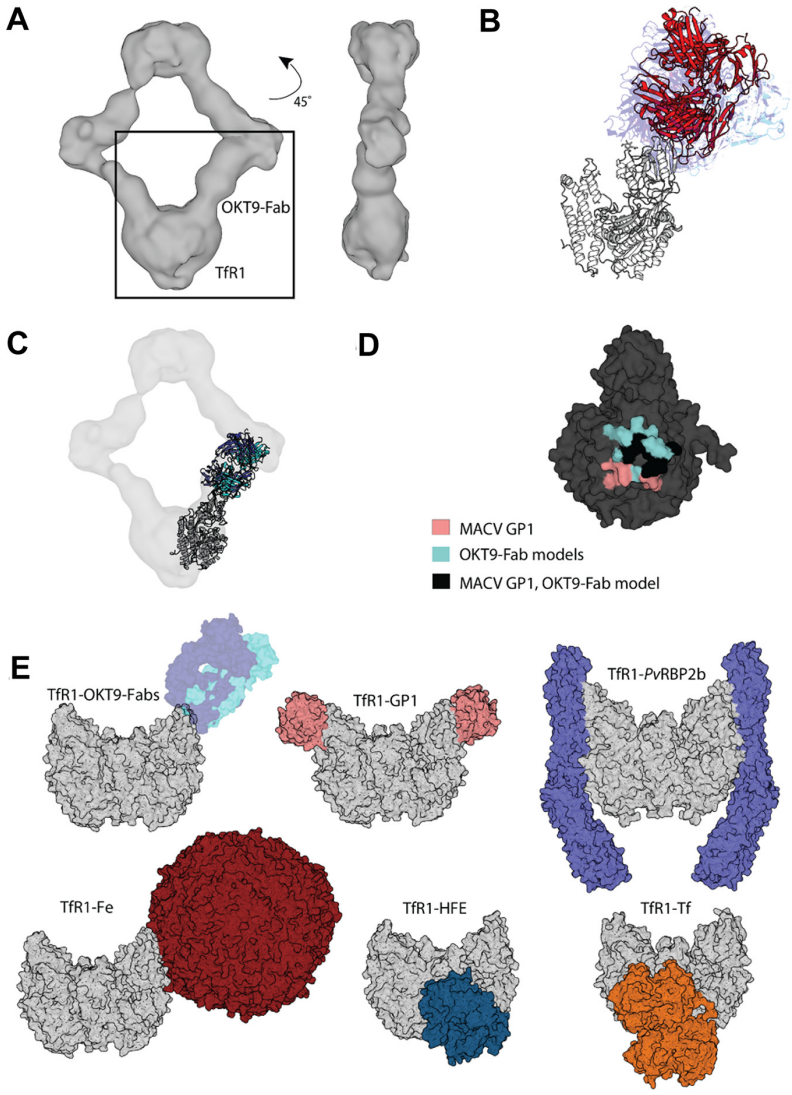


FIG 4 Docking model of hTfR1-OKT9 complex and relationship to natural ligands and human pathogen molecular ligands. (A) Isosurface rendering of hTfR1-OKT9 3D density map with an estimated resolution of 11 Å. Inset shows a well-resolved region of the 3D density, corresponding to a monomeric transferrin receptor bound by an OKT9-Fab. Front (left) and side (right) views are shown. (B) Front views of all 10 ClusPro-generated models using OKT9-Fab and 3KAS hTfR1 apical domain monomer (white). Best-fit models used for analysis are highlighted in red; other models are shown in hues of blue. (C) Overlay of ClusPro generated model using OKT9-Fab and crystal structure PDB entry 3KAS with hTfR1-OKT9 3D density. (D) Surface representation of the 3KAS apical domain (dark gray) colored to illustrate the binding interface with MACV GP1 protein (salmon), OKT9-Fab model (blue), or both (black). (E, top) Transferrin receptor (gray) bound to natural ligands (transferrin, PDB entry 1SUV; hereditary hemochromatosis factor [HFE] [43], PDB entry 1DE4; and ferritin, PDB entry 6GSR). (Bottom) Transferrin receptor bound to OKT9-Fabs and human pathogens (MACV-GP1, PDB entry 3KAS; *P. vivax*, PDB entry 6D04).

The variable region of OKT9 binds hTfR1 with nanomolar affinity, allowing full-length OKT9 to inhibit cellular entry by NWM with nanomolar IC_{50}. This efficient inhibition of NWM internalization is enabled by the high overall avidity of full-length OKT9, with an EC_{50} value an order of magnitude better than that for OKT9-Fab. The affinity estimated for the binding of OKT9-Fab to hTfR1 is on par with that previously reported for other TfR1-binding antibodies, including ch128.1, with equilibrium dissociation

constant (K_D) values of 4.82 nM and 5.7 nM, respectively (9). OKT9-Fab also binds hTfR1 with equal or greater affinity than virus-neutralizing antibody variable fragments targeting GP1 of JUNV (5 nM) and MACV (16.3 nM) (22). These comparisons illustrate the opportunity for high-affinity inhibition of viral entry by targeting hTfR1, provided the targeting agent effectively precludes binding of NWM GP1 to the receptor. The potency of OKT9 viral entry inhibition is further demonstrated on a competent attenuated JUNV, where treatment with the antibody led to reduced plaque formation and viral detection by RT-PCR. This approach succeeded despite the potential for non-TfR1-mediated viral entry, possibly involving C-type lectins DC-SIGN/L-SIGN (24, 25), the phosphatidylserine receptors Axl (26), TIM-1 (27), or fluid-phase uptake.

Use of single variable-domain antibodies or antibody fragments is warranted when potent enough to act within the timescale of their half-life. Antibody fragments are especially preferable when their corresponding full-length antibody is associated with a risk of immunotoxic effects in humans (28). For these reasons, use of OKT9-Fab could be favored over its full-length counterpart given a history of potential homeostatic dysregulation by full-length OKT9: incubation of the erythroleukemia cell line K562 to full-length OKT9 induces cellular redistribution and increased degradation of TfR1 (29). Potential autoimmune reactions add to concerns over homeostatic dysregulation induced by full-length OKT9, as with other full-length antibodies. The superior avidity of full-length OKT9 over OKT9-Fab can also be a detriment, as has been observed in relation to transcytosis across the blood-brain barrier (30). These concerns are especially pressing for hTfR1 targeting given its ubiquitous expression across tissue types. However, several of these potential issues can be mitigated by engineering of full-length antibodies to suppress binding of immune effector molecules to their Fc region and/or introducing mutations that reduce the affinity of each variable region.

The overall site of interaction between the OKT9 variable region and the apical domain of hTfR1 stands near a previously identified loop thought to be a key determinant for MACV GP1 binding on hTfR1. In fact, based on negative-stain single-particle electron microscopy and computational docking models, residues D204, Y211, and N348 on hTfR1 are hypothesized to be bound by both MACV GP1 and OKT9 and are key determinants of zoonotic transmission across arenaviruses (31). Direct competitive inhibition of NWM GP1 binding to hTfR1 by the OKT9-Fab is supported by biolayer interferometry-based competition assays and presents a molecular framework for dissecting NWM blockade by hTfR1-targeting antibodies. Therefore, hTfR1 blockade by the OKT9 variable region alone is sufficient for blocking viral entry and permits the discovery of new NWM viruses that utilize TfR1 as a port of entry into human cells. We demonstrate this by investigating the recently reported and lethal Sabia variant (SABV-L), whose genome sequence has been determined, and its S segments coding for its glycoproteins share 87% homology with those of SABV strain SPH114202 (14). OKT9 blockade of GP-mediated cellular entry of this new virus not only confirms its use of hTfR1 as a portal of entry into human cells but also could inform on the impact of hTfR1 binding by other pathogens, such as the malaria parasite *Plasmodium vivax* (Fig. 4E) (32). Importantly, the unique geometry of the ternary complex formed by dimeric receptor and OKT9 in solution suggest efficient steric occlusion of the hTfR1 apical domain with limited cross-linking at the cell surface.

Monoclonal antibody-based therapies are now front-line treatments for a variety of diseases. Only four monoclonal antibodies have been FDA approved to treat an infectious disease. Three of these directly target infectious pathogens, including palivizumab for respiratory syncytial virus prophylaxis (46) and raxibacumab and obiltoxaximab for treatment of inhalational anthrax (33, 34). In contrast, ibalizumab targets CD4 to block the entry of HIV-1 into human cells (35). Because of the rare occurrence of NWHFs as well as their relatively short manifestation in humans, broadly neutralizing antibodies represent feasible antiviral therapeutics for existing and emerging clade B NWMs. Our approach further enables a possible route to treatment for those infected

or at risk of getting infected by hemorrhagic fever viruses that share hTfR1 as a host receptor, including viral strains that are yet to be identified.

MATERIALS AND METHODS

OKT9 IgG1 production. Murine hybridomas expressing the monoclonal antibody OKT9 were obtained from the ATCC (catalog number CRL-8021). Cultures initially grew in Iscove's modified Dulbecco's medium (IMDM) supplemented with fetal bovine serum (FBS) (10%), GlutaMAX, and penicillin-streptomycin. Once sufficiently proliferated, the cells were passed to 2-liter roller bottles and weaned from FBS dependence using increasing volumes of SFM (Gibco, USA) before the medium was harvested. Cells were spun down, and the supernatant was filtered by gravity filtration against cellulose paper (Millipore, USA) to remove any large particulate matter. This was then passed through a 0.22-μm Millipore filter, and the clarified medium was purified by affinity chromatography using a protein G column HiTrap protein G HP (GE LifeSciences, USA) equilibrated to Tris-buffered saline (TBS) with 200 mM NaCl at pH 7.4. The OKT9 was eluted using a gradient of TBS with 100 mM glycine and 200 mM NaCl at pH 2.6; 100 μl of 1 M TBS at pH 9.0 was added to each fraction. These fractions were analyzed by SDS-PAGE, and the purest samples were consolidated and concentrated to ~200 to 500 μl using an Amicon Ultra-4 10K centrifugal filter device (Millipore, USA). The resulting protein was further purified via SEC using an Enrich SEC 650 10- by 300-mm column (Bio-Rad, USA) connected to an NGC Quest 100 chromatography system instrument (Bio-Rad, USA) equilibrated with TBS. SEC fractions were analyzed by SDS-PAGE and the purest bands were concentrated using an Amicon Ultra-4 10K centrifugal filter device to ~500 μl.

OKT9-Fab production. Purified OKT9 was digested with ficin using the mouse IgG_1 Fab and $F(ab)_2$ preparation kit (Pierce, Thermo Scientific, USA) according to the manufacturer's instructions. After digestion, the sample was purified by SEC using an Enrich SEC 650 10- by 300-mm column (Bio-Rad, USA) and NGC Quest 100 chromatography system chromatography equipment (Bio-Rad, USA) equilibrated with TBS. These fractions were collected and analyzed by SDS-PAGE to confirm purity and integrity. Fractions were subsequently concentrated to ~100 μl using an Amicon Ultra-4 10K centrifugal filter device for use in crystallization trays.

OKT9 variable region sequence analysis. OKT9 heavy- and light-chain variable region residue compositions were determined using mRNA isolated from the hybridoma cell line subjected to RT-PCR with mouse-specific primer sets to amplify the target regions by using a proprietary sequencing procedure (LakePharma, USA).

Production and purification of human sTfR1. Baby hamster kidney (BHK) cells expressing the histidine-tagged soluble extracellular domain of human TfR1 (sTfR1) were cultured at 37°C in IMDM with 1× GlutaMAX and 1× penicillin-streptomycin antibiotic mixture and certified 10% fetal bovine serum (Thermo Fisher, USA). BHK cells were trypsinized with TrypLE express (Thermo Fisher, USA), centrifuged, and replated after harvest of the sTfR1-containing supernatant. Supernatant was clarified with a 0.22-μm Millipore vacuum filter and passed through a His trap (GE Healthcare, Millipore, USA) column via fast protein liquid chromatography (FPLC) on a BioLogicDuoFlow (Bio-Rad, USA). The column was washed with a 90:10 mixture of TBS (250 mM NaCl, 20 mM Tris, pH 7.4) and imidazole buffers (250 mM NaCl, 20 mM Tris, 500 mM imidazole) and then eluted via a gradient of increasing imidazole in TBS. Elution fractions were analyzed via SDS-PAGE, and similar samples were pooled and further purified via an Enrich SEC 650 10- by 300-mm column (Bio-Rad, USA) and NGC Quest 100 chromatography system chromatography equipment (Bio-Rad, USA).

Size exclusion chromatography of OKT9-Fab alone and in complex with soluble human TfR1. Three hundred micrograms of OKT9-Fab in 500 μl of TBS (50 mM Tris, 150 mM NaCl, pH 7.6) was injected into a 2-ml loop of a medium-pressure liquid chromatography system (NGC; Bio-Rad, USA) and run through an Enrich SEC 650 column (Bio-Rad, USA) using TBS at a flow rate of 0.8 ml/min. A similar procedure was performed using 900 μg of sTfR1 alone. A complex of sTfR1 and OKT9-Fab was allowed to form by mixing their solutions at a 3:1 molar ratio of receptor to Fab, followed by incubation at room temperature for 20 min prior to loading into the injection loop. All fractions were collected and analyzed by SDS-PAGE: 15 μl of each sample was added to 3.5 μl of NuPAGE LDS sample buffer, heated to 98°C for 10 min before being loaded into a precast NuPAGE 4 to 12% bis-acrylamide gel; sizing was analyzed against a See Blue 2 protein standard. Samples were run for 1 h at 100 V and then stained with SimplyBlue SafeStain before being imaged on an AzureBiosystems 300c gel imager using the AzureBiosystems gel analysis software. Analysis of gel images was performed in ImageJ.

Binding of OKT9 to human sTfR1 by ELISA. A 96-well plate was covered with 50 μl/well of a 1.5-μg/ml solution of human sTfR1 in carbonate buffer (0.015 M Na_2CO_3, 0.035 M $NaHCO_3$, pH 9.3) for 16 h at 4°C. The plate was washed 3 times with 200 μl/well phosphate-buffered saline (PBS), 0.05% Tween 20; 200 μl of blocking solution (5% skim milk powder, 1% bovine serum albumin (BSA), and 0.02% sodium azide in PBS, pH 7.4) was added to each well and incubated for 1 h at room temperature. The plate was then washed 3 times with 200 μl/well of PBS, 0.05% Tween 20, and 50 μl/well of OKT9 (0.025 to 25.8 nM) or OKT9-Fab (0.077 to 76.9 nM) was added to the wells and incubated for 2 h at room temperature. At the end of the incubation time, the plate was washed 3 times with 200 μl/well of PBS, 0.05% Tween 20 and incubated with an anti-mouse IgG horseradish peroxidase (Vector Laboratories) (1:2,000) in a 1:3 dilution of blocking solution for 1 h at 37°C. The plate was then washed 4 times with 200 μl/well of PBS, 0.05% Tween 20 and 80 μl/well of TMB substrate solution (BD Biosciences) was added and incubated for 30 to 45 min at room temperature. The reaction was stopped with 80 μl/well of 2 M H_2SO_4, and the plate was read by absorbance at an optical density at 450 nm with an enzyme-linked immunosorbent assay

(ELISA) plate reader (Thermo Scientific, USA). EC_{50} values were obtained by nonlinear regression fitting to a variable slope, four-parameter dose-response model using GraphPad Prism 8 (www.graphpad.com).

Binding analysis of OKT9 to human sTfR1 by BLI. All BLI assays were performed on a ForteBio Octet 96RED at a volume for all the solutions of 200 μl/well, with agitation set to 1,000 rpm at 30°C in solid black 96-well plates (Greiner, USA). HIS1K (Anti-Penta-HIS, USA) probes were equilibrated for 600 s in TBS, pH 7.6 (50 mM Tris-Cl, pH 7.6, 150 mM NaCl), prior to sample loading; 100 μg/ml sTfR1-His ligand was loaded for 60 s, followed by a biosensor baseline equilibration step in TBS for 60 s. Typical capture levels were between 0.4 and 0.5 nm and variability within a row of eight tips did not exceed 0.1 nm. Loading was followed by a 60-s association step, during which probes were exposed to OKT9-Fab at concentrations of 250 nM, 125 nM, and 62.5 nM in TBS. Dissociation of Fab was performed in TBS for 900 s. A similar experiment was performed in which 2-fold dilutions from 857.1 μg/ml to 13.4 μg/ml sTfR1-His were immobilized to HIS1K probes for 90 s, followed by a baseline equilibration step in TBS for 90 s. Association of OKT9-Fab at 70 nM for 60 s was next, followed by dissociation in TBS buffer for 600 s.

Data analysis and curve fitting were conducted using Octet 96RED analysis software. Correction of any systematic baseline drift was achieved by subtracting the shift recorded for a reference sensor loaded without sTfR1-His and incubated with OKT9-Fab. All data were filtered using the Savitzky-Golay algorithm and fitted with binding equations that assumed a 1:1 interaction. Global nonlinear least-squares fitting was performed on data sets that included association and dissociation steps.

MACV GP1-Fc BLI competition assay. The soluble human transferrin receptor 1 (sTfR1) for these studies was expressed in HEK293S cells. The ectodomain of human transferrin receptor 1 (GenBank accession no. NM_003234.3, residues 121 to 760) was cloned into the pHLsec expression vector downstream of the pHLsec (36) secretion signal. sTfR1 was transfected in HEK293S GnTI$^{-/-}$ cells (ATCC CRL-3022TM) using linear polyethylenimine (PEI). HEK293S GnTI$^{-/-}$ cells were cultured as suspension cells in FreeStyle 293 expression medium (Gibco, Thermo Scientific, USA) supplemented with 2% (vol/vol) ultra-low IgG FBS (Thermo Fisher, USA) and penicillin-streptomycin. Supernatant was harvested 72 h post-transfection. sTfR1 was purified by affinity chromatography using human transferrin-coupled Sepharose as previously described (8, 37, 38) and eluted in buffer containing 2 M potassium chloride and 50 mM HEPES (pH 7.5). sTfR1 was further purified by size-exclusion chromatography on a Superdex 200 increase column (GE Healthcare Life Sciences, USA), eluting at the expected retention time. DNA encoding the MACV GP1 subunit (GenBank accession no. NC_005078, residues 87 to 250) was cloned into a pVRC8400 vector containing human IgG Fc (a gift from Aaron Schmidt). The resulting vector was transfected in HEK293T cells grown in suspension. The MACV GP1-Fc fusion protein was purified by protein A affinity purification according to the manufacturer's protocol (Thermo Fisher Scientific), followed by size-exclusion chromatography on a Superdex 200 increase column. The assay was performed using an Octet RED96 system (ForteBio, USA). MACV GP1-Fc was loaded onto anti-human IgG Fc capture biosensor tips (ForteBio, USA) at 17 nM for 100 s in kinetic buffer (PBS containing 0.02% [vol/vol] Tween and 0.1% BSA), followed by a baseline measurement taken for 120 s. Next, sTfR1 at 1.5 μM, sTfR1 at 1.5 μM plus OKT-9 at 2 μM, or sTfR1 at 1.5 μM plus transferrin at 2 μM were associated for 300 s. The sensor tip was then placed in kinetic buffer for 300 s for a subsequent dissociation step.

Generation of pseudotyped virus presenting NWHF glycoproteins. The procedures and vectors used to make the pseudoviruses used here were described previously (7). Briefly, pseudotyped virus particles presenting the surface glycoproteins of JUNV, MACV, CHAV, GTOV, SABV, SABV-L, and LASV were generated by cotransfecting HEK-293T cells with vectors for the expression of the polyprotein gag/pol of the nonreplicating murine leukemia virus, GP1 and GP2 of each of the viruses, and the retroviral vector pQCXIX expressing eGFP (enhanced green fluorescent protein). Transfection was achieved by mixing equimolar quantities of all three vectors in the presence of calcium phosphate (7) or Lipofectamine 2000 (Invitrogen, USA). HEK-293T cells were incubated with the vector cocktail at 37°C with 5% CO_2. Twenty-four hours posttransfection, medium was exchanged, and supernatant with the released pseudoviruses was collected 48 h and 72 h posttransfection. The medium containing each of the pseudoviruses was clarified by centrifugation at 10,000 $\times$ *g* and subsequently filtered with a 0.45-μm polyvinylidene difluoride filter (Jetbiofil, China), and pseudovirus particles were further concentrated using a Vivaspin 1,000-kDa filter (Sartorius, USA). Aliquots were used immediately or stored at −80°C.

Cellular internalization of pseudotyped virus presenting NWHF glycoproteins. HEK-293T cells at approximately 40% confluence were preincubated with medium containing a serial dilution (from 0.001 to 100 nM) of OKT9 or with 100 nM OKT9-Fab and 50 nM anti-EGFR (negative control) for 30 min at 37°C with 5% CO_2 in 48-well plates (CellATTACH Biofil, China). One hundred microliters of the concentrated solution of pseudovirus particles of JUNV, MACV, CHAV, GTOV, SABV, SABV-L, and LASV (control) in the presence of the different antibody conditions was added to each well. Cells were incubated for 16 h and washed with fresh medium. After 48 h of incubation at 37°C with 5% CO_2, cells were inspected by wide-field fluorescence microscopy on an Olympus CKX41 inverted microscope (Olympus, Japan) equipped with an LCAch N 20×/0.40 Php objective (Olympus, Japan), and bright-field and fluorescence images were acquired with a Q-Color5 digital imaging system (Olympus, Japan). Cells were then fixed with 2% paraformaldehyde, and the inhibition of internalization was quantified by flow cytometry using a FACScanto (BD Biosciences, USA) equipped with an argon ion laser (488 nm), with filter settings for GFP (530/30 nm), and the flow data were analyzed with the software FlowJo 7.6 (BD Biosciences, USA). Relative internalization was determined as the percentage of cells positive for GFP normalized to cells treated in the absence of antibodies. The IC_{50} data were calculated from each pseudovirus dose-response data set. For this analysis, we used a dose-response model assuming that the log (inhibitor) versus response curves follow a symmetrical sigmoidal shape. The IC_{50} of OKT9 for JUNV, MACV, CHAV, GTOV, SABV, and SABV-L was determined as the concentration that provokes a response halfway

between the maximal (top) response and the maximally inhibited (bottom) response. We did not include the estimated IC_{50} of LASV, because it did not exhibit inhibition in the OKT9 tested concentration range. The fitting model used is Y = bottom + top-bottom)/(1 + $10^{(X\text{-LogIC50})}$), where IC_{50} is the concentration of OKT9 that gives a response halfway between bottom and top; and top and bottom are plateaus in the units of the Y axis. Statistical differences between OKT9, OKT9-Fab, and anti-EGFR compared to non-treated control conditions were analyzed using nonpaired t test with two tails and assuming different variances between groups.

Cell labeling with fluorescent transferrin and a fluorescent membrane probe. The fluorescent labeling of HEK-293T cells with Tf-TMR (Invitrogen, USA) and the lipophilic fluorophore Vybrant DiO cell-labeling solution (Invitrogen, USA) was performed by following the manufacturer's instructions. Briefly, HEK-293T cells at 50% confluence and grown in coverslips were incubated with medium containing 100 nM OKT9, 100 nM OKT9-Fab, 100 nM anti-EGFR, or no antibody for 48 h. Medium was exchanged, and 350 μl of 50 μg/ml Tf-TMR (red) diluted in medium without serum was added to each well. Cells were incubated 30 s and washed one time with PBS. After that, cells were fixed with 4% paraformaldehyde for 20 min. Cells were washed 3 times for 3 min with PBS, and 350 μl of Vybrant DiO (green) cell-labeling solution diluted 1:200 was added and incubated for 20 min at 37°C. Next, cells were washed with PBS and mounted on a slide. Cells were inspected by wide-field fluorescence microscopy on an Olympus IX83 motorized confocal DSU (disk spinning unit) module (Olympus, Japan) equipped with a PlanAPO N 60×/1.42 objective. Bright-field and fluorescence images were acquired with ORCA FLASH 4.0 V2 digital CMOS (4 MP, 16 bits) (Hamamatsu, Japan) and CellSens Dimensions software (Olympus, Japan). The green and red image overlay was generated using the ImageJ Merge function.

***In vitro* inhibition of JUNV infection.** A549 cells (ATCC CCL-185) were plated in 24-well plates at a density of 10^5 cells per well. After 24 h, the cells were incubated for 1 h with 200 nM OKT9, OKT9-Fab, or a control IgG (anti-EGFR) or medium alone (control). After incubation, cells were infected for 1 h with JUNV IV4544 viral strain (15) at a multiplicity of infection (MOI) of 0.01 in the presence of antibodies. After removing the inoculum, cell monolayers were incubated with the respective antibody-supplemented medium or medium alone. The supernatants were harvested 24 h postinfection, and JUNV production was titrated by a standard plaque assay in Vero cells.

Quantitative real-time PCR determination of viral load was assessed as follows. A549 monolayers were harvested, and RNA was extracted using Tri Reagent (Molecular Research Center, USA) according to the manufacturer's instructions. cDNA then was generated by use of murine reverse transcriptase M-MLV (Promega, USA) and random primers (Biodynamics, Argentina). This cDNA was amplified by real-time PCR using SYBR green complete mix detection (Roche, USA). Real-time PCR was carried out with an initial incubation at 95°C during 5 min, followed by 45 cycles of 30 s at 95°C, 45 s at 60°C, and 30 s at 72°C and a final step of 5 min at 72°C. Amplification plots were analyzed with Bio-Rad software, and the comparative threshold cycle method was used to determine viral gene expression relative to the β-actin cellular gene. Primer sequences are the following: Actin Fw, 5′-GAGACCTTCAACACCCCAGCC-3′; Actin Rv, 5′-GGCCATCTCTTGCTCGAAGTC-3′; Z Fw, 5′-ATGGGCAACTGCAACGGGGCATC-3′; Z Rv, 5′-CTATGGT GGTGGTGCTGTTGGCT-3′.

Statistical analysis was carried out using InfoStat software (http://www.infostat.com.ar). Randomized block analysis of variance (ANOVA) was performed in titration assays, while ANOVA was performed in RNA quantification assays. In both, statistical significance was assessed at a level of a P value of <0.05, followed by Dunnett's or Duncan's *post hoc* test (InfoStat; http://www.infostat.com.ar). Graphs were created using GraphPad Prism software (La Jolla, CA; www.graphpad.com).

Crystallographic analysis of the OKT9-Fab. The OKT9-Fab was concentrated to ~10 mg/ml in 50 mM Tris-HCl, pH 7.5, with an Amicon Ultra-4 10K centrifugal filter device (Millipore, USA) and then filtered using a 0.22-μm centrifuge filter (Corning, USA) (Table 3). A LabTech mosquito crystal nanodispenser TTP robot (TTP LabTech, UK) was used to evaluate 576 different conditions in a pendant drop vapor diffusion system. The plates were incubated in an environment without vibrations at approximately 18°C. Several conditions produced crystals of various qualities, and a $MgCl_2$-containing mother liquor was chosen for optimization. In a 24-well optimization plate (Hampton Research, USA), vapor diffusion/hanging-drop conditions in 1:1 and 1:2 drops of protein to mother liquor were prepared using various concentrations of Tris, $MgCl_2$, and polyethylene glycol (PEG) 8000. Large, blade-like crystals grew under conditions containing 100 mM Tris, 200 mM $MgCl_2$, and 20%, wt/vol, PEG 8000. These crystals were collected and passed through glycerol as a cryoprotectant, mounted on a microloop, and flash frozen into liquid nitrogen for transport. X-ray diffraction data were obtained at beamline 24-ID-E at the APS synchrotron (Advanced Photon Source, Argonne National Laboratory). Diffraction patterns were reduced in XDS and then merged and scaled in XSCALE (39). The structure was phased by molecular replacement using PHASER in the PHENIX suite of crystallographic software (40) with a chimeric model generated from two antibody Fabs (PDB entries 4Q0X and 4OTX). Following initial refinement in BUSTER (https://www.globalphasing.com/), final refinement was carried out in PHENIX, with model building in COOT (41). Visualization of models and structural analyses were performed in PyMOL (https://pymol.org/2/).

Electron microscopy of OKT9-sTfR1 complexes. Purified OKT9 and sTfR1 were diluted to 10 μg/ml and 2 μg/ml, respectively. A 1:1 ratio of diluted OKT9 and sTfR1 was then mixed and further diluted by a factor of two. OKT9-sTfR1 negative-stained grids were prepared for electron microscopy by placing a 2.5-μl drop of complex onto a glow-discharged ultrathin C film on Lacey Carbon 300 mesh (Ted Pella, Inc., USA) and washed twice with 2% uranyl acetate. Negative-stain grids were imaged at room temperature using a Tecnai F30 operating at 300 keV at a magnification of ×120,000 on a TVIPS XF-416 camera with a pixel size of 16 μm and a sensor with 4,096 by 4,096 pixels. These settings corresponded to a pixel size of approximately 2.0 Å/pixel. A total of 5 data sets were collected at various tilt angles to facilitate

views of the complex and avoid orientation bias. In addition to zero tilt, data were collected at a tilt of 10, 25, 35, and 45°.

Single-particle analysis of OKT9-sTfR1 complexes. Global contrast transfer function was calculated using Gctf implemented in RELION 3.0.7 for micrographs in all data sets. Approximately 982 (0° tilt), 946 (10° tilt), 1,389 (25° tilt), 2,044 (35° tilt), and 909 (45° tilt) particles were manually picked from the respective data sets. OKT9-sTfR1 particles were extracted in RELION using a 250-pixel box size. After extraction, STAR files were joined to combine particles from all data sets. Particles were initially classified using $K = 20$ classes and angular sampling of 5°. Several rounds of 2D classification allowed for removal of poorly aligning particles. Final 2D classifications were performed using $K = 4$ with a final particle count of 4,191. RELION extracted particles were also imported into cryoSPARC 2.12.2 for 2D classification. Selected 2D classes containing 5,217 particles were used as inputs for an *ab initio* reconstruction. The resulting best class and volume containing 3,436 particles were used for homogenous refinement.

ClusPro model generation. The docking model of OKT9-Fab to hTfR1 was performed using the web server ClusPro docking algorithm (18). This software runs a rigid body docking, followed by a root mean standard deviations clustering ranked by the number of structures with the lowest energy and refined by energy minimization to remove steric clashes (18). The structures of OKT9-Fab and hTfR1 were uploaded to the ClusPro v2.0 server in antibody-antigen mode for docking (42). The hTfR1 monomeric structure extracted from the PDB entry 3KAS was defined as the ligand, while the newly determined structure of OKT9-Fab with a modeled Y103 side chain was defined as the receptor. To minimize the search region for docking, we relied on the 3D density of the complex of OKT9-hTfR1. Based on this density, we identified the overall region of the hTfR1 apical domain bound by OKT9. We selected an attraction mask that included hTfR1 surface residues D204-L212, N348-E350, S368-K371, and K374. We also applied a repulsion mask surrounding this interaction area on the hTfR1 apical domain that included the following surface residues: V228-L232, Y247-G252, I277, A340-E343, G351, D352, and M365-T367. For OKT9-Fab, we used a repulsion mask that excluded the surface residues of the heavy- and light-chain CDRs and the surface residues surrounding them. The repulsion mask included light-chain residues D1-C23, L39-K51, L53, I54, G63-G72, T78-C94, and G105-E219 and the heavy-chain residues Q1-A24, M34, W36-E46, I48, G49, T58, Y60-S76, T78-A97, and W109-R219. Residues in the attraction and repulsion masks were identified using PyMOL.

Fit of ClusPro models into 3D density for OKT9-hTfR1 complex. 3D density corresponding to the final homogenous refinement map was loaded onto Chimera for further analysis. The ectodomain of the dimeric human transferrin crystal structure (PDB entry 1CX8) was manually placed within the map density, followed by the fit-in-map function to finalize placement of the structure. Monomeric hTfR1 (PDB entry 3KAS) and ClusPro inputs were placed using MatchMaker, given the dimer structure as a reference. All generated ClusPro models were subsequently fit using MatchMaker and 3KAS as a reference and were later manually inspected for fit onto the 3D density.

Data availability. The coordinates for the structure of OKT9-Fab have been deposited in the Worldwide Protein Data Bank (wwPDB) with accession code 6WX1.

ACKNOWLEDGMENTS

We thank Michael Sawaya and Duilio Cascio (UCLA). BHK cells producing the ectodomain of the hTfR1 in soluble form (BHK-sTfR1) were a generous gift from Manuel Penichet and Tracy Daniels (UCLA).

This work was performed as part of STROBE, an NSF Science and Technology Center through grant DMR-1548924. This work was also supported by DOE grant DE-FC02-02ER63421 and NIH-NIGMS grants R35 GM128867, 5R21AI135851-02, and P41GM136508, Programa Bec.Ar, Fundación Bunge y Born, INC 2018-2020, PID 2015-0034, IP-COVID19 939, Fundación Williams, PICT 2018-3954, PICT 2017-1136, 20020190100013BA, and Fundación René Barón. G.H., C.C.G., and S.C. are members of Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). S.F., C.P., and C.A.V. were funded by CONICET fellowships, and M.D.F. was funded by Ruth L. Kirschstein NRSA grant GM007185, a Whitcome predoctoral fellowship, and a National Science Foundation graduate research fellowship. J.A.R. is supported as a Searle Scholar, a Pew Scholar, a Beckman Young Investigator, a Packard Fellow, and a Sloan Fellow.

We have no competing interests to declare.

G.H., S.M.C., C.C.G., J.A., and J.A.R. directed the project; J.Z., S.Z., C.S., and D.F. produced and purified proteins; S.Z. and C.S. grew crystals and performed crystallographic analyses; L.E.C. and J.A. performed biolayer assays; S.F., G.H., C.S., D.F., J.A.R., and M.C. performed binding and kinetics analyses; S.F., M.V.B., and C.P. performed pseudovirus cellular uptake assays; S.M.C., C.C.G., and C.A.V. performed infectivity assays; G.H., J.A.R., and M.D.F. performed electron microscopy; M.D.F. performed single-particle analysis of electron microscopy images; G.H., M.D.F., and

S.F. performed computational docking and modeling. All authors contributed to writing of the article.

REFERENCES

1. Shao J, Liang Y, Ly H. 2015. Human hemorrhagic fever causing arenaviruses: molecular mechanisms contributing to virus virulence and disease pathogenesis. Pathogens 4:283–306. https://doi.org/10.3390/pathogens4020283.
2. Borio L, Inglesby T, Peters CJ, Schmaljohn AL, Hughes JM, Jahrling PB, Ksiazek T, Johnson KM, Meyerhoff A, O'Toole T, Ascher MS, Bartlett J, Breman JG, Eitzen EM, Jr, Hamburg M, Hauer J, Henderson DA, Johnson RT, Kwik G, Layton M, Lillibridge S, Nabel GJ, Osterholm MT, Perl TM, Russell P, Tonat K, Working Group on Civilian Biodefense. 2002. Hemorrhagic fever viruses as biological weapons: medical and public health management. JAMA 287:2391–2405. https://doi.org/10.1001/jama.287.18.2391.
3. Centers for Disease Control and Prevention. 2012. Possession, use, and transfer of select agents and toxins; biennial review. Final rule. Fed Regist 77:61083–61115.
4. Zeitlin L, Geisbert JB, Deer DJ, Fenton KA, Bohorov O, Bohorova N, Goodman C, Kim D, Hiatt A, Pauly MH, Velasco J, Whaley KJ, Altmann F, Gruber C, Steinkellner H, Honko AN, Kuehne AI, Aman MJ, Sahandi S, Enterlein S, Zhan X, Enria D, Geisbert TW. 2016. Monoclonal antibody therapy for Junin virus infection. Proc Natl Acad Sci U S A 113:4458–4463. https://doi.org/10.1073/pnas.1600996113.
5. Westover JB, Sefing EJ, Bailey KW, Van Wettere AJ, Jung KH, Dagley A, Wandersee L, Downs B, Smee DF, Furuta Y, Bray M, Gowen BB. 2016. Low-dose ribavirin potentiates the antiviral activity of favipiravir against hemorrhagic fever viruses. Antiviral Res 126:62–68. https://doi.org/10.1016/j.antiviral.2015.12.006.
6. Mahmutovic S, Clark L, Levis SC, Briggiler AM, Enria DA, Harrison SC, Abraham J. 2015. Molecular basis for antibody-mediated neutralization of new world hemorrhagic fever mammarenaviruses. Cell Host Microbe 18:705–713. https://doi.org/10.1016/j.chom.2015.11.005.
7. Radoshitzky SR, Abraham J, Spiropoulou CF, Kuhn JH, Nguyen D, Li W, Nagel J, Schmidt PJ, Nunberg JH, Andrews NC, Farzan M, Choe H. 2007. Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. Nature 446:92–96. https://doi.org/10.1038/nature05539.
8. Abraham J, Corbett KD, Farzan M, Choe H, Harrison SC. 2010. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. Nat Struct Mol Biol 17:438–444. https://doi.org/10.1038/nsmb.1772.
9. Helguera G, Jemielity S, Abraham J, Cordo SM, Martinez MG, Rodriguez JA, Bregni C, Wang JJ, Farzan M, Penichet ML, Candurra NA, Choe H. 2012. An antibody recognizing the apical domain of human transferrin receptor 1 efficiently inhibits the entry of all new world hemorrhagic fever arenaviruses. J Virol 86:4024–4028. https://doi.org/10.1128/JVI.06397-11.
10. Cohen-Dvashi H, Amon R, Agans KN, Cross RW, Borenstein-Katz A, Mateo M, Baize S, Padler-Karavani V, Geisbert TW, Diskin R. 2020. Rational design of universal immunotherapy for TfR1-tropic arenaviruses. Nat Commun 11:67. https://doi.org/10.1038/s41467-019-13924-6.
11. Sutherland R, Delia D, Schneider C, Newman R, Kemshead J, Greaves M. 1981. Ubiquitous cell-surface glycoprotein on tumor cells is proliferation-associated receptor for transferrin. Proc Natl Acad Sci U S A 78:4515–4519. https://doi.org/10.1073/pnas.78.7.4515.
12. Reinherz EL, Kung PC, Goldstein G, Levey RH, Schlossman SF. 1980. Discrete stages of human intrathymic differentiation: analysis of normal thymocytes and leukemic lymphoblasts of T-cell lineage. Proc Natl Acad Sci U S A 77:1588–1592. https://doi.org/10.1073/pnas.77.3.1588.
13. Panaccio M, Zalcberg JR, Thompson CH, Leyden MJ, Sullivan JR, Lichtenstein M, McKenzie IF. 1987. Heterogeneity of the human transferrin receptor and use of anti-transferrin receptor antibodies to detect tumours in vivo. Immunol Cell Biol 65:461–472. https://doi.org/10.1038/icb.1987.55.
14. de Mello Malta F, Amgarten D, Nastri A, Ho YL, Boas Casadio LV, Basqueira M, Selegatto G, Cervato MC, Duarte-Neto AN, Higashino HR, Faustino Medeiros FA, Gendler J, Levin AS, Pinho JRR. 2020. Sabia virus-like mammarenavirus in patient with fatal hemorrhagic fever, Brazil, 2020. Emerg Infect Dis 26:1332–1334. https://doi.org/10.3201/eid2606.200099.
15. Contigiani MS, Sabattini MS. 1977. Virulencia diferencial de cepas de virus Junin por marcadores biológicos en ratones y cobayos. Medicina 37:244–251.
16. Byrne SL, Leverence R, Klein JS, Giannetti AM, Smith VC, MacGillivray RT, Kaltashov IA, Mason AB. 2006. Effect of glycosylation on the function of a soluble, recombinant form of the transferrin receptor. Biochemistry 45:6663–6673. https://doi.org/10.1021/bi0600695.
17. Neiveyans M, Melhem R, Arnoult C, Bourquard T, Jarlier M, Busson M, Laroche A, Cerutti M, Pugniere M, Ternant D, Gaborit N, Chardes T, Poupon A, Gouilleux-Gruart V, Pelegrin A, Poul MA. 2019. A recycling anti-transferrin receptor-1 monoclonal antibody as an efficient therapy for erythroleukemia through target up-regulation and antibody-dependent cytotoxic effector functions. MAbs 11:593–605. https://doi.org/10.1080/19420862.2018.1564510.
18. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. 2017. The ClusPro web server for protein-protein docking. Nat Protoc 12:255–278. https://doi.org/10.1038/nprot.2016.169.
19. De Feo CJ, Weiss CD. 2012. Escape from human immunodeficiency virus type 1 (HIV-1) entry inhibitors. Viruses 4:3859–3911. https://doi.org/10.3390/v4123859.
20. Cheng Y, Zak O, Aisen P, Harrison SC, Walz T. 2004. Structure of the human transferrin receptor-transferrin complex. Cell 116:565–576. https://doi.org/10.1016/S0092-8674(04)00130-8.
21. Amanat F, Duehr J, Huang C, Paessler S, Tan GS, Krammer F. 2020. Monoclonal antibodies with neutralizing activity and Fc-effector functions against the Machupo virus glycoprotein. J Virol 94:e01741-19. https://doi.org/10.1128/JVI.01741-19.
22. Clark LE, Mahmutovic S, Raymond DD, Dilanyan T, Koma T, Manning JT, Shankar S, Levis SC, Briggiler AM, Enria DA, Wucherpfennig KW, Paessler S, Abraham J. 2018. Vaccine-elicited receptor-binding site antibodies neutralize two New World hemorrhagic fever arenaviruses. Nat Commun 9:1884. https://doi.org/10.1038/s41467-018-04271-z.
23. Montemiglio LC, Testi C, Ceci P, Falvo E, Pitea M, Savino C, Arcovito A, Peruzzi G, Baiocco P, Mancia F, Boffi A, Des Georges A, Vallone B. 2019. Cryo-EM structure of the human ferritin-transferrin receptor 1 complex. Nat Commun 10:1121. https://doi.org/10.1038/s41467-019-09098-w.
24. Goncalves AR, Moraz ML, Pasquato A, Helenius A, Lozach PY, Kunz S. 2013. Role of DC-SIGN in Lassa virus entry into human dendritic cells. J Virol 87:11504–11515. https://doi.org/10.1128/JVI.01893-13.
25. Martinez MG, Bialecki MA, Belouzard S, Cordo SM, Candurra NA, Whittaker GR. 2013. Utilization of human DC-SIGN and L-SIGN for entry and infection of host cells by the New World arenavirus, Junin virus. Biochem Biophys Res Commun 441:612–617. https://doi.org/10.1016/j.bbrc.2013.10.106.
26. Fedeli C, Torriani G, Galan-Navarro C, Moraz ML, Moreno H, Gerold G, Kunz S. 2018. Axl can serve as entry factor for Lassa virus depending on the functional glycosylation of dystroglycan. J Virol 92:e01613-17. https://doi.org/10.1128/JVI.01613-17.
27. Brouillette RB, Phillips EK, Patel R, Mahauad-Fernandez W, Moller-Tank S, Rogers KJ, Dillard JA, Cooney AL, Martinez-Sobrido L, Okeoma C, Maury W. 2018. TIM-1 mediates dystroglycan-independent entry of Lassa virus. J Virol 92:e00093-18. https://doi.org/10.1128/JVI.00093-18.
28. Brennan FR, Morton LD, Spindeldreher S, Kiessling A, Allenspach R, Hey A, Muller PY, Frings W, Sims J. 2010. Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies. MAbs 2:233–255. https://doi.org/10.4161/mabs.2.3.11782.
29. Weissman AM, Klausner RD, Rao K, Harford JB. 1986. Exposure of K562 cells to anti-receptor monoclonal antibody OKT9 results in rapid redistribution and enhanced degradation of the transferrin receptor. J Cell Biol 102:951–958. https://doi.org/10.1083/jcb.102.3.951.
30. Bien-Ly N, Yu YJ, Bumbaca D, Elstrott J, Boswell CA, Zhang Y, Luk W, Lu Y, Dennis MS, Weimer RM, Chung I, Watts RJ. 2014. Transferrin receptor (TfR) trafficking determines brain uptake of TfR antibody affinity variants. J Exp Med 211:233–244. https://doi.org/10.1084/jem.20131660.
31. Abraham J, Kwong JA, Albarino CG, Lu JG, Radoshitzky SR, Salazar-Bravo J, Farzan M, Spiropoulou CF, Choe H. 2009. Host-species transferrin receptor 1 orthologs are cellular receptors for nonpathogenic new world clade B arenaviruses. PLoS Pathog 5:e1000358. https://doi.org/10.1371/journal.ppat.1000358.

32. Gruszczyk J, Huang RK, Chan LJ, Menant S, Hong C, Murphy JM, Mok YF, Griffin MDW, Pearson RD, Wong W, Cowman AF, Yu Z, Tham WH. 2018. Cryo-EM structure of an essential Plasmodium vivax invasion complex. Nature 559:135–139. https://doi.org/10.1038/s41586-018-0249-1.
33. Mazumdar S. 2009. Raxibacumab. MAbs 1:531–538. https://doi.org/10.4161/mabs.1.6.10195.
34. Greig SL. 2016. Obiltoxaximab: first global approval. Drugs 76:823–830. https://doi.org/10.1007/s40265-016-0577-0.
35. Bettiker RL, Koren DE, Jacobson JM. 2018. Ibalizumab. Curr Opin HIV AIDS 13:354–358. https://doi.org/10.1097/COH.0000000000000473.
36. Aricescu AR, Lu W, Jones EY. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62:1243–1250. https://doi.org/10.1107/S0907444906029799.
37. Lawrence CM, Ray S, Babyonyshev M, Galluser R, Borhani DW, Harrison SC. 1999. Crystal structure of the ectodomain of human transferrin receptor. Science 286:779–782. https://doi.org/10.1126/science.286.5440.779.
38. Sahtoe DD, Coscia A, Mustafaoglu N, Miller LM, Olal D, Vulovic I, Yu TY, Goreshnik I, Lin YR, Clark L, Busch F, Stewart L, Wysocki VH, Ingber DE, Abraham J, Baker D. 2021. Transferrin receptor targeting by de novo sheet extension. Proc Natl Acad Sci U S A 118:e2021569118. https://doi.org/10.1073/pnas.2021569118.
39. Kabsch W. 2010. Xds. Acta Crystallogr D Biol Crystallogr 66:125–132. https://doi.org/10.1107/S0907444909047337.
40. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75:861–877. https://doi.org/10.1107/S2059798319011471.
41. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.
42. Brenke R, Hall DR, Chuang GY, Comeau SR, Bohnuud T, Beglov D, Schueler-Furman O, Vajda S, Kozakov D. 2012. Application of asymmetric statistical potentials to antibody-protein docking. Bioinformatics 28:2608–2614. https://doi.org/10.1093/bioinformatics/bts493.
43. Bennett MJ, Lebron JA, Bjorkman PJ. 2000. Crystal structure of the hereditary haemochromatosis protein HFE complexed with transferrin receptor. Nature 403:46–53. https://doi.org/10.1038/47417.
44. Maier KE, Jangra RK, Shieh KR, Cureton DK, Xiao H, Snapp EL, Whelan SP, Chandran K, Levy M. 2016. A new transferrin receptor aptamer inhibits new world hemorrhagic fever mammarenavirus entry. Mol Ther Nucleic Acids 5:e321. https://doi.org/10.1038/mtna.2016.32.
45. Zong M, Fofana I, Choe H. 2014. Human and host species transferrin receptor 1 use by North American arenaviruses. J Virol 88:9418–9428. https://doi.org/10.1128/JVI.01112-14.
46. Cardenas S, Auais A, Piedimonte G. 2005. Palivizumab in the prophylaxis of respiratory syncytial virus infection. Expert Rev Anti Infect Ther 3:719–726. https://doi.org/10.1586/14787210.3.5.719.

Appendix Chapter 2

This chapter is based on the preprint “Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at its Core“.

Richards, L. S., **Flores, M. D.**, Zink, S., Schibrowsky, N. A., Sawaya, M. R., & Rodriguez, J. A. (2023). Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at its Core. BioRxiv : The Preprint Server for Biology. https://doi.org/10.1101/2023.02.08.527771

Title: Cryo-EM Structure of a Human LECT2 Amyloid Fibril Reveals a Network of Polar Ladders at its Core

Authors: Logan S. Richards[1], Maria D. Flores[1], Samantha Zink[1], Natalie A. Schibrowsky[1], Michael R. Sawaya[1] and Jose A. Rodriguez[1*]

Affiliations:

[1] Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center; University of California, Los Angeles (UCLA); Los Angeles, CA 90095, USA.

* Correspondence to Jose A. Rodriguez (jrodriguez@mbi.ucla.edu)

Summary

ALECT2 systemic amyloidosis is caused by deposition of the leukocyte cell-derived chemotaxin-2 (LECT2) protein in the form of fibrils. In ALECT2 amyloidosis, ALECT2 fibril deposits can be found in the glomerulus, resulting in renal failure. Affected patients lack effective treatment options outside of renal transplant or dialysis. While the structure of LECT2 in its globular form has been determined by X-ray crystallography, structures of ALECT2 amyloid fibrils remain unknown. Using single particle cryo-EM, we now find that human LECT2 forms robust twisting fibrils with canonical amyloid features. At their core, ALECT2 fibrils contain two mating protofilaments, the ordered core of each protofilament spans residues 55-75 of the LECT2 sequence. The overall geometry of the ALECT2 fibril displays features in line with other pathogenic amyloids. Its core is tightly packed and stabilized by a network of hydrophobic contacts and hydrogen-bonded uncharged polar residues, while its outer surface displays several charged residues. The robustness of ALECT2 fibril cores is illustrated by their limited dissolution in 3M urea and their persistence after treatment

with proteinase K. As such, the ALECT2 fibril structure presents a potential new target for treatments against ALECT2 systemic amyloidosis.

Key words

Amyloid; Cryo-EM; ALECT; LECT2; kidney disease

Significance statement

Amyloid diseases are serious conditions characterized by the formation and deposition of insoluble protein fibrils in various tissues. ALECT2 systemic amyloidosis is a recently identified amyloid disease that is characterized by the deposition of amyloid fibrils in the kidneys composed of misfolded human LECT2 protein. ALECT2 amyloidosis inevitably leads to renal failure in patients and to date there exist no treatments for the disease and limited structural information on the amyloid state of the LECT2 protein. Here, we begin to close this gap in knowledge by using cryo-EM to determine the first high-resolution structure of the human protein LECT2 in an amyloid fold and identify key regions of interaction that may be important for the fibril stability. This work lays a foundation of structural knowledge upon which we hope future treatments for ALECT2 amyloidosis can be developed.

Main Text

Introduction

Amyloid diseases are linked to the formation and persistence of large, multimeric structures in various tissues. Amyloid fibrils are characterized by a cross-beta scaffold in which identical protein

molecules mate tightly as beta-strands to form a long, unbranched fibril[1,2]. Before amyloid-forming proteins assemble into fibrils, they can display a globular fold, requiring partial or total unfolding to convert into an amyloid state[3,4]. Debilitating neurodegenerative diseases such as Alzheimer's and Parkinson's disease are most well-known to involve amyloidogenesis[5,6], while many other examples of amyloid diseases involve the systemic deposition of persistent amyloid fibrils throughout various organs including the heart, liver, kidneys, skin, digestive tract, and nervous system[7]. While the persistence of amyloid aggregates is a unifying feature of these diseases, the factors contributing to the misfolding, retention and toxicity of each disease-associated fibril are distinct and likely dependent on their structures[8]. This underscores drive to characterize amyloid fibril structures.

In 2008, Benson *et al*. identified a new form of systemic amyloid disease associated with impaired renal function, renal failure and nephrotic syndrome[9]. Biochemical characterization of the fibrils observed in the glomeruli of the kidneys identified the protein leukocyte cell-derived chemotaxin-2 (LECT2) as a major component of the fibrils, inspiring the disease name, ALECT2 amyloidosis[9]. Since its identification, ALECT2 amyloidosis has been characterized as a common and likely underdiagnosed form of renal amyloidosis[10] with high prevalence among patients of Mexican descent in the southwest United States, those of Punjabi descent, First Nations peoples in British Columbia, Egyptians, Chinese of Han ethnicity, and Native Americans[11]. There remains a need for diagnostic tools to distinguish ALECT2 amyloidosis from other renal amyloid diseases in patients and avoid misdiagnosis and improper treatment[12]. In contrast to some other systemic amyloid diseases, there are no available treatments for ALECT2 amyloidosis and no molecules that effectively target the amyloid state of its namesake.

Nascent LECT2 is a 151 amino acid polypeptide with an 18-residue N-terminal signal peptide that is cleaved before its secretion into the blood as a globular protein[13]. In circulation, LECT2 performs multiple biological functions. It can act as a chemotactic factor to promote the migration of

neutrophils to sites of infection[13,14]. It can also act as a regulator of cartilage growth[15], and it can act as a hepatokine important for metabolic homeostasis[16]. Structurally, LECT2 displays an M23 metalloendopeptidase fold that coordinates Zn(II), but lacks a critical histidine residue that would enable its catalysis[17]. While the clinical details of ALECT2 amyloid disease have been well characterized, the molecular details of ALECT2 amyloid aggregation remain unclear. Loss of this bound Zn(II) may play an important role in the conversion of properly folded LECT2 into an amyloid form *in vitro* as it is thought to destabilize a central beta-barrel motif within the protein[18]. Many patients who suffer from ALECT2 amyloidosisare also homozygous for the common I40V sequence polymorph of the protein[19]. This mutation appears to be important, but not sufficient for inducing amyloid formation, and does not destabilize the metal binding properties of the functional protein; accordingly, its role in amyloid conversion remains unclear[18]. Amyloidogenic segments of LECT2 have been identified, and predicted amyloid-forming LECT2 peptides assemble into amyloid-like fibrils in isolation[20]. However, the identity of the ALECT2 amyloid core remains unknown.

Recent advances in single-particle cryo-EM using helical reconstruction have permitted the determination of near-atomic resolution structures of a wide variety of amyloid fibrils[3]. These approaches have now allowed us to determine the structure of amyloid fibrils formed by recombinant full-length LECT2. This ~2.7Å resolution cryo-EM structure of an ALECT2 fibril reveals a tightly mated amyloid core that spans residues 55 to 75 and harbors a network of polar ladders. Its structure is also a first step toward a more comprehensive molecular understanding of ALECT2 amyloidosis and a starting point for the design of targeted therapeutics or diagnostic agents.

Results

Assembly of recombinant human ALECT2 amyloid fibrils

Amyloid fibrils were grown from recombinantly expressed, full-length LECT2 protein encoding the I40V sequence polymorph. Purified LECT2 was allowed to assemble into fibrils over a 48-hour period, while shaking at 2400 rpm at a temperature of 37°C. After an initial lag phase, fibrils grew at an exponential rate and then stabilized, as monitored by changes in Thioflavin T (ThT) fluorescence over a 48-hour time course (Figure S1D). Unbranched fibrils were observed by negative stain transmission electron microscopy in samples prepared in an equivalent manner, but without ThT (Figure S1E). The ALECT2 fibrils exhibited a regular helical twist with an ~850Å spacing between crossovers. These features were consistent with other amyloid fibrils, and suggested that ALECT2 fibrils harbored a canonical amyloid core supported by steric zipper motifs. Indeed, certain regions of the LECT2 sequence were predicted to have a high propensity to form steric zippers (Figure S1A), and the amyloid nature of the fibrils was confirmed by the appearance of a signature cross-beta diffraction pattern when X-ray diffraction was collected from aligned ALECT2 fibrils (Figure S2A).

ALECT2 amyloid fibrils are urea and protease-resistant

Amyloid aggregates are often highly stable and can resist chemical or proteolytic denaturation. To evaluate the stability of ALECT2 fibrils, a nephelometric assay was performed that monitored light scattering induced by fibrils in solution. Fibril dissolution was correlated with decreased nephelometry signal and assessed under conditions where fibrils were exposed to either water, fibrillization buffer alone, fibrillization buffer with 3M urea, or 40 nM Proteinase K in 150 mM NaCl, 50 mM MOPS pH 6.5 buffer (Figure S4A). LECT2 fibrils persisted in solution when exposed to fibrillization buffer or water over a 24-hour period. In contrast, exposure to 3M urea induced an immediate, slight decrease in fibril content followed by a stable lower signal over the 24-hour period. Incubation with Proteinase K induced a larger drop in signal over the first two hours of incubation following a stable, lower signal phase during the remaining incubation period. These initial drops in signal could have resulted from the dissolution of larger fibril clumps or the removal

of more susceptible fibrils, but in all cases, there remained a stable subset of fibrils in solution (Figure S4B-E). The presence of a species that was resistant to protease digestion correlated with the appearance of a ~7kDa band as observed by SDS PAGE (Figure S9A). The identity of this fragment was investigated 60min after exposing fibrils to the protease using bottom-up mass spectrometry. That analysis identified segments covering 25% of the LECT2 sequence (Supplementary Table 2), including a fragment of the ALECT2 fibril core spanning residues N67-R74 and portions of its lysine-rich C-terminal region (Figure S9A).

Structural determination of an ALECT2 amyloid fibril core

After biochemical characterization, ALECT2 fibrils were prepared for high-resolution structure determination by single particle cryo-EM. The conditions for grid preparation were optimized to retain regularly twisting fibrils and discourage their adsorption to the air-water interface (Figure S4A and S5). Cryo-electron micrographs of these fibrils produced 2D class averages with fibril characteristics consistent with negative stain images (Figure S3A). Frozen hydrated fibrils exhibited one of two consistent helical crossovers. Fibrils with a 400Å helical crossover, which we termed the fast-twisting polymorph, accounted for only about 10% of fibrils present in the sample (Figure S3C). The remaining 90% of fibrils exhibited an 850Å helical crossover and were termed the slow-twisting polymorph (Figure S3D). The two polymorphs share very similar features apart from their different crossover distances, but given its abundance, the slow-twisting polymorph proved more amenable to high-resolution 2D classification and 3D structure determination. Subdivision of the fibril into two symmetry-related protofilaments could be inferred by the appearance of mirror symmetry in 2D class averages. Fibril images gave the impression of being mirrored across the fibril axis when it aligned parallel to the image plane. In this view, very near the fibril axis, the density of one protofilament appeared staggered relative to the other, resembling the teeth of a zipper (Figure S3E and S3F). These observations indicated that the two protofilaments might be related by a symmetry that approximated a 2_1-screw axis. X-ray fibril

diffraction from the ALECT2 fibrils showed a strong 4.69Å reflection; corresponding to the true spacing between ALECT2 molecules within a protofilament (Figure S2A and S2B). This spacing was confirmed by analysis of Fourier Transforms from high-resolution images of the fibrils (Figure S3B). The 2D classes were sufficient to generate a *de novo* 3D initial model in RELION and this map was used as an initial model for 3D classification (Figure S3G and S3H).

A three-dimensional reconstruction from the ALECT2 fibril images displayed an ordered amyloid core with an estimated resolution of ~2.7Å, based on the 0.143 Fourier shell correlation (FSC) criterion (Figure S6B). The map was of sufficient quality to allow for building of an unambiguous atomic model. There were no breaks in connectivity of the tube of density attributed to each molecule and there was clear separation between layers of density corresponding to individual LECT2 molecules, whether stacked along the fibril axis or mating with the opposite protofilament (Figure 1A and 1B). Further, the resolution of the map allowed for the unambiguous assignment of side chains and peptide backbone oxygen atoms, allowing *de novo* assignment of the residues at its core (Figure 1C)[21,22]. The modelled portion of the amyloid core exhibited a clear pseudo-2_1-screw symmetry with the repeating unit being a 21-residue sequence stretching from Methionine 55 to Isoleucine 75 (Figure 2A and 2B). To assure that the assignment of this sequence was correct, Rosetta was used to calculate configurations for all possible 21 residue segments of LECT2 constrained to match the observed density. This was done by threading each sequence onto the peptide backbone of the model to test which threading segment produces the most stable structure. This analysis confirmed that the M55-I75 sequence was the most energetically favorable sequence assignment for the geometry of the fibril core (Figure S7A).

Stability-promoting molecular features in the ALECT2 fibril core

At the core of the ALECT2 fibril, a segment from each of two protofilaments mate in a largely dry interface. The two protofilaments are nearly 2-fold symmetric in the fibril core, allowing hydrophobic zippers to coalesce a network of uncharged polar residues stabilized by polar

ladders. Each hydrophobic pocket within the fibril core contains tightly packed steric zipper structures composed of residues I56, V73, and I75; residue I69 anchors the very center of the fibril core near the helical axis (Figure 3A and 3B). A pair of glycines, G58 and G72, spaced only 4Å apart segregate the hydrophobic patches at the edges of the fibril core from the network of polar residues within it (Figure 3A). These polar residues include Q59, N65, N67, and N71, all of which are incorporated into stabilizing polar ladders (Figure 3C). Residues N67 and N71 form a polar clasp that holds together not only layers above and below but also stitches together the molecules from opposite protofilaments (Figure 3C). This polar clasp also seals off the central pair of I69 residues and pins together the two largest beta-strands in the fibril structure. That combination of features results in an overall energetically favorable structure with a calculated standard free energy of solvation[2,23] of ΔG_o of -15.2 kcal/mol per layer and -0.36 kcal/mol per residue (Figure 2C). The solvation energy and small size of the ALECT2 amyloid core are similar to that of the human RIPK1-RIPK3 hetero-amyloid, which is hypothesized to form stable aggregates *in vivo* and signal TNF-induced necroptosis in cells[24]. These metrics could be further influenced by other stabilizing interactions, such as disulfide bridges or metal coordination sites, that in the present structure are absent from the fibril core, but exist within the fibril fuzzy coat and remain unresolved.

Discussion

In search of amyloid folds adopted by the LECT2 protein, we find that recombinant LECT2 polypeptides encoding the I40V sequence polymorphism readily form amyloid fibrils that display a canonical beta-sheet rich architecture with stacked layers of the LECT2 protein separated by a rise of 4.69 Å per layer and obeying a regular, left-handed, helical twist. The amyloid core resolved

within these fibrils is composed of two identical protofilaments that each contain 21 of the 133 residues in full-length LECT2: from Methionine 55 to Isoleucine 75. Interestingly, this sequence has not been predicted by previous work to form amyloid fibrils. However, other studies identified the regions immediately preceding and following it as possible amyloidogenic segments[20], and the newly identified core segment is predicted to form steric zipper structures by the amyloid-prediction algorithm in ZipperDB[25](Figure S1A).

While a significant fraction of the LECT2 sequence is predicted to be amyloidogenic, much of the protein's globular fold is stabilized by three disulfide bonds, two in the N-terminal region and one in C-terminal region (Figure 2A). These bonds would likely prevent many of the predicted amyloidogenic regions from assembling into fibrils. However, the core segment identified here is not involved in any of the native disulfide bonds and represents a portion of the central beta-barrel motif of LECT2[17]. Specifically, the ALECT2 fibril core encompasses most of beta-strands 3 and 4 and all of loop 2 in the LECT2 globular fold (Figure S8A). That loop is solvent-facing and sits adjacent to the Zn(II) binding pocket[17]. This same region of the protein has previously been hypothesized to be susceptible to reductions in pH that could facilitate the loss of the bound zinc coordinated to the first and second beta-strands and may destabilize the beta-barrel structure, leading to the exposure of amyloidogenic segments[18].

A comparison of native and fibrillar ALECT2 structures reveals that the beta-strand conformation of N- and C-terminal segments of the fibrillar fold are retained from beta-strands 3 and 4 of the globular fold. However, the central segment of the fibril core represents a large ordering of secondary structure, namely rearrangement of loop 2 into 3 beta-strand segments, the largest of which stretches from N65 to N71 (Figure 2B). The outer surface of the ALECT2 fibril is decorated with larger amino acids including basic residues K61, K66, and R74 as well as one acidic residue, E60, and a pi-stacked Y63 (Figure 2B). These residues are also in solvent-exposed positions in the globular fold of LECT2. Residues I56, V73, and I75, which form the outer steric zipper

segments of the fibril, are all buried within the central beta-barrel of globular LECT2, and I69, the central hydrophobic residue, is natively buried as part of loop 2 abutted to alpha-helix 1 (Figure S8B). In contrast, the polar-ladder-forming residues Q59, N65, N67, and N71 undergo a large conformational and environmental change from the globular fold. All four of these residues are solvent-facing in loop 2 of the globular structure of LECT2, and all are packed into a relatively dry interface in the fibril structure (Figure S8B). Polar ladder conformations like those revealed in this structure have long been hypothesized to stabilize amyloid assemblies[26,27] and were also favored during energy minimization of the ALECT2 fibril fold in Rosetta (Figure S7B). The structural transition to a polar ladder within a dry environment confers increased stability by joining chains along the fibril axis while fully satisfying hydrogen bonds within the fibril core. This stabilizing effect is further amplified when the two polar ladders in separate protofilaments interdigitate into a zipper conformation, as is observed between residues N67 and N71 in the ALECT2 fibril.

The structural conversion of the loop 2 region of globular LECT2 plays a central role in generating the ALECT2 amyloid core observed here, and the dimeric nature of the fibril draws some comparisons to its globular counterpart. In addition to the dimer resolved in the crystal structure of LECT2, previous studies have observed that stable, SDS resistant, dimers and higher order oligomers of LECT2 readily form with or without the presence of zinc[28]. One hypothesis for this dimerization is that LECT2 could undergo domain swapping. This arrangement has been observed for amyloid proteins previously, such as with β_2-microglobulin, which has been shown to form domain-swapped amyloid fibrils with disulfide linkages[29]. Interestingly, the amyloid core arrangement of loop 2 spans the gap between monomers of LECT2 in the crystallographic dimer well. A model of a domain-swapped LECT2 dimer through the loop 2 region using the amyloid core structure resolved here shows that this region could facilitate the domain-swapped conformation with minimal disruption to the rest of the globular LECT2 fold (Figure S10A). Further,

this conformation maintains the general amyloid core arrangement and could adopt the fibrillar fold following destabilization of the rest of the dimeric interface (Figure S10B).

It remains to be seen whether recombinant LECT2 amyloid fibrils exactly match those present in the kidneys of ALECT2 systemic amyloidosis patients; however, structures of the former and the segments at their core provide a first glimpse at stabilizing interactions within ALECT2 fibrils. Further, the exact role of the I40V polymorphism in amyloid aggregation remains unclear as the residue was not resolved as part of the ALECT2 amyloid core. The segment at the core of this recombinant ALECT2 fibril is derived from part of the central beta-barrel structure of globular LECT2, supporting the theory that loss of zinc may destabilize this portion of the protein, exposing the amyloidogenic segments resolved here and allowing for aggregation. This information opens the possibility for rational design of aggregation inhibitors which could either stabilize this region of globular LECT2 or inhibit the extension of fibrils of ALECT2 by targeting residues at its core.

Acknowledgements

This work is supported by DOE Grant DE-FC02-02ER63421, NSF Grant DMR-1548924, the NIH-NIGMS Grant R35 GM128867. L.S.R. is supported by the USPHS National Research Service Award 5T32GM008496. J.A.R. is also supported as a Sloan Fellow, a Pew Scholar, and a Packard Fellow. Part of this work was performed at the Stanford-SLAC Cryo-EM Center (S2C2). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the following S2C2 personnel for their invaluable support and assistance: Dr. Patrick Mitchell, Dr. Yee-Ting Lee, and Professor Wah Chiu who are supported by the National Institutes of Health Common Fund Transformative High-Resolution Cryo-Electron Microscopy program (U24 GM129541). In addition, we thank the staff of the UCLA Molecular Instrumentation Center (MIC), supported by the UCLA Division of Physical Sciences.

Disclosure Statement

All the authors declared no competing interests.

Data Statement

All data are available in either the main text or the Supplementary Material. The reconstructed cryo-EM map was deposited in the Electron Microscopy Data Bank with the accession code EMD-29682, while coordinates for the refined atomic model were deposited in the Protein Data Bank under the accession code 8G2V.

References

1. Eisenberg, D., and Jucker, M. (2012). The Amyloid State of Proteins in Human Diseases. Cell *148*, 1188–1203. 10.1016/j.cell.2012.02.022.

2. Sawaya, M.R., Hughes, M.P., Rodriguez, J.A., Riek, R., and Eisenberg, D.S. (2021). The expanding amyloid family: Structure, stability, function, and pathogenesis. Cell *184*, 4857–4873. 10.1016/j.cell.2021.08.013.

3. Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., and Scheres, S.H.W. (2017). Cryo-EM structures of tau filaments from Alzheimer's disease. Nature *547*, 185–190. 10.1038/nature23002.

4. Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., Westermark, P., Zacharias, M., and Fändrich, M. (2019). Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. Nat Commun *10*, 5008. 10.1038/s41467-019-13038-z.

5. Selkoe, D.J. (1991). The molecular pathology of Alzheimer's disease. Neuron *6*, 487–498. 10.1016/0896-6273(91)90052-2.

6. Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M., and Goedert, M. (1998). α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia

with Lewy bodies. Proceedings of the National Academy of Sciences *95*, 6469–6473. 10.1073/pnas.95.11.6469.

7. Muchtar, E., Dispenzieri, A., Magen, H., Grogan, M., Mauermann, M., McPhail, E.D., Kurtin, P.J., Leung, N., Buadi, F.K., Dingli, D., et al. (2021). Systemic amyloidosis from A (AA) to T (ATTR): a review. Journal of Internal Medicine *289*, 268–292. 10.1111/joim.13169.

8. Shi, Y., Zhang, W., Yang, Y., Murzin, A.G., Falcon, B., Kotecha, A., van Beers, M., Tarutani, A., Kametani, F., Garringer, H.J., et al. (2021). Structure-based classification of tauopathies. Nature *598*, 359–363. 10.1038/s41586-021-03911-7.

9. Benson, M.D., James, S., Scott, K., Liepnieks, J.J., and Kluve-Beckerman, B. (2008). Leukocyte chemotactic factor 2: A novel renal amyloid protein. Kidney International *74*, 218–222. 10.1038/ki.2008.152.

10. Larsen, C.P., Ismail, W., Kurtin, P.J., Vrana, J.A., Dasari, S., and Nasr, S.H. (2016). Leukocyte chemotactic factor 2 amyloidosis (ALECT2) is a common form of renal amyloidosis among Egyptians. Mod Pathol *29*, 416–420. 10.1038/modpathol.2016.29.

11. Picken, M.M. (2020). The Pathology of Amyloidosis in Classification: A Review. AHA *143*, 322–334. 10.1159/000506696.

12. Mann, B.K., Bhandohal, J.S., Cobos, E., Chitturi, C., and Eppanapally, S. (2022). LECT-2 amyloidosis: what do we know? Journal of Investigative Medicine *70*, 348–353. 10.1136/jim-2021-002149.

13. Yamagoe, S., Yamakawa, Y., Matsuo, Y., Minowada, J., Mizuno, S., and Suzuki, K. (1996). Purification and primary amino acid sequence of a novel neutrophil chemotactic factor LECT2. Immunology Letters *52*, 9–13. 10.1016/0165-2478(96)02572-2.

14. Yamagoe, S., Kameoka, Y., Hashimoto, K., Mizuno, S., and Suzuki, K. (1998). Molecular Cloning, Structural Characterization, and Chromosomal Mapping of the Human LECT2 Gene. Genomics *48*, 324–329. 10.1006/geno.1997.5198.

15. Hiraki, Y., Inoue, H., Kondo, J., Kamizono, A., Yoshitake, Y., Shukunami, C., and Suzuki, F. (1996). A Novel Growth-Promoting Factor Derived from Fetal Bovine Cartilage, Chondromodulin II: PURIFICATION AND AMINO ACID SEQUENCE*. Journal of Biological Chemistry *271*, 22657–22662. 10.1074/jbc.271.37.22657.

16. Willis, S.A., Sargeant, J.A., Yates, T., Takamura, T., Takayama, H., Gupta, V., Brittain, E., Crawford, J., Parry, S.A., Thackray, A.E., et al. (2020). Acute Hyperenergetic, High-Fat Feeding Increases Circulating FGF21, LECT2, and Fetuin-A in Healthy Men. The Journal of Nutrition *150*, 1076–1085. 10.1093/jn/nxz333.

17. Zheng, H., Miyakawa, T., Sawano, Y., Asano, A., Okumura, A., Yamagoe, S., and Tanokura, M. (2016). Crystal Structure of Human Leukocyte Cell-derived Chemotaxin 2 (LECT2) Reveals a Mechanistic Basis of Functional Evolution in a Mammalian Protein with an M23 Metalloendopeptidase Fold*. Journal of Biological Chemistry *291*, 17133–17142. 10.1074/jbc.M116.720375.

18. Ha, J.-H., Tu, H.-C., Wilkens, S., and Loh, S.N. (2021). Loss of bound zinc facilitates amyloid fibril formation of leukocyte-cell-derived chemotaxin 2 (LECT2). Journal of Biological Chemistry *296*, 100446. 10.1016/j.jbc.2021.100446.

19. Larsen, C.P., Kossmann, R.J., Beggs, M.L., Solomon, A., and Walker, P.D. (2014). Clinical, morphologic, and genetic features of renal leukocyte chemotactic factor 2 amyloidosis. Kidney International *86*, 378–382. 10.1038/ki.2014.11.

20. Tsiolaki, P.L., Nasi, G.I., Baltoumas, F.A., Fishman, S., Tu, H.-C., and Iconomidou, V.A. (2019). Delving into the amyloidogenic core of human leukocyte chemotactic factor 2. Journal of Structural Biology *207*, 260–269. 10.1016/j.jsb.2019.06.001.

21. Falcon, B., Zivanov, J., Zhang, W., Murzin, A.G., Garringer, H.J., Vidal, R., Crowther, R.A., Newell, K.L., Ghetti, B., Goedert, M., et al. (2019). Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. Nature *568*, 420–423. 10.1038/s41586-019-1026-5.

22. Scheres, S.H.W. (2020). Amyloid structure determination in RELION-3.1. Acta Crystallogr D Struct Biol *76*, 94–101. 10.1107/S2059798319016577.

23. Eisenberg, D., and McLachlan, A.D. (1986). Solvation energy in protein folding and binding. Nature *319*, 199–203. 10.1038/319199a0.

24. Mompeán, M., Li, W., Li, J., Laage, S., Siemer, A.B., Bozkurt, G., Wu, H., and McDermott, A.E. (2018). The Structure of the Necrosome RIPK1-RIPK3 Core, a Human Hetero-Amyloid Signaling Complex. Cell *173*, 1244-1253.e10. 10.1016/j.cell.2018.03.032.

25. Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proceedings of the National Academy of Sciences *107*, 3487–3492. 10.1073/pnas.0915166107.

26. Perutz, M.F., Johnson, T., Suzuki, M., and Finch, J.T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proceedings of the National Academy of Sciences *91*, 5355–5358. 10.1073/pnas.91.12.5355.

27. Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.Ø., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-β spine of amyloid-like fibrils. Nature *435*, 773–778. 10.1038/nature03680.

28. Okumura, A., Suzuki, T., Miyatake, H., Okabe, T., Hashimoto, Y., Miyakawa, T., Zheng, H., Unoki-Kubota, H., Ohno, H., Dohmae, N., et al. (2013). Leukocyte cell-derived chemotaxin 2 is a zinc-binding protein. FEBS Letters *587*, 404–409. 10.1016/j.febslet.2013.01.025.

29. Liu, C., Sawaya, M.R., and Eisenberg, D. (2011). β2-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. Nat Struct Mol Biol *18*, 49–55. 10.1038/nsmb.1948.

30. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Cryst D *66*, 486–501. 10.1107/S0907444910007493.

31. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci *30*, 70–82. 10.1002/pro.3943.

32. Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., et al. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol *2*, 1–13. 10.1038/s42003-019-0437-z.

33. Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology *180*, 519–530. 10.1016/j.jsb.2012.09.006.

34. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Cryst D *75*, 861–877. 10.1107/S2059798319011471.

35. Chaudhury, S., Lyskov, S., and Gray, J.J. (2010). PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics *26*, 689–691. 10.1093/bioinformatics/btq007.

36. Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567. 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

37. Wagner, T., Lusnig, L., Pospich, S., Stabrin, M., Schönfeld, F., and Raunser, S. (2020). Two particle-picking procedures for filamentous proteins: SPHIRE-crYOLO filament mode and SPHIRE-STRIPER. Acta Cryst D *76*, 613–620. 10.1107/S2059798320007342.

38. He, S., and Scheres, S.H.W. (2017). Helical reconstruction in RELION. Journal of Structural Biology *198*, 163–176. 10.1016/j.jsb.2017.02.003.

39. Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ *6*, 5–17. 10.1107/S205225251801463X.

40. Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253–267. 10.1107/S2052252520000081.

41. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput *13*, 3031–3048. 10.1021/acs.jctc.7b00125.

Figures:

Figure 1: Views of the recombinant LECT2 fibril core revealed by cryo-EM helical reconstruction. A) Model fibril core structure of LECT2 (teal) modelled within the refined 2.7Å resolution cryo-EM map (grey mesh). B) View of the LECT2 fibril core with helical symmetry applied to show its full 850Å twist and 60Å width. C) Side view of the fibril model in the final cryo-EM map demonstrates its fit into density within the fibril layers.

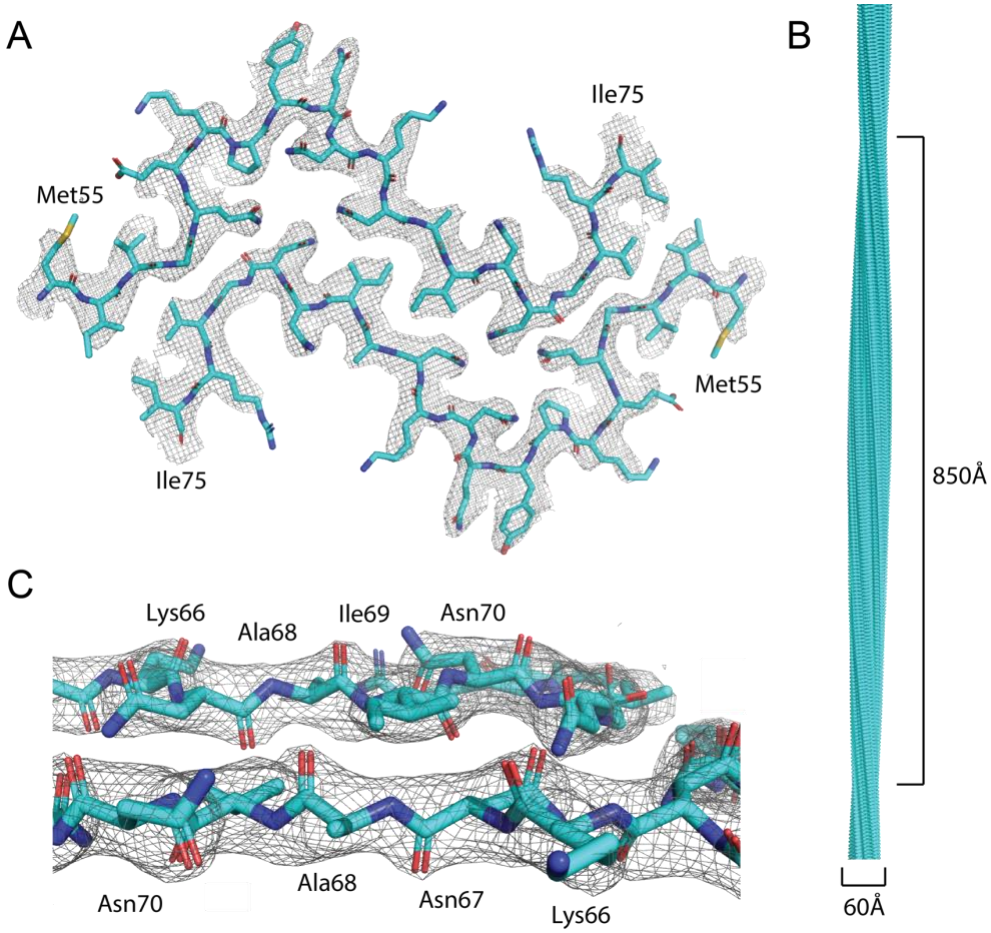


Figure 2: LECT2 fibril core sequence and structure overview. A) Sequence of LECT2 showing native secondary structure elements as yellow arrows for beta-strands and green rectangles for alpha helices with disulfides shown as orange brackets and the I40V polymorphism noted by a red star. The sequence of the fibril core is highlighted in red. B) Depiction of the fibril core showing the distribution of polar and nonpolar residues. C) Depiction of the fibril core showing stabilizing residues and the calculated solvation energy of the fibril core structure. Depictions of the fibril core were generated using the Amyloid Illustrator web service (https://srv.mbi.ucla.edu/AmyloidAtlas/Illustrator/).

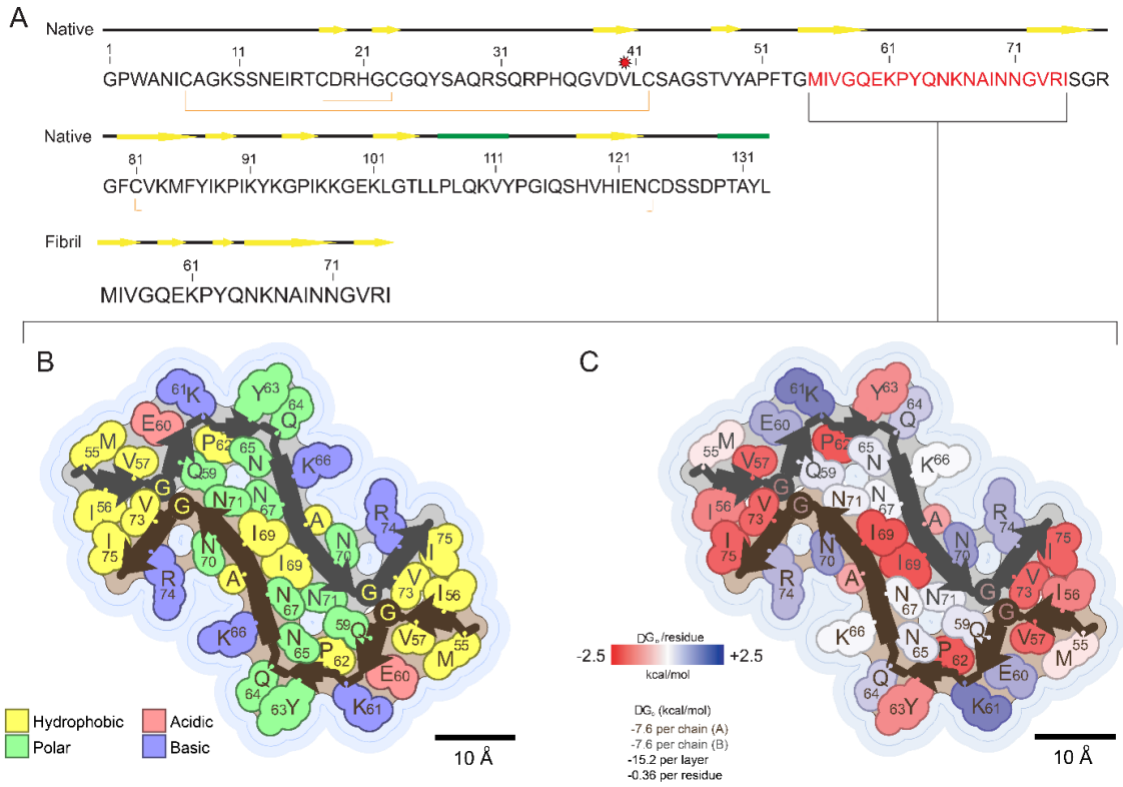
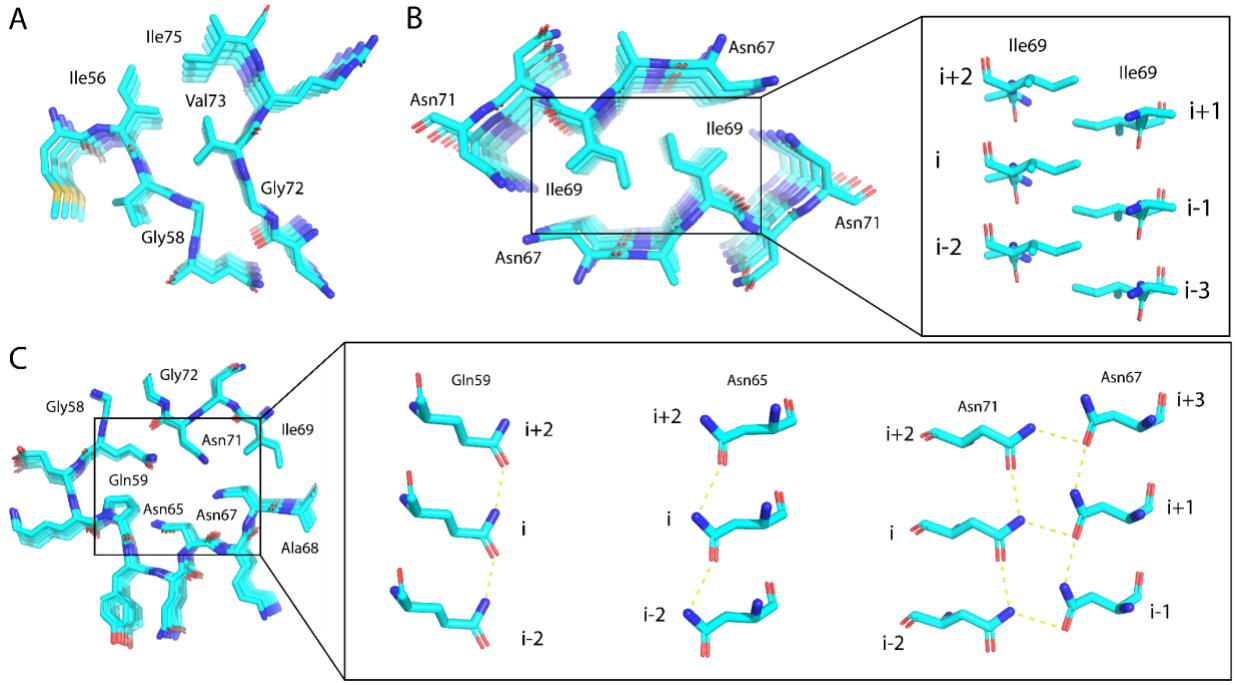


Figure 3: Stabilizing contacts at the core of the LECT2 fibril. A) The outer steric zipper interface formed by residues Gly58, Gly72, Val73, Ile56, and Ile75. B) The central steric zipper interface formed by residue Ile69 is flanked by a paired polar ladder. Inset shows the pseudo-2_1 stacking of fibril layers. C) Hydrophilic pocket within the LECT2 fibril core contains polar ladders formed by uncharged polar residues Gln59, Asn65, and Asn67 with Asn71, spaced 4.69Å apart along the fibril axis (inset).



STAR Methods

Key Resources Table

Reagent or Resource	**Source**	**Identifier**
Bacterial and Viral strains		
BL21 Star™ (DE3) Chemically Competent *E. coli*	ThermoFisher Scientific	CAS: C601003
Recombinant DNA		
Codon-optimized synthetic gene encoding residues 19-151 of the human LECT2 protein (GenBank AAC17734.1)	Genewiz	N/A
Chemicals and recombinant proteins		
LECT2 protein	This work.	N/A
Proteinase K protein	Sigma-Aldrich	CAS: 39450-01-6
Thioflavin T	ThermoFisher Scientific	CAS: 2390-54-7
Deposited Data		
Recombinant LECT2 Fibril	This work.	PDB ID: 8G2V EMBD: EMD-29682
Software and Algorithms		
Pymol Molecular Graphics	System Schrodinger, LLC	https://pymol.org/2/
Coot	Emsley et al. [30]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
UCSF Chimera	UCSF RBVI [31]	https://www.cgl.ucsf.edu/chimera/
CrYOLO	Wagner et al. [32]	https://cryolo.readthedocs.io/en/stable/index.html
RELION	Sjors H.W. Scheres [33]	https://relion.readthedocs.io/en/release-4.0/
Phenix	Liebschner et al. [34]	https://phenix-online.org/
PyRosetta	Chaudhury et al. [35]	https://www.pyrosetta.org/
MASCOT	Perkins et al. [36]	https://proteomicsresource.washington.edu/mascot/
Other		
Quantifoil (R1.2/1.3) 300M Au Cryo-EM Grids	Ted Pella inc.	CAS: 658-300-AU-100

Method Details

Protein Expression and Purification

Full-length, mature human sequence LECT2 protein encoding the I40V sequence polymorphism and an N-terminal 6-His tag was transformed into and expressed in BL21-gold E. coli cells. Two 1L cultures grown at 37°C to an O.D. of 0.75 were induced with 1mM IPTG and allowed to shake overnight at 23°C. Cells were then collected by centrifugation at 10,000xg for 10 minutes, resuspended in 1X PBS with 6M guanidine, and lysed on an Avestin Emulsiflex C3 (ATA Scientific). Insoluble cell material was removed by centrifugation at 9000 rpm for 50 minutes and the supernatant was collected, filtered using a 0.8µm size cutoff, and run over a His-trap column and washed with 3 column volumes of buffer. Protein was eluted with an elution buffer of 6M guanidine, 1X PBS, and 500 mM imidazole. Eluate fractions were evaluated for protein purity by gel electrophoresis then pooled and concentrated to 40 mg/ml (Figure S1B). Concentrated stocks were flash frozen for storage at -80°C in 1X PBS with 6M guanidine and 20mM TCEP.

Thioflavin T Fluorescence

To generate a Thioflavin T (ThT) fibril growth kinetics curve, LECT2 protein was prepared to 2mg/ml in fibrillization buffer as described above. Equimolar amounts of Thioflavin T, also dissolved in fibrillization buffer, was added to this solution. This preparation along with a control of only fibrillization buffer and ThT were set up in triplicates of 100 µL in a 96 well plate. The samples were then subjected to continuous shaking on a Varioskan LUX (ThermoFisher Scientific) as the ThT fluorescence was measured every 15 minutes for 48 hours using an excitation wavelength of 445 nm and measuring emission at 482 nm.

Cryo-EM Sample Preparation and Data Collection

To grow fibrils for cryo-EM analysis, stock LECT2 protein was diluted to 2 mg/ml (120 µM) into a fibrillization buffer containing 50 mM MOPS 6.5, 30 mM TCEP, 10 mM EDTA, and 150 mM NaCl. A volume of 100 µL of protein solution per well was then agitated in a 96 well plate on an acoustic shaker running at 60 Hz (2400 rpm) at 37°C for 48 hours. The solution was collected and spun down at 10,000xg for 5 minutes to remove large aggregates not amenable to single particle analysis. The solution was then diluted 6-fold in buffer and glycerol was added to the remaining supernatant to a concentration of 0.25%. A volume of 1.5 µL of this solution was added to each side of a gold Quantifoil R 1.2/1.3 cryo-EM grid (Ted Pella Inc.) within a Vitrobot System (ThermoFisher Scientific) with humidity set to 100% and a temperature of 4°C. The grid was then blotted for 1.5 seconds with the blot force set to -1, plunge frozen in liquid ethane, and stored for data collection in liquid nitrogen. High-resolution cryo-EM data was collected at the Stanford-SLAC Cryo-EM Center (S^2C^2) over the course of 48 hours using EPU for automated data collection. A total of 13830 movies were obtained taking three shots per hole on the grid. Movies were collected on a Titan Krios G3i microscope (ThermoFisher Scientific) equipped with a BioQuantum K3 camera (Gatan, Inc.) using a pixel size of 0.79 Å/pixel, a total dose of 52 $e^-/Å^2$ over 40 frames, and a defocus range between −0.8 and −1.8 µm.

Cryo-EM Data Processing

Collected images were input into MotionCor2 for drift correction and CTFFIND4 was used to calculate Contrast Transfer Functions (CTF). Fibrils were automatically picked using filament mode in crYOLO[32,37] after being trained on a manually selected pool of fibril segments picked from 100 movies. This yielded 1,622,942 segments using an overlap of 90% between neighboring segments and a box size of 384 pixels. The segments were then transferred to RELION 3.1[33,38,39] for all subsequence 2D and 3D classification. The segments were split into six equally sized groups and subjected to iterative rounds of reference-free two-dimensional (2D) classification using T = 8 and K = 100 to remove poorly aligned classes. After narrowing each split of segments, they were combined again (T = 8 and K = 100) resulting in 308,578 segments contributing to well-

defined 2D classes to be used in 3D classification. These classes were used to generate a *de novo* 3D initial model which was used as a reference for further 3D classification. Initial classification (K = 4, T = 4) did not impose pseudo-2_1-screw symmetry and used an initial helical twist of -1.06° and a helical rise of 4.8 Å. The best class from this classification was selected out and further subclassified by removing any segments with a CTF estimate poorer than 4 Å resolution. The resulting 51,543 segments were used for further 3D classification (K = 3, T = 4) allowing local optimization of helical twist and rise. The best resulting class showed clear strand separation along the fibril axis and the pseudo-2_1-screw symmetry became apparent without it having been imposed. The model and data from these 24,770 segments were used for high-resolution gold-standard 3D auto-refinement with the pseudo-2_1-screw symmetry now enforced. The resulting model underwent iterative Bayesian polishing[39] and CTF refinement[40] to further improve the map (Figure S6A). A final refined helical twist of 179.49° was derived from the auto-refine map and a helical rise of 2.345 Å was imposed on the post-processed map based on X-ray fibril diffraction data showing 4.69 Å strand separation in the fibril samples. The final resolution was ultimately calculated to be 2.715 Å from gold-standard Fourier shell correlations at 0.143 between two independently refined half-maps (Figure S6B).

Model Building

A poly-alanine model was initially built into the post-processed map in Coot[30]. Side chains were modified manually to test the fit of different registrations of the LECT2 sequence in the density. A close approach between tubes of density could be interpreted only as the G58-G72 interaction between symmetric strands, as the closeness of the approach would sterically exclude all other side chains. This feature helped to identify the likely segment of ALECT2 and the entire 21 residue chain from Met55-Ile75 was built out from there. Final atomic refinements and statistical calculations for Supplementary Table 1 were performed in Phenix[34]. While the map seemed to agree clearly with this assignment, we decided to confirm its validity by testing the fit and favorability of all possible 21-residue segments from the LECT2 sequence within the map using Rosetta. Threading was performed with a custom python script utilizing the PyRosetta software package[35]. In a sliding window approach, each sequence window was threaded onto a poly-alanine backbone, placed in the density and Fast Relaxed in PyRosetta. The pipeline used the REF2015 score function[41] with an increased electrostatics weight (fa_elec = 1.5) in combination with a density score term (elec_dens_fast = 25). Each sequence was energy minimized in triplicate and the resulting scores were averaged. Symmetry was applied to pose objects to increase speed of computation. The Fast Relaxed structure was then optimized and refined in Coot[36] and Phenix[34].

X-ray Fibril Diffraction

ALECT2 amyloid fibrils were grown as described above for cryo-EM sample preparation. A 3 uL droplet of the mature fibril solution was pipetted between two glass rods held about a millimeter apart and allowed to evaporate. This process was repeated multiple times until a visible fibril bundle was observed bridging the glass rods. This bundle was placed on an in-house Rigaku FRE+ rotating anode X-ray beamline and exposed for 5 minutes onto a Rigaku HTC detector with varimax confocal optics. Proteinase K and ice ring diffraction patterns were used to calibrate the detector distance and accurately measure the 4.69 Å reflection observed from the fibril bundle.

Fibril Stability Assays

ALECT2 amyloid fibrils were grown as described above for cryo-EM sample preparation and then pelleted at 14,300xg for 10 minutes. The fibril pellets were then resuspending to a protein concentration of 40 μM in water, fibrillization buffer, 3M urea with fibrillization buffer, and fibrillization buffer with 1:1000 Proteinase K (Sigma-Aldrich) to a final concentration of 40 nM. These solutions were immediately pipetted in triplicate into a 96 well tray alongside buffer control wells for each and placed on a NEPHELOstar Plus plate reader (BMG Labtech). The tray was

stirred every ten minutes to homogenize the mixture and light scattering of particles in solution was measured immediately after each agitation. For analysis of protease resistant LECT2 fragments, fibrils at 120 μM were incubated with 10 μM Proteinase K for various timepoints and then evaluated by SDS-PAGE (Figure S9A). The protein gel band from the one-hour timepoint was extracted, digested with Trypsin, and analyzed by gel liquid chromatography tandem mass spectrometry.

In-gel digestion and peptide mass fingerprinting of ALECT2 using GeLC-MS/MS

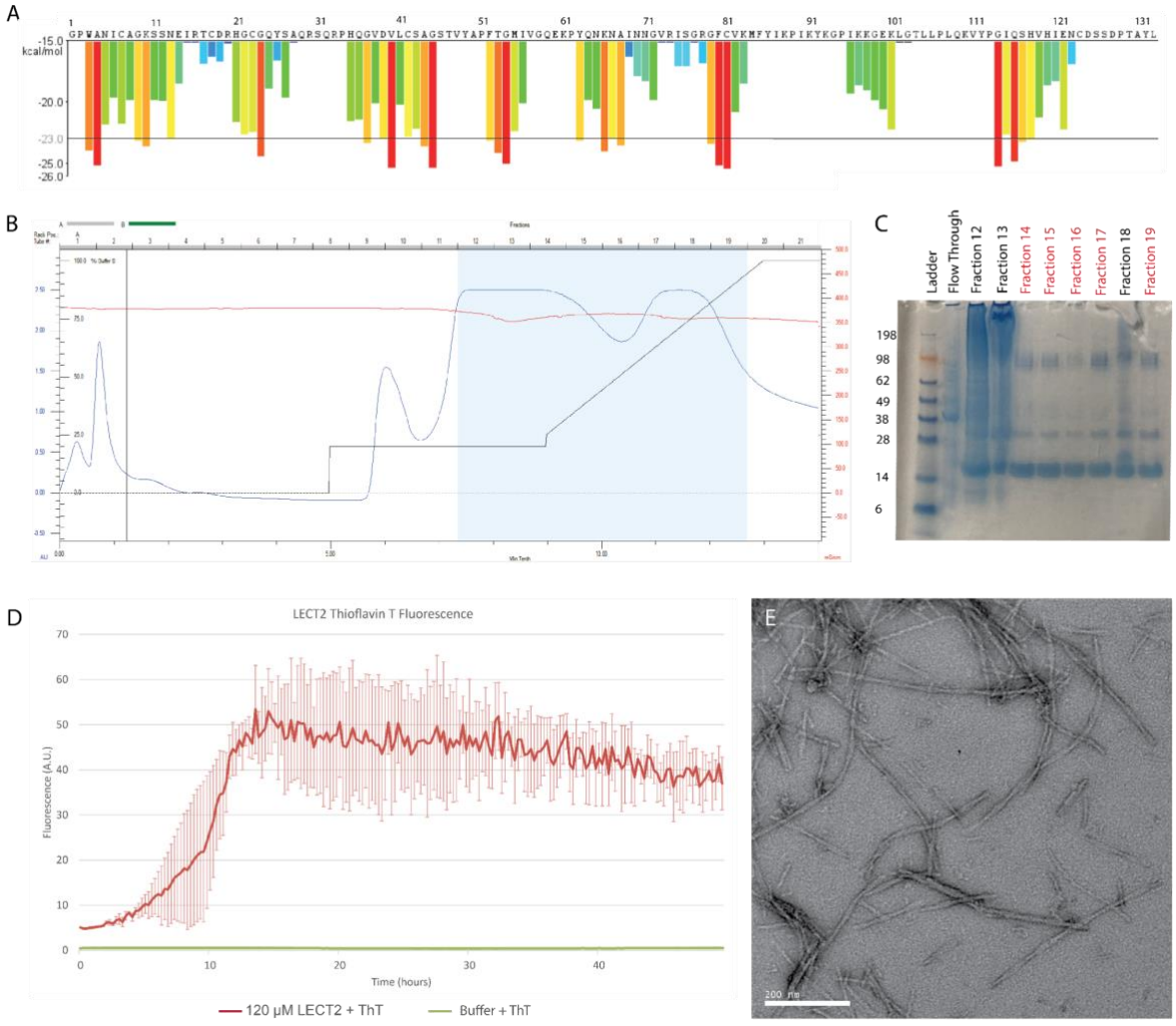
Gel Liquid Chromatography tandem mass spectrometry spectra collected on Proteinase K digested ALECT2 fibrils were acquired on a ThermoFisher Q-Exactive Plus (UCLA Molecular Instrumentation Center, Los Angeles, CA, USA). LECT2 fibrils at 120 μM were incubated with 10 μM Proteinase K for one hour. These samples were removed at various timepoints, boiled for 10 minutes at 98 °C to halt digestion, and loaded on a 4–12% Bis-Tris SDS-PAGE gel. The gel band from the one-hour timepoint was excised and digested with 200 ng trypsin at 37°C overnight. The digested products were then extracted from the gel bands in 50% acetonitrile/49.9% H2O/ 0.1% trifluoroacetic acid (TFA) followed by desalting with C18 StageTips. Extracted peptides were then injected on an EASY-Spray HPLC column (25 cm × 75 μm ID, PepMap RSLC C18, 2 μm, ThermoScientific) and tandem mass spectra were acquired with a quadrupole orbitrap mass spectrometer (Q-Exactive Plus Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source. The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer (VER. 1.4, Thermo Scientific and analyzed by a MASCOT[36] sequence database search.

Supplementary Information:

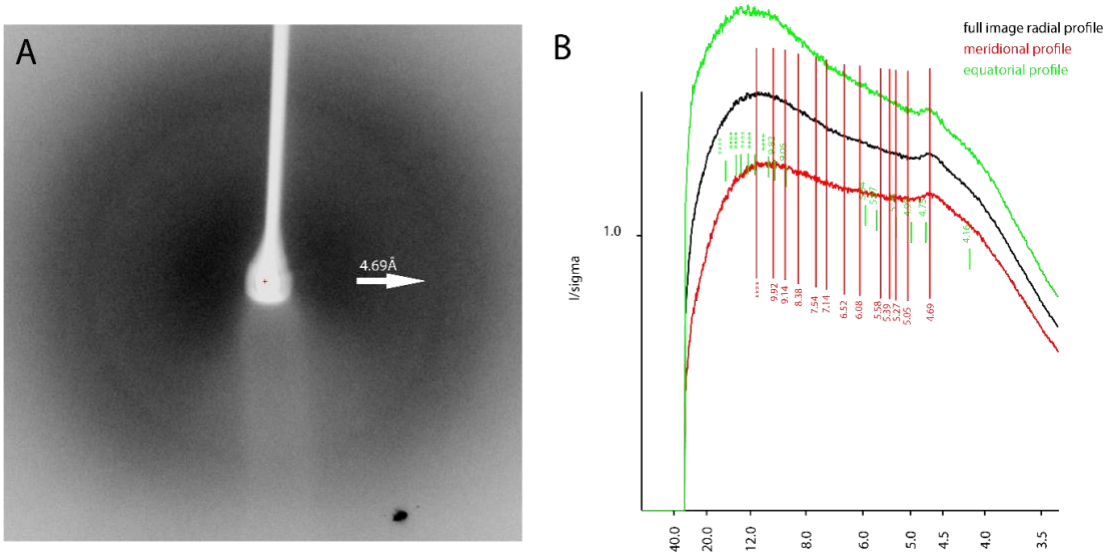
Supplementary Table 1: Cryo-EM data collection and structure information.

Data Collection	
Magnification	110,000x
Defocus Range (um)	-0.8 to -1.8
Voltage (kV)	300
Microscope	Titan Krios G3i
Camera	BioQuantum K3 camera
Frame exposure time (s)	2.5
No. of movie frames	40
Total electron dose ($e^-/Å^2$)	52
Pixel Size (Å)	0.799
Reconstruction	
Box size (pixel)	384
Inter-box distance (Å)	30
No. of segments extracted	1,622,942
No. of segments after Class2D	308,578
No. of segments after Class3D	24,770
Map Resolution (Å)	2.715
FSC threshold	0.143
Map sharpening B-factor ($Å^2$)	-69.7
Helical rise (Å)	2.345
Helical twist (°)	179.49
Symmetry imposed	C1
Atomic Model	
Initial model used	*De novo*
No. of unique non-hydrogen atoms	166
R.m.s.d. bonds (Å)	0.006
R.m.s.d. angles (°)	1.055
Molprobity score	1.88
Favored rotamers (%)	94.74
Ramachandran favoured (%)	5.26
Ramachandran outliers (%)	0
CB deviations > 0.25 Å (%)	0
Bad bonds (%)	0
Bad angles (%)	0

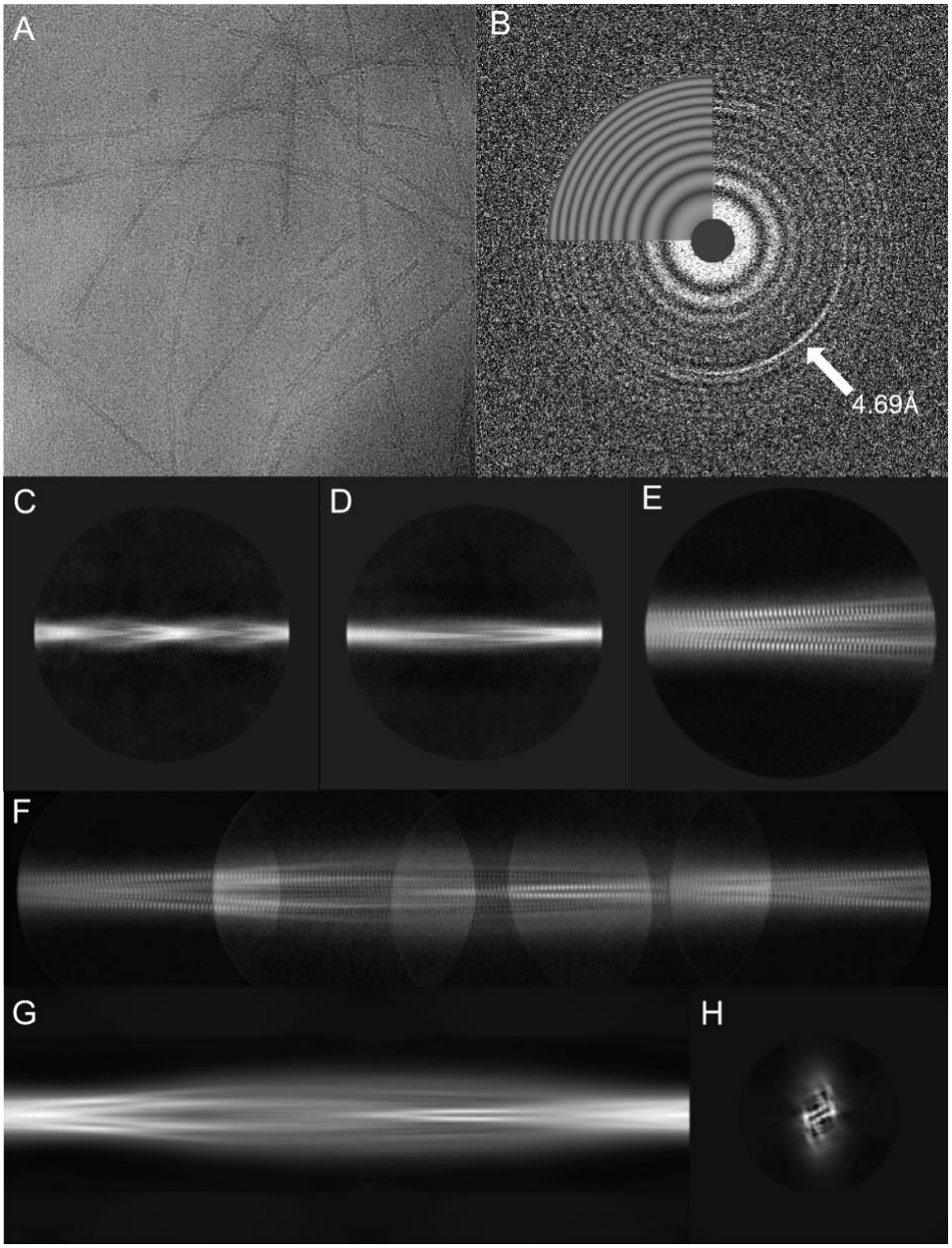
Supplementary Figure 1: Production of recombinant LECT2 amyloid fibrils. A) Predicted amyloid forming six-residue segments (kcal/mol < -23.0) are shown for the LECT2 sequence, calculated using ZipperDB. B) A His-tagged, full-length LECT2 polypeptide was expressed and purified using a His-Trap column under the illustrated gradient. C) Purity of the fractions was evaluated by gel electrophoresis. The purest fractions: 14-17, 19 (red) were pooled and concentrated. D) The recombinant protein was used to form fibrils and Thioflavin T fluorescence was monitored over 48 hours of growth. E) The resulting fibrils from Thioflavin T analysis were imaged by negative stain TEM.



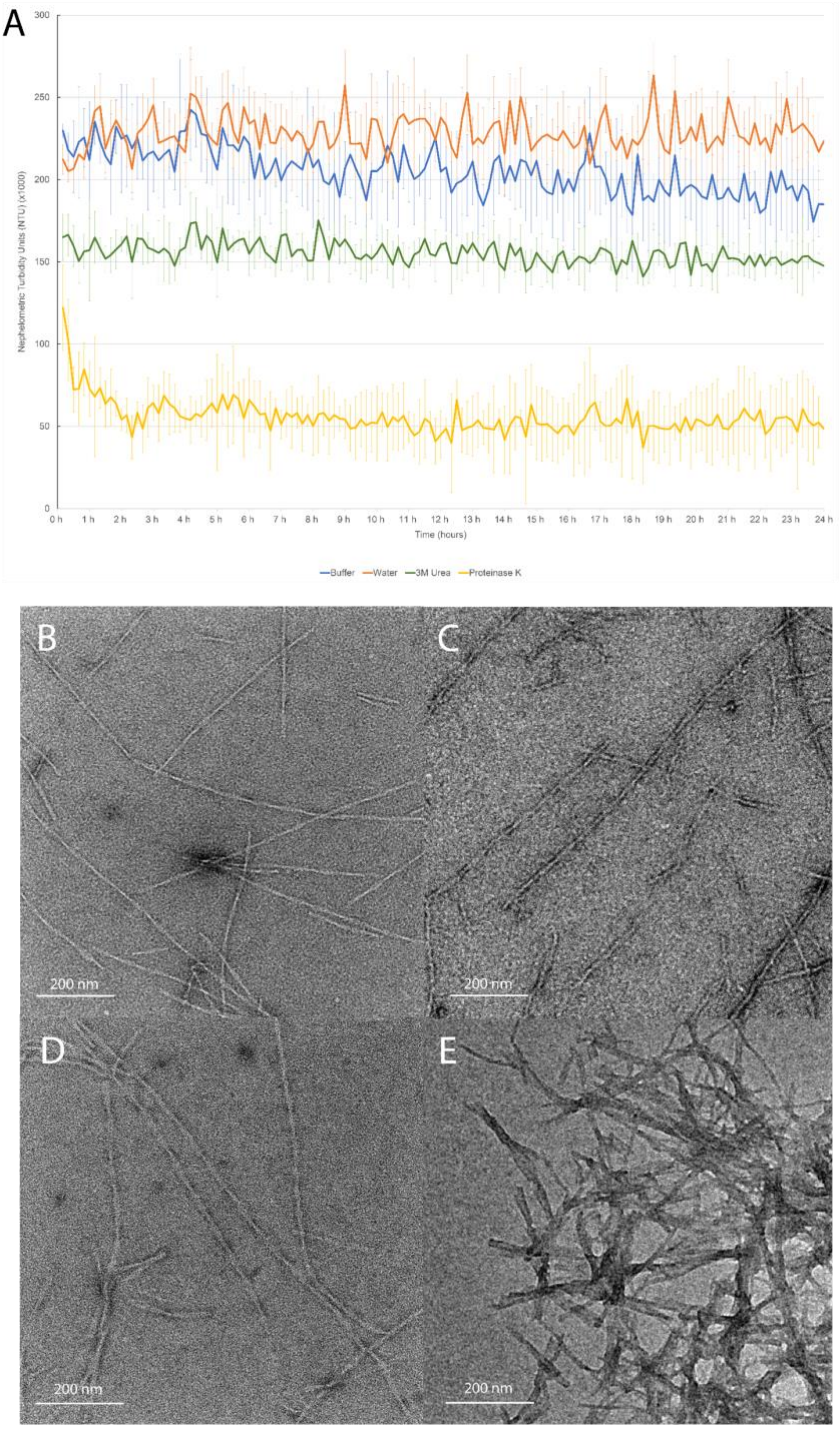
Supplementary Figure 2: X-ray Diffraction from LECT2 fibrils A) X-ray diffraction exposure of bundles recombinant LECT2 fibrils. B) Radial profile generated from the diffraction image indicates a strong diffraction ring at 4.69Å.



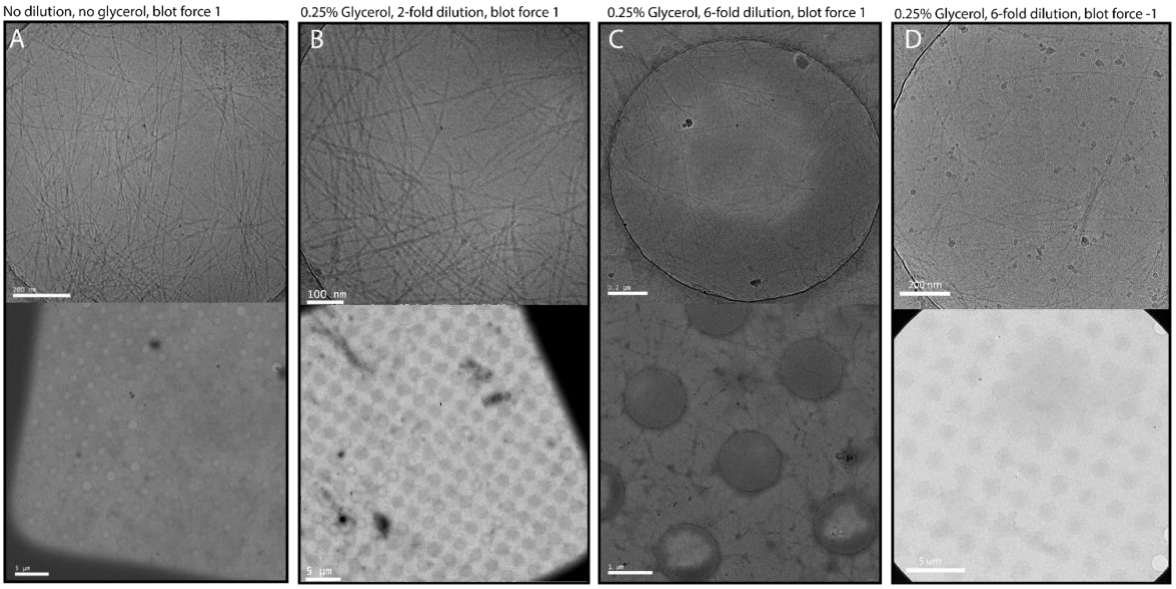
Supplementary Figure 3: Vitrified LECT2 fibrils, CTF analysis, and 2D classification. A) Example image of the vitrified fibrils used for cryo-EM data collection. B) Example CTF image showing the signature fibril diffraction profile at 4.69Å. C) Representative 2D class of the fast-twisting fibril species from a 1024 pixel box. D) Representative 2D class of the slow-twisting fibril species from a 1024 pixel box. E) A slow-twisting fibril image from 2D class after classification using a 384 pixel box in Relion. F) Composite image of 384 pixel box size 2D classes from the slow-twisting fibril polymorph stitched together to show its full twist. G) *De novo* 3D model generated from 2D classes in Relion viewed from the side and H) as a cross section.



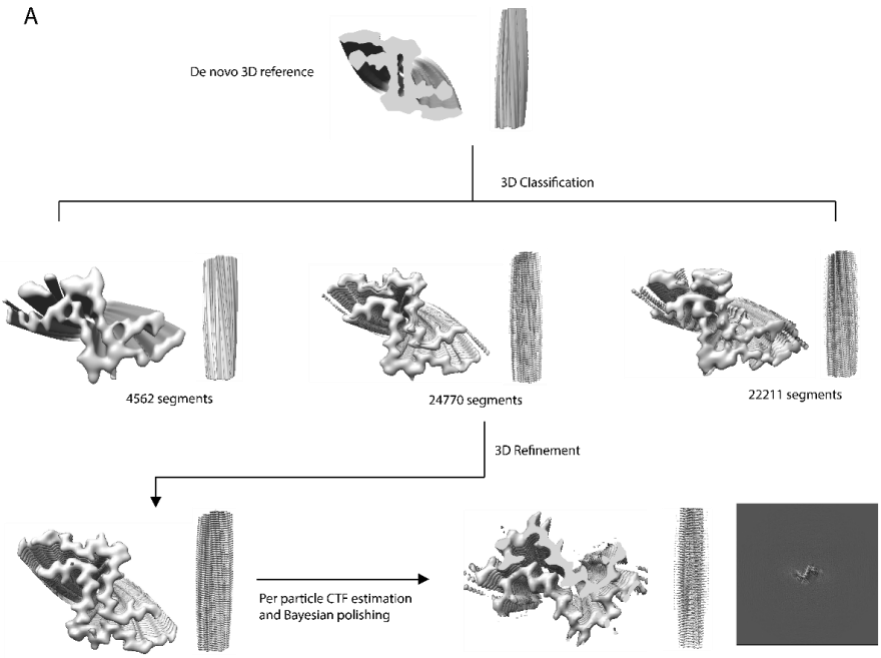
Supplementary Figure 4: LECT2 fibril stability. A) Nephelometer turbidity readings for the incubated fibril samples over 24 hours. B-E) Negative stain TEM images of the fibrils remaining after 24 hours incubation in fibrillization buffer (B), water (C), 3M urea (D), and 40nM Proteinase K (E).

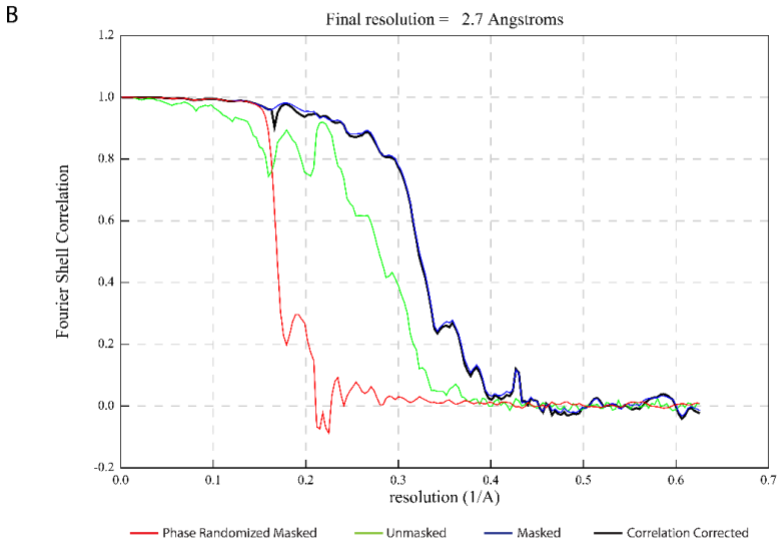


Supplementary Figure 5: Cryo-EM grid optimization. Recombinant fibrils were plunge frozen in buffer at different dilutions and Vitrobot blot settings for optimal particle dispersion, representative micrographs are shown. A) No dilution results in clumping and overcrowding. B) Addition of 0.25% glycerol improves ice distribution, but 2-fold dilution does not reduce crowding. C) 6-fold dilution improves fibril distribution in grid holes. D) Reducing the blot force improves ice distribution and maintains fibril distribution.

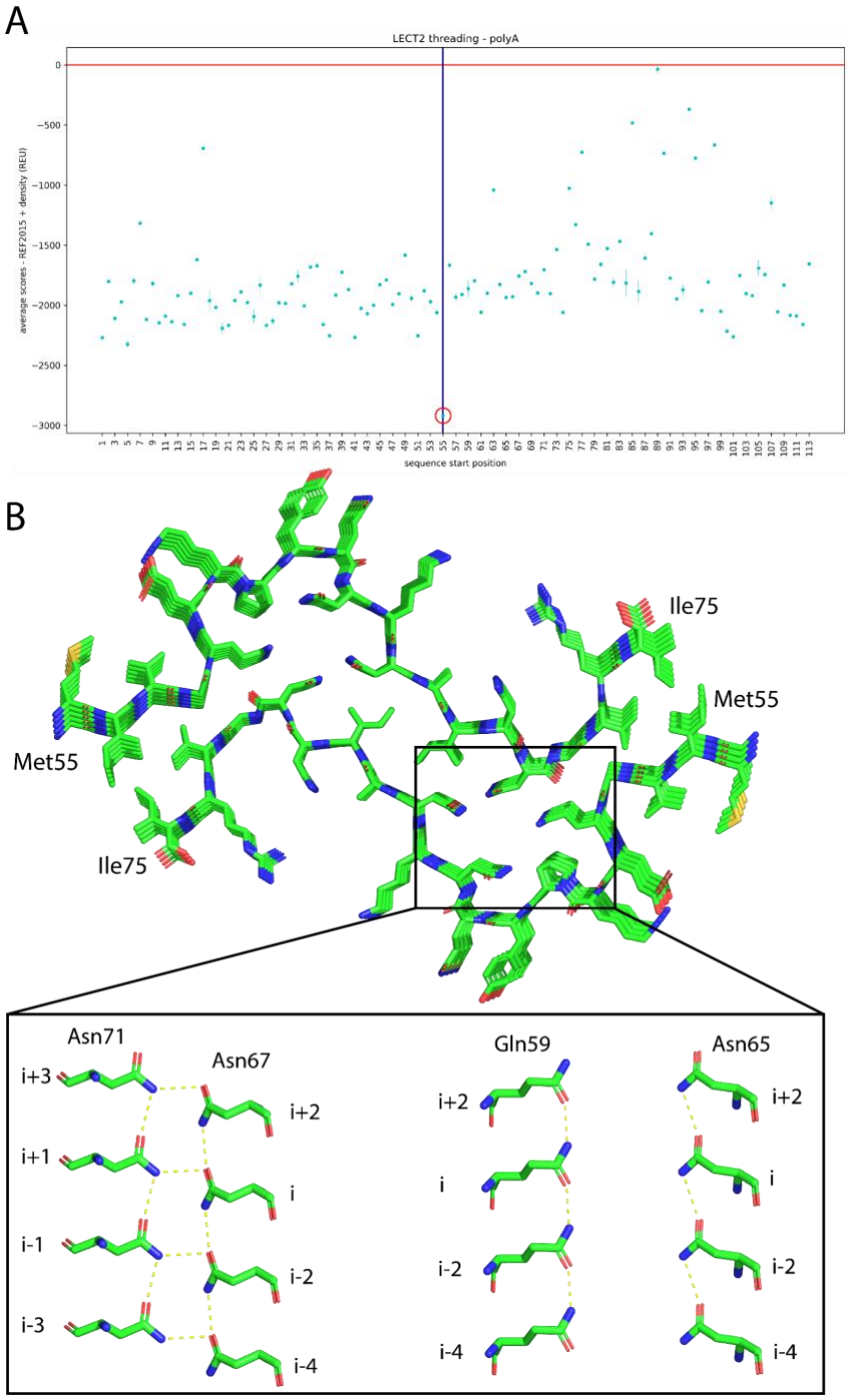


Supplementary Figure 6: Cryo-EM data processing and model comparisons. A) Overview schematic representation of helical reconstruction of fibril species performed in RELION showing maps of fibril sides and cross sections. B) FSC curves between independently refined half-maps which were phase randomized (red), unmasked and corrected (green), masked (navy blue), or masked and corrected (black). Final resolution of 2.7Å was calculated using a cutoff of FSC=0.143.

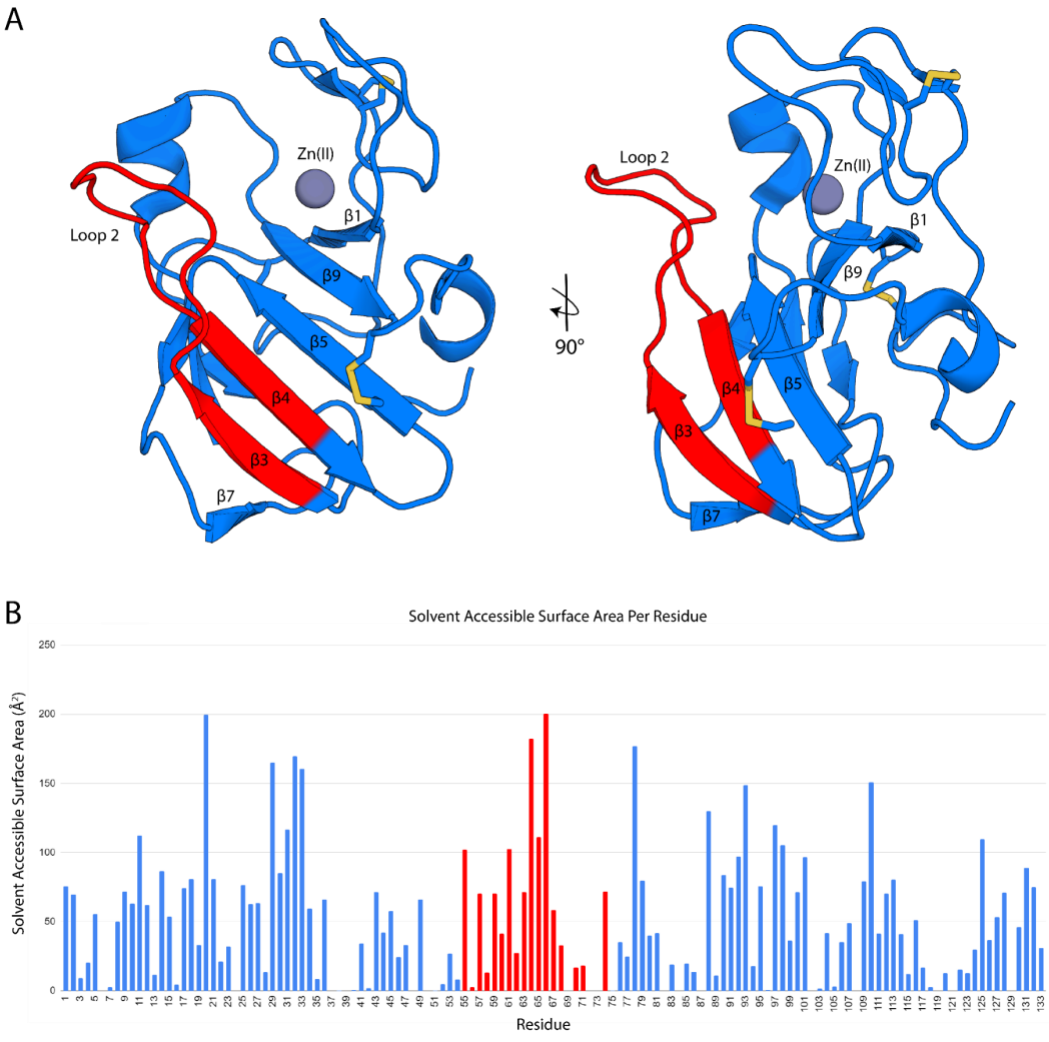




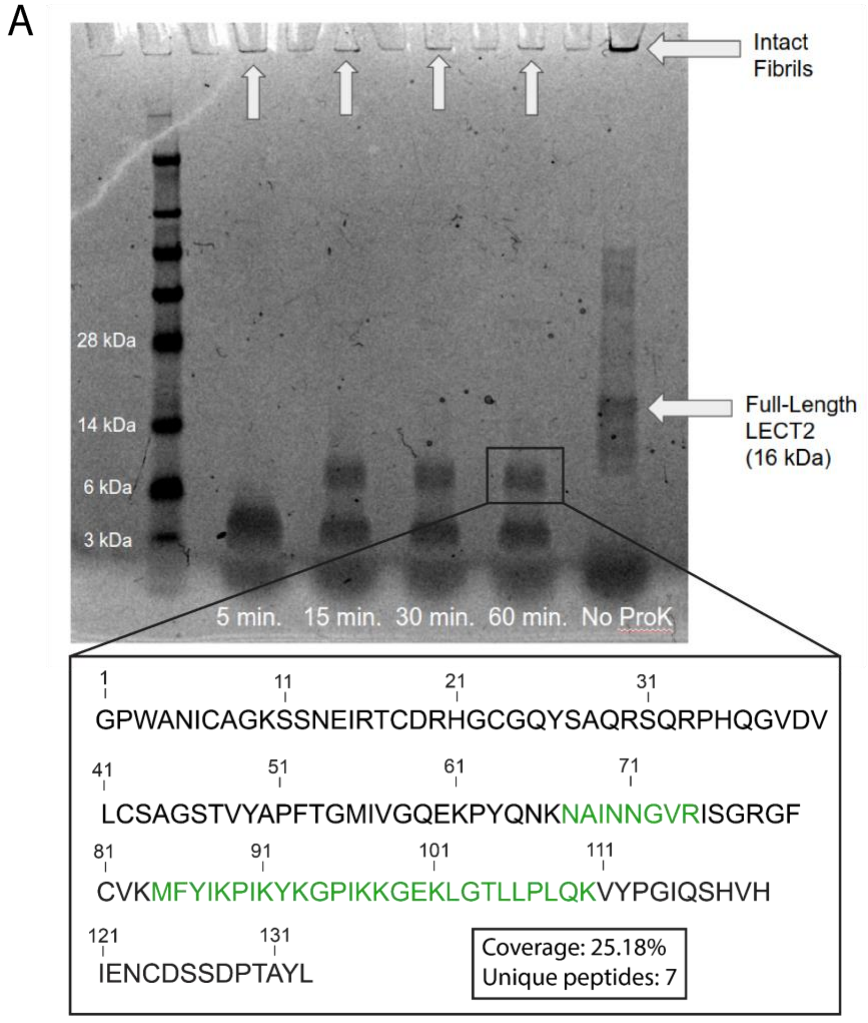
Supplementary Figure 7: Identification of sequence matching LECT2 fibril structure by threading and energy minimization in Rosetta. A) Threading analysis indicates that the sequence beginning at Met55 is the most energetically favorable among possible assignments for the cryo-EM map (red circle). B) Energy-minimized, fast relaxed model from Rosetta predicts overall fibril core morphology as well as the stabilizing polar ladders in the hydrophilic pockets (inset).



Supplementary Figure 8: LECT2 globular fold and solvent accessibility of residues 55-75. A) Views of LECT2 in its globular form (PDB:5B0H)[17]. Residues 55-75 (colored in red) are identified as forming the amyloid core of LECT2 in its fibril form. B) Solvent accessible surface area per residue of LECT2 in its globular fold calculated using the Ochanomizu University program (ver. 1.2).



Supplementary Figure 9: Mass spectrometry of Proteinase K digested fibrils. A) SDS-PAGE shows stable protein species present over one hour of Proteinase K digestion. Boxed band was excised and peptides were extracted for GeLC-MS/MS (inset) with sequence segments detected colored in green.

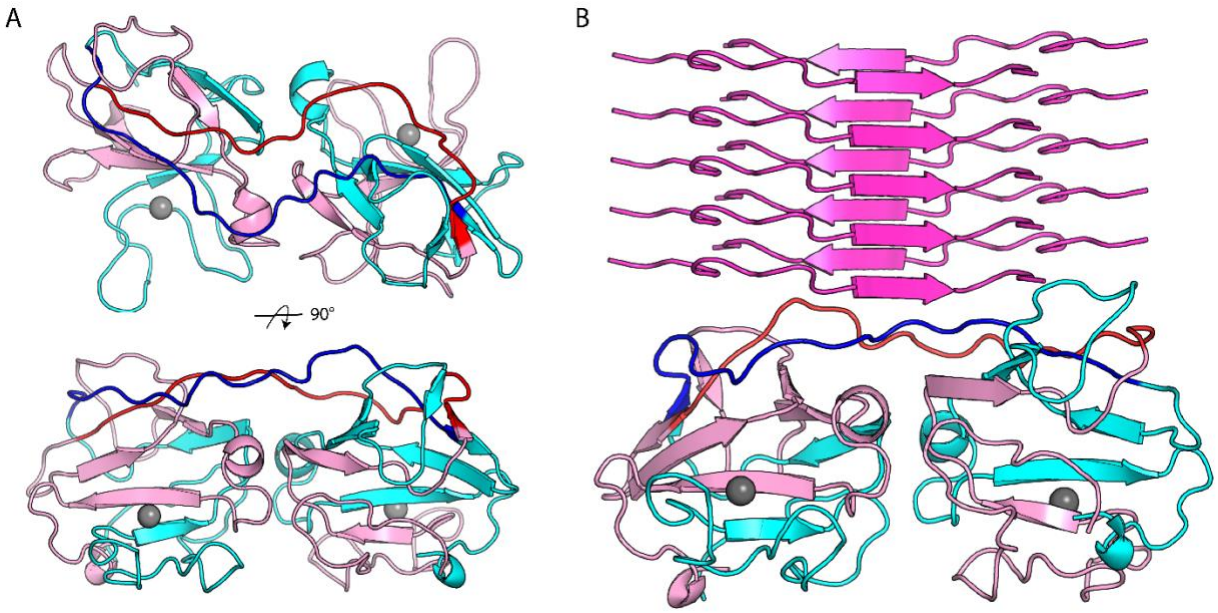


Supplementary Table 2: Summary of Proteinase K digested LECT2 peptides identified by GeLC-MS/MS.

Annotated Sequence	Modifications	MH+ [Da]	Theo. MH+ [Da]	Charge	m/z [Da]	XCorr
[K].NAINNGVR.[I]		857.4591	857.4588	2	429.2332	1.62
[K].mFYIKPIK.[Y]	M1(Oxidation)	1055.596	1055.596	3	352.5368	1.09
[K].mFYIKPIK.[Y]	M1(Oxidation)	1055.596	1055.596	2	528.3015	1.85
[K].mFYIKPIKYK.[G]	M1(Oxidation)	1346.753	1346.754	3	449.5891	1.43
[K].mFYIKPIK.[Y]	M1(Oxidation)	1055.599	1055.596	3	352.538	1.26
[K].mFYIKPIK.[Y]	M1(Oxidation)	1055.596	1055.596	2	528.3016	0.61
[K].KGEKLGTLLPLQK.[V]		1424.884	1424.884	3	475.6328	0.45
[K].mFYIKPIK.[Y]	M1(Oxidation)	1055.596	1055.596	3	352.537	0.79
[K].mFYIKPIKYKGPIK.[K]	M1(Oxidation)	1742.006	1742.007	4	436.257	0.95
[K].GEKLGTLLPLQK.[V]		1296.789	1296.789	3	432.9346	1.48
[K].LGTLLPLQK.[V]		982.6294	982.6295	2	491.8184	0.5
[K].LGTLLPLQK.[V]		982.6304	982.6295	2	491.8188	1.56
[K].LGTLLPLQK.[V]		982.6304	982.6295	2	491.8189	1.3

*Table includes only peptides mapped to the human LECT2 protein sequence used in these experiments.

Supplementary Figure 10: Modelled domain-swapping for globular LECT2 dimer. A) Modelled conformation of the LECT2 dimer (cyan, pink) domain-swapped through loop 2 using the ALECT2 fibril core structure (blue, red) B) ALECT2 fibril structure pictured above the domain-swapped model to highlight conformational similarities.



Supplementary Code: PyRosetta Sequence Threading

```
import os
import numpy as np
import pandas as pd
from pathlib import Path
import csv

from pyrosetta import *
from pyrosetta.rosetta import *
from pyrosetta.teaching import *
from pyrosetta.toolbox import *

from rosetta.protocols.cryst import *
from rosetta.protocols.rosetta_scripts import *

np.set_printoptions(threshold=sys.maxsize)

init('-ignore_unrecognized_res -load_PDB_components false -ignore_zero_occupancy false')

#########################################################
### INPUTS ###
#########################################################
### protein name
protName = 'lect2'

### to build symmetry pose
symmInfo = 'lect2.symm'
pose_file = 'lect2_INPUT.pdb'

### start/end residues in pose numbering for base layer of symmetric pose
base_start = 85
base_end = 105

### for job distribution
job = int(sys.argv[1])

### sliding window size
window_size = 21

### full sequence txt file to be threaded
seq_file = 'seq.txt'

### output name for scores file
scores_output = 'lect2_scores.csv'

#########################################################
### PREPROCESSING ###
#########################################################
### pre prosess sequence file
seq = Path(seq_file).read_text()
seq = seq.replace('\n', '')
seq_list = list(seq)
```

```
#########################################################
### FUNCTIONS ###
#########################################################
scorefxn = get_fa_scorefxn()

def symmetrize_pose(pose):
    """ set up symmetric pose """
    pose_symm_data = core.conformation.symmetry.SymmData(pose.total_residue(), pose.num_jump())
    pose_symm_data.read_symmetry_data_from_file(symmInfo)
    core.pose.symmetry.make_symmetric_pose(pose, pose_symm_data)
    sym_info = pose.conformation().Symmetry_Info()

    print("AssymUnit? equivalent_res")
    for i in range(1, pose.size()+1):
        print(i, sym_info.bb_is_independent(i), sym_info.bb_follows(i))

    print("Total Subunits:", sym_info.subunits())

    return pose

def def_windows(x, wnd_sz):
    """ define sequence positions for each window """
    window_list = []

    start_pos = 0
    end_pos = start_pos + wnd_sz

    num_windows = len(x) - wnd_sz + 1

    for j in range(0, num_windows):
        window_tmp = []
        for i in range(start_pos, end_pos):
            window_tmp.append(x[i])
        start_pos += 1
        end_pos += 1
        window_list.append(window_tmp)

    return window_list

def thread(pose, x, base_start, base_end):
    """ perform threading
        x = sequence window list """

    j = 0
    for i in range(base_start, base_end+1):
         mutate_residue(pose, i, x[job][j])
         j += 1

def FRwDensity(pose, rpt_num):
    """ set up and perform FastRelax with cryoEM density map  """

    ## set up density FastRelax
```

```
    setup_dens = XmlObjects.static_get_mover('<LoadDensityMap name = "loaddens"
mapfile="postprocess.mrc"/>')
    setup_dens.apply(pose)
    setup_dens_pose = rosetta.protocols.electron_density.SetupForDensityScoringMover()
    setup_dens_pose.apply(pose)

    ## set up score function with correct weights
    score = get_score_function()
    score_dens_cart = create_score_function('ref2015_cart')
    score_dens_cart.set_weight(rosetta.core.scoring.elec_dens_fast, 25)
    score_dens_cart.set_weight(fa_elec, 1.5)

    ## set up fast relax parameters
    mmf = pyrosetta.rosetta.core.select.movemap.MoveMapFactory()
    mmf.all_bb(setting=True)
    mmf.all_bondangles(setting=True)
    mmf.all_chi(setting=True)
    mmf.all_jumps(setting=True)
    mmf.set_cartesian(setting=True)

    FR = pyrosetta.rosetta.protocols.relax.FastRelax(scorefxn_in=score_dens_cart, standard_repeats=1)
    FR.cartesian(True)
    FR.set_movemap_factory(mmf)
    FR.min_type("lbfgs_armijo_nonmonotone")

    ## apply FR, score relaxed pdb, and output file
    FR.apply(pose)
    output_name = 'lect2_FR_thread_' + str(job) + '_' + str(rpt_num) + '.pdb'
    pose.dump_pdb('./output/' + output_name)
    scores = output_name, score_dens_cart(pose), scorefxn(pose)

    return scores

#########################################################
### RUN ###
#########################################################

window_list = def_windows(seq_list, window_size)  ## build list of sequence windows
seq_list_df = pd.DataFrame(window_list)          ## convert to df

for i in range(0, 3):
    pose = pose_from_pdb(pose_file)                    ## create pose object
    pose = symmetrize_pose(pose)                       ## symmetrize the pose

    thread(pose, window_list, base_start, base_end)         ## perform threading; mutate side chains on
pose object to reflect the current sequence window
    scores = FRwDensity(pose, i)                       ## perform FastRelax on current sequence window
```