

UCLA

UCLA Electronic Theses and Dissertations

Title

Structures at the Core of Mammalian Prions

Permalink

<https://escholarship.org/uc/item/0p65550b>

Author

Glynn, Calina

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structures at the Core of Mammalian Prions

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

by

Calina Glynn

2021

© Copyright by

Calina Glynn

2021

ABSTRACT OF THE DISSERTATION

Structures at the Core of Mammalian Prions

by

Calina Glynn

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2021

Professor Jose A. Rodriguez, Chair

Prion diseases, also known as Transmissible Spongiform Encephalopathies, are neurodegenerative diseases that pose a major threat to both humans and animals. They fall into a category of misfolding diseases known as amyloidoses where accumulation of fibrous protein aggregates correlates with disease symptoms. Unlike other amyloid diseases, prion diseases are infectious and not strictly linked with age. Victims of prion diseases experience dementia, hallucinations, and an inability to take care of themselves among other symptoms before inevitably succumbing to disease. While the time between presentation of symptoms and death is often less than one year, symptoms can take decades to appear. This makes the immediate cause of sporadic disease difficult to determine.

Prion diseases currently lack any form of treatment or means of prevention outside of selective breeding in sheep, which takes advantage of a disease-preventing sequence polymorphism. Infectious prions share β -sheet rich cores that produce a cross- β diffraction pattern not observed in natively folded PrP^C. Knowledge of the atomic structures adopted by prions will aid not only in

structure-based drug design and prion disease prevention, but also provide answers to the centuries old question of what makes a protein infectious.

The aim of this dissertation is to uncover the structural characteristics of prions that distinguish them from other amyloids. The dissertation is also aimed at uncovering molecular explanations for species barriers, whereby PrP^{Sc} from one species converts PrP^C of another species to a misfolded form with an efficiency dependent on both the original and newly infected host prions. These aims will be achieved through a combination of technological advancements made in atomic-level amyloid structure determination, with a focus on micro-crystal electron diffraction (MicroED) of the building blocks that make up prion cores (chapters 1 and 2) and single particle cryo-electron microscopy (cryo-EM) of prion filaments (chapters 3 and 4).

The dissertation of Calina Glynn is approved.

David S. Eisenberg

Tamir Gonen

Todd O. Yeates

Z. Hong Zhou

Jose A. Rodriguez, Committee Chair

University of California, Los Angeles

2021

DEDICATION

To my family for their unconditional love and support and all my mentors for their dedication and inspiration

LIST OF FIGURES

Figure 1.1	Structure of proto-PrP ^{Sc} from sub-ångström diffraction of nanoscale crystals by MicroED.....	24
Figure 1.2	Filament structure of proto-PrP ^{Sc} and its stabilizing network of hydrogen bonds.....	24
Figure 1.3	Sub-ångström cryo-EM maps show stabilizing hydrogen bonds in proto-PrP ^{Sc}	25
Figure S1.1	Sequence analysis of the common core of prion protofilaments.....	30
Figure S1.2	Characterization of proto-PrP ^{Sc} aggregates.....	31
Figure S1.3	Physical and chemical denaturation of proto-PrP ^{Sc} aggregates.....	32
Figure S1.4	Microfocal x-ray structure determination of proto-PrP ^{Sc} and comparison to MicroED structure of proto-PrP ^{Sc}	33
Figure S1.5	Ab initio structure determination of proto-PrP ^{Sc}	34
Figure S1.6	<i>B</i> factors in proto-PrP ^{Sc}	35
Figure S1.7	Hydrogen bond networks formed by polar side chains in β -sheets.....	36
Figure S1.8	Refinement of proto-PrP ^{Sc} as a function of resolution.....	37
Figure S1.9	Comparison of ultrahigh-resolution features in maps solved by MicroED maps and X-ray crystallography.....	38
Figure S1.10	Ab initio structure determination of carbamazepine.....	39
Figure S1.11	Comparison of ultrahigh-resolution features in MicroED maps of proto-PrP ^{Sc} refined in different software packages.....	40
Figure 2.1	The β 2 α 2 loop misfolds into amyloid.....	71
Figure 2.2	Structures of crystalline mammalian prion β 2 α 2 loop aggregates.....	72
Figure 2.3	Denaturation of β 2 α 2 loop crystals.....	73

Figure 2.4	Structural Conservation in the $\beta 2\alpha 2$ Loop.....	74
Figure 2.5	$\beta 2\alpha 2$ Loop Structures Fall Into Two Structural Categories.....	75
Figure 2.6	Comparison of naked mole rat and rabbit loop structures.....	76
Figure 2.7	Fiber Diffraction of $\beta 2\alpha 2$ loop aggregates.....	77
Figure 2.8	$\beta 2\alpha 2$ loop influence on fibril structure.....	78
Figure 3.1	PrP domain architecture and structure of the rPrP ^{Res} fibril, the protease-resistant core of rPrP ⁹⁴⁻¹⁷⁸	84
Figure 3.2	Secondary structure of rPrP ^{Res} in contrast to that of PrP ^C and topology of the rPrP ^{Res} protofilament.....	85
Figure 3.3	Stability of the core of rPrP ^{Res}	87
Figure 3.4	Compatibility of sequence variants at key residues in rPrP ^{Res}	88
Figure S3.1	Production, isolation and characterization of rPrP ⁹⁴⁻¹⁷⁸ fibrils.....	93
Figure S3.2	Partial protease digestion of rPrP ⁹⁴⁻¹⁷⁸ fibrils.....	94
Figure S3.3	Classification of major species of protease-resistant rPrP ⁹⁴⁻¹⁷⁸ fibrils.....	95
Figure S3.4	2D classification of minor morphological populations formed by rPrP ⁹⁴⁻¹⁷⁸	96
Figure S3.5	Agreement of rPrP ¹⁰⁶⁻¹⁴⁵ model with core density in rPrP ^{Res}	97
Figure S3.6	Structural agreement of hexapeptide prion zipper structures to rPrP ^{Res}	98
Figure S3.7	Molecular contacts at the core of rPrP ^{Res}	99
Figure S3.8	Alignment of fast-relaxed mutant sequences to rPrP ^{Res}	100
Figure 4.1	Formation and Characterization of rBvPrP ⁹⁴⁻¹⁷⁸ Fibrils.....	113
Figure 4.2	Cryo-EM Reconstruction of rBvPrP ⁹⁴⁻¹⁷⁸ Fibrils.....	116
Figure 4.3	Formation of rElkPrP ⁹⁴⁻¹⁷⁸ Fibrils.....	119
Figure 4.4	3D Classification and Refinement of Short Pitch rElkPrP ⁹⁴⁻¹⁷⁸ Polymorphs...	123

Figure 4.5	3D Classification and Refinement of a Thin Medium Pitch rElkPrP ⁹⁴⁻¹⁷⁸ Polymorph.....	125
Figure 4.6	3D Classification of Thick Medium Pitch rElkPrP ⁹⁴⁻¹⁷⁸ Polymorphs.....	127
Figure 4.7	Schematic of Types of Polymorphism Seen in rElkPrP ⁹⁴⁻¹⁷⁸ Fibrils.....	131

LIST OF TABLES

Table S1.1	Crystallographic structure determination of proto-PrP ^{Sc}	41
Table S1.2	Hydrogen bonds in proto-PrP ^{Sc}	42
Table S1.3	Crystallographic structure determination of carbamazepine.....	43
Table S1.4	Refinement of proto-PrP ^{Sc} as a function of resolution.....	44
Table 2.1	Crystallization Conditions for β 2 α 2 Loop Peptides.....	79
Table 2.2	Data collection and refinement statistics.....	80
Table 3.1	Cryo-EM data collection, refinement and validation statistics.....	85
Table 3.2	X-ray data collection and refinement statistics.....	86
Table S3.1	Properties of PrP ¹¹³ AGAAAA ¹¹⁸ and ¹¹⁹ GAVVGG ¹²⁴	102
Table S3.2	Rosetta energy and deviation from rPrP ^{Res} of repacked mutants and peptides.....	102
Table S3.3	Solvation energy of rPrP ^{Res} compared to a subset of other amyloid fibril structures.....	102
Table A.1	All Protein Constructs.....	170

ACKNOWLEDGEMENTS

Reaching back to my undergraduate education, I would like to thank my mentors Prof. Deborah Perlstein, Eric Camire, Dr. John Grossman, and Dr. Amanda Vo for teaching me, being examples of what I could aspire to be, and giving me an excellent foundation in biochemistry lab techniques that served me well in graduate school. Prof. Deborah Perlstein encouraged me to participate in my first scientific conferences, which showed me the amazing breadth of science being done across the country. With the help of Prof. Perlstein and Boston University's Conference Participation Travel Award, I was able to attend the American Society for Biochemistry and Molecular Biology meeting across the country in San Diego California as an undergraduate, where I met my thesis committee member Prof. Todd Yeates. This opportunity facilitated my introduction to structural biology.

The work presented here was supported in part by the Cellular and Molecular Biology Pre-Doctoral Training Grant (NIH T32), the Ruth L. Kirschstein Predoctoral Individual NRSA (NIH F31, 1F31 AI143368), and the Fowler Fellowship in Protein Science (UCLA).

For chapters 1 and 2, I would especially like to thank coauthors Evelyn Hernandez and Jennifer Miao for starting the project and their early and continued efforts to crystallize many of the $\beta 2\alpha 2$ loop peptides. Evelyn Hernandez carried out all biophysical characterization of $\beta 2\alpha 2$ loop crystals.

I would like to thank Dr. Duilio Cascio, Dr. Michael Sawaya, and Mike Collazo for their assistance, teaching, and advice in all aspects of crystallography. Their above and beyond efforts have benefitted not only myself, but every student, postdoc, faculty, and staff member who utilized structural biology at UCLA. I would also like to thank Prof. Christina Sigurdson and Dr. Marcin Apostol for helpful discussions and advice on not only science, but also career development.

Prof. David Eisenberg provided critical feedback, valuable insight, and words of encouragement on all scientific endeavors. His boundless knowledge of all things amyloid and structural biology gave me a great appreciation for the rich history surrounding prions and the work of all who came before me. The work presented in this thesis is built upon a strong foundation laid by David and his lab.

I would like to thank Prof. Tamir Gonen for use of electron microscopes at the Janelia Research Campus accessed through the Janelia Visiting Researcher Program. I would also like to thank former Gonen lab member Dr. Dan Shi for his teachings and imparting his wisdom in all aspects of electron microscopy at all hours of the day and night.

For chapters 3 and 4 I would like to thank Dr. Mark Arbing for advice and assistance designing and purifying the tag-free prion constructs. Prof. Z. Hong Zhou encouraged the pursuit of ambitious structural targets and provided access to the electron microscopes at the Electron Imaging Center for Nanomachines at UCLA. Dr. Peng Ge provided crucial expertise and instruction in all aspects of single particle cryo-EM including sample preparation, data collection, and processing of data from amyloid filaments.

I would like to thank Dr. Marcus Gallagher-Jones for his endless encouragement and willingness to teach me at any pace I needed throughout my graduate career. He and Jose created a welcoming environment where science was discussed with enthusiasm at every opportunity and made a field I was initially weary about – electron microscopy – seem less daunting. I could not have asked for a better labmate, scientist, and friend to look up to and learn with.

I would like to thank my labmates for their support, encouragement, and discussion. The exceptionally broad scientific interests and backgrounds of the lab brought new perspectives, methods, and approaches into view. I am glad to have been in a lab with such lively discussions, whether it be about science or tea.

Most importantly, I would like to thank my mentor Prof. Jose Rodriguez. Jose's enthusiasm for science was infectious and inspiring and working with him was the best decision I could have made. His early investment allowed me to become the scientist I am today. I was given the opportunity to travel across the country to the HHMI Janelia Research Campus where I learned MicroED sample preparation and data collection in an immersive one-on-one environment with Jose and Dr. Dan Shi over the course of a few days with almost no prior electron microscopy experience. Jose's trust forced me to become independent in these daunting tasks immediately with the added responsibility of returning high quality data to all members of the lab during each monthly trip. This combined with synchrotron trips, one-on-one lessons in MicroED data processing, and a conference where I presented a poster were all part of my lab rotation experience before even joining the lab. I would not have been afforded so many opportunities anywhere else, and I am forever grateful to Jose for it.

VITA

Education

2012-2016 Bachelor of Arts in Biochemistry and Molecular Biology, Boston University

Research and Work Experience

01/17 - 9/21 Graduate Student Researcher with Jose Rodriguez, UCLA

01/19 - 03/19 Teaching Assistant for Structural Molecular Biology: X-ray and Electron Crystallography and EM graduate course with David Eisenberg, Todd Yeates, and Jose Rodriguez, UCLA

04/17 - 08/18 Visiting Student Researcher with Tamir Gonen, Janelia Research Campus

01/17 - 06/17 Teaching Assistant for Biochemical Methods undergraduate lab with Stacie Nakamoto, UCLA

03/13 – 05/16 Undergraduate Student Researcher with Deborah Perlstein, Boston University

Honors and Awards

2020 Audree Fowler Fellowship in Protein Science

2020 Roberts A. Smith Excellence in Research Award

2019-2021 Ruth L. Kirschstein Predoctoral Individual National Research Service Award (NIH NIAID F31)

2019-2020 UCLA Clinical and Translational Science Institute (CTSI) Core Voucher Award

2017-2019 Ruth L. Kirschstein National Research Service Award GM007185 (Cellular and Molecular Biology Training Grant, UCLA, NIH T32)

2016-2018 UCLA Graduate Dean's Scholar Award

2016 Boston University Conference Participation Travel Award

2015 Boston University Student Research Award

2015 Boston University Funded Research Opportunity Grant

2012-2016 Boston University Grant

PUBLICATIONS

Catharine Shipps, H. Ray Kelly, Peter J. Dahl, Sophia M. Yi, Dennis Vu, David Boyer, **Calina Glynn**, Michael R. Sawaya, David Eisenberg, Victor S. Batista and Nikhil S. Malvankar. Intrinsic electronic conductivity of individual atomically-resolved amyloid crystals reveals micrometer-long hole hopping via tyrosines. *PNAS*, (2020).

Calina Glynn, Michael R. Sawaya, Peng Ge, Marcus Gallagher-Jones, Connor W. Short, Ronquiah Bowman, Marcin Apostol, Z. Hong Zhou, David S. Eisenberg and Jose A. Rodriguez. Cryo-EM structure of a human prion fibril with a hydrophobic, protease resistant core. *Nature Structural and Molecular Biology*, (2020).

Calina Glynn and Jose A. Rodriguez. Data-driven challenges and opportunities in crystallography. *Emerging Topics in Life Sciences*, (2019).

Marcus Gallagher-Jones, Colin Ophus, Karen C. Bustillo, David R. Boyer, Ouliana Panova, **Calina Glynn**, Chih-Te Zee, Jim Ciston, Kevin Canton Mancina, Andrew M. Minor and Jose A. Rodriguez. Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction. *Communications Biology*, (2019).

Chih-Te Zee, **Calina Glynn**, Marcus Gallagher-Jones, Jennifer Miao, Carlos G. Santiago, Duilio Cascio, Tamir Gonen, Michael R. Sawaya and Jose A. Rodriguez. Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ. *IUCrJ*, (2019).

John D. Grossman, Eric J. Camire, **Calina A. Glynn**, Christopher M. Neil, Bryan O. Seguinot and Deborah L. Perlstein. The Cfd1 Subunit of the Nbp35-Cfd1 Iron Sulfur Cluster Scaffolding Complex Controls Nucleotide Binding. *Biochemistry*, (2019).

Johan Hattne, Dan Shi, **Calina Glynn**, Chih-Te Zee, Marcus Gallagher-Jones, Michael Martynowycz, Jose A. Rodriguez, Tamir Gonen. Analysis of global and site-specific radiation damage in protein nanocrystals by MicroED. *Structure*, (2018).

Marcus Gallagher-Jones[†], **Calina Glynn**[†], David R. Boyer, Michael W. Martynowycz, Evelyn Hernandez, Jennifer Miao, Chih-Te Zee, Irina V. Novikova, Lukasz Goldschmidt, Heather T. McFarlane, Gustavo F. Helguera, James E. Evans, Michael R. Sawaya, Duilio Cascio, David Eisenberg, Tamir Gonen and Jose A. Rodriguez. Sub-ångstrom CryoEM structure of a prion protofibril reveals a polar clasp. *Nature Structural & Molecular Biology*, (2018).

Michael Martynowycz, **Calina Glynn**, Jennifer Miao, M. Jason de la Cruz, Johan Hattne, Dan Shi, Duilio Cascio, Jose Rodriguez, Tamir Gonen. MicroED Structures from Micrometer Thick Protein Crystals. *bioRxiv*, 152504 (2017).

Amanda Vo, Nicholas M. Fleischman, Mary J. Froehlich, Claudia Y. Lee, Jessica A. Cosman, **Calina A. Glynn**, Zanut O. Hassan and Deborah L. Perlstein. Identifying the protein interactions of the cytosolic iron sulfur cluster targeting complex essential for its assembly and recognition of apo-targets. *Biochemistry*, (2017).

[†] Authors contributed equally to this work

OVERVIEW

A History of Prion Structure: The importance of establishing a structural basis for prion infectivity

When protein-based infectious agents associated with the transmissible sheep wasting disease, scrapie, were first coined as prions in 1982¹, the quest to reveal the structure of an infectious protein began. In 1983 it was noted that isolation of the infectious scrapie agent from hamster brains yielded a single detectable protein - the prion protein - and protease-resistant rod-shaped aggregates that bound the amyloid binding dye Congo red². These same fibrillar aggregates were observed directly in plaques from hamster brains infected with scrapie and were recognized by antibodies raised against the prion protein (PrP). DNA sequencing of the PrP gene revealed that PrP is a natively expressed protein across mammals, but lacks protease-resistance outside of the disease context³. Likewise, the secondary structure composition of the healthy, cellular form of the prion protein (PrP^C)⁴ is notably distinct from that observed in the scrapie form (PrP^{Sc})⁵. The transition from a PrP^C structure essentially devoid of beta sheets to a PrP^{Sc} structure with over 46% estimated beta sheet content⁵ pointed to a massive structural rearrangement beyond the small transitions previously seen in enzymes and protein complexes.

At this point the existence of at least two possible structures for PrP was clear and two questions arose: 1) What is the structure of an infectious protein? And 2) What are the details of the conversion process from PrP^C to PrP^{Sc}? The path to answer these questions has not been trivial due to numerous factors. Even PrP^C is not readily expressed in large quantities in a soluble form, and purification of even truncated forms of the protein was found to require careful selection of bacterial expression system^{6,7}. In spite of extensive efforts, PrP^C segregates to inclusion bodies and tricky refolding procedures are needed to obtain pure, natively folded PrP^C without abundant truncation products⁸. These challenges created substantial hurdles for structural biology

approaches of the 1990s. Without large quantities of pure, untruncated material, crystallographic approaches to uncovering the structure of PrP^C were unsuccessful and Nuclear Magnetic Resonance (NMR) became the method of choice for visualizing the native fold of PrP^C⁹⁻¹⁴.

Attempts at deciphering the structure of PrP^{Sc} came with even greater challenges. The tendency of PrP^{Sc} to form dense meshworks of insoluble fibrils ruled out studies by both crystallography and solution state NMR. Moreover, the sheer size of the aggregates, which could be well over a megadalton, precluded clear measurement of more than a few residue interactions via solid state NMR (ssNMR). This led to fiber diffraction experiments becoming a primary source of structural information about PrP^{Sc}. This type of information, however, precludes knowledge of the structure taken on by specific residues within PrP and which secondary structures are changing between the PrP^C and PrP^{Sc} conformations. This gap in knowledge was tackled through fiber diffraction studies of synthetic peptides corresponding to segments of particular interest, including the hydrophobic region, ¹¹³AGAAAAGA¹²⁰, the first alpha helix containing residues 109-122, and a naturally occurring truncation observed in disease containing the first two alpha helices in PrP^C (residues 90-145)¹⁵. These studies demonstrated that all targets studied refolded into filamentous beta sheet structures, laying the groundwork for the hypothesis that the entirety of PrP^C undergoes a misfolding event to form PrP^{Sc}. However, how exactly these beta sheets were arranged in three dimensions to build the mysterious PrP^{Sc} structure remained confined to speculation.

While imaging fibrils of PrP^{Sc} by electron microscopy in 2002, Holger Wille observed what appeared to be regular two dimensional crystalline aggregates branching off of PrP^{Sc} rods^{16,17}. These were believed to be 2D crystals of PrP^{Sc}. The combined constraints from these images of 2D crystals and a survey of known protein folds dominated by beta sheets pinpointed a beta helix structure as most likely. Based on the ability of known beta helix structures to oligomerize and

form filaments¹⁸⁻²⁰, this was not an unreasonable conclusion at a time before any structures representing an amyloid fold had been observed or conceived of, and thus wouldn't be options considered for structure prediction.

Building Prion Structures from Peptides Up

In the early 2000s, prion researchers were not the only group facing challenges in structure determination of fibrillar aggregates. Melinda Balbirnie took a new approach to extracting atomic level structural information from amyloids²¹. After seeking out large segments within the 123 residue prion domain of the yeast prion Sup35 that would form crystals amenable to unambiguous structure determination, only fibril forming segments were found. It took synthesis of a seven residue repeating motif within a region critical to the prion properties of Sup35 to permit crystal formation alongside fibril formation. Both the fibrillar form and crystalline form of this segment - GNNQQNY - exhibited similar x-ray powder diffraction patterns with strong rings at $\sim 4.7\text{\AA}$ and $\sim 10\text{\AA}$, indicative of a shared cross beta structure for both types of aggregates. The improved resolution of the powder diffraction patterns from crystals as opposed to fibrils offered a great deal of new information about the underlying structure of the aggregates. Measurement of the unit cell along with density measurements allowed an estimate of the number of water molecules present per peptide - just four. This surprisingly low number compared to the number of residues in need of hydrogen bonding partners indicated that a vast majority of the hydrogen bonds within the structure must be satisfied by the side chains within the same peptide or from neighboring peptides instead of by water molecules. This dry structure held together by polar side chain interactions - called polar zippers first coined by Max Perutz²²⁻²⁴ - would have much lower entropy than a structure whose hydrogen bonds were largely satisfied by water molecules²⁵. This was proposed as a likely explanation for the profound stability of amyloid aggregates rich in polar residues²¹.

In the next few years, Ruben Diaz-Avalos and Melinda Balbirnie made efforts to increase the size of the GNNQQNY crystals, but struggled to form crystals larger than a micron in diameter that would be amenable to x-ray crystallography²⁶. Using electron diffraction, it was demonstrated that individual crystals were sufficiently ordered to produce discrete Bragg reflections out to $\sim 2\text{\AA}$, but crystalline domains < 1 micron across were not amenable to structure determination by synchrotron sources at the time²⁶. Likewise, these small crystalline domains posed a challenge for X-ray Free Electron Lasers, where Bragg reflections captured in each pattern may belong to different crystalline domains. This began a concurrent pursuit for the GNNQQNY structure using electron crystallography, but this quest was also wrought with challenges. Theoretical work on the severe impacts of dynamical scattering on accuracy of structure factor intensities from titans of electron microscopy cast doubt on the possibility of solving structures from three dimensional crystals²⁷⁻³². For this reason, electron diffraction was not heavily pursued for atomic structure determination of protein or peptides until a decade later. Soon after the attempts with electron crystallography, microfocus beamlines became available at synchrotrons, reducing the lower size limit for crystalline samples from tens of microns to roughly 5 microns or less in some cases. This opened up the possibility for determining crystal structures using more familiar techniques and allowed Rebecca Nelson and Michael Sawaya to determine the first atomic structures of cross-beta spines from the peptide GNNQQNY³³.

In 2007, Michael Sawaya expanded the known diversity of structures taken on by amyloid spines, which included the first atomic structure of a human prion peptide, ¹⁷⁰SNQNNF¹⁷⁵³⁴. This segment belongs to the $\beta 2\alpha 2$ loop of the mammalian prion protein, which resides in a hotspot for sequence diversity³⁵ and was computationally predicted³⁶ to form an amyloid structure similar to NNQQNY, coined a “steric zipper”. Microfocus beamlines became the primary method for deriving atomic level structural information from peptidic amyloid cores - including prions³⁷⁻³⁹ - for the next decade.

But determination of structures derived from larger amyloid core segments was hindered by the natural tendency of beta sheets to twist to favor fibril formation over crystal formation. At this point structural biology had been pushed to its limits for amyloids and even microfocus beamlines were unsuitable for determination of structures from a growing fraction of disease-relevant amyloid spines. Technological advancements were needed to provide conformational context for these amyloid backbones.

The first of these advancements came to light in 2013 with the advent of micro-crystal electron diffraction (MicroED)⁴⁰. This cryo-electron microscopy (cryo-EM) based method took advantage of the strong interaction between the electron beam and matter to allow for meaningful data collection from many fewer unit cells - or much smaller crystals - than was possible with even microfocus x-ray beamlines. In 2015, Jose Rodriguez collected continuous rotation⁴¹ MicroED data from crystals just 100-300 nm in thickness from two peptides within the toxic core of alpha synuclein⁴². Of note, these crystals were many orders of magnitude thicker than the expected usable thickness for recording reliable, quantitative, electron diffraction patterns based on calculations on the interference of multiple scattering events in reflection intensities²⁸⁻³¹. This was also the first novel structure determined by MicroED, and instilled confidence in electron diffraction of three-dimensional crystals as a reliable method for protein structural biology. The resulting 1.4Å structures from the NACore and preNACore peptides were the highest resolution achieved by a cryo-EM method at the time. Based on the combined crystal structures of these peptide building blocks, a putative model for a multi-sheet core of alpha synuclein fibrils was proposed⁴². In the next few years, MicroED unleashed a flood of new amyloid structures from crystals previously too small to be determined using x-ray sources and became a central technique for amyloid structural biologists. As these structures made their debut, concurrent work described in the appendix explored the potential of MicroED⁴³ to determine structures from crystals ranging in thickness

beyond theoretical predictions⁴⁴, as a means of quantifying sidechain- and atom-specific effects of radiation damage in cryo-EM⁴⁵, and alternative phasing methods⁴⁶.

In chapter 1, I delve into one of early drops in the ocean of amyloid MicroED structures - a 9 residue segment from the $\beta 2\alpha 2$ loop of the bank vole prion protein⁴⁷. The 0.75Å structure reached a record resolution for cryo-EM of proteinaceous material at the time and revealed the positions of all atoms in the structure, including hydrogens. The tightly packed core of the structure was inaccessible to solvents or denaturants and featured numerous intricate three-dimensional hydrogen bond networks, providing a much sought after atomic explanation for how the $\beta 2\alpha 2$ loop may contribute to the trademark stability of prions. Similar structures were observed for $\beta 2\alpha 2$ loop crystals of other mammalian prions with several key species-specific differences, which are described in detail in chapter 2.

These 9 residue peptide structures determined through technological advancements in electron diffraction built on information obtained from their 6 residue counterparts determined using microfocus x-ray beamlines, but still lacked holistic information about the structure taken on by the entire prion protein to create an infectious entity.

The Resolution Revolution and its Relevance for Amyloid Structure Determination

Great strides were made in visualizing amyloid structures using ssNMR, but the problem of structural heterogeneity within a sample left the structures taken on by individual polymorphs uncertain. To circumvent this, electron microscopy was used to average images of filaments of the same morphology to achieve higher resolution than from individual images alone. In the 1990s and 2000s, images were recorded on film, which came with severe challenges that hindered high-resolution structure determination. This meant that some of the earliest attempts at revealing

amyloid fibril structures were confined to low resolutions that didn't allow for unambiguous beta sheet position assignment⁴⁸⁻⁵⁰. As charge coupled devices (CCDs) started to replace film in the late 1990s and 2000s, some of the limitations of film could begin to be addressed⁵¹⁻⁵³. Without the need to develop each film, more quantitative assessments of structure started to become possible. The immediate readout of images from CCDs also permitted rapid data analysis and evaluation of the imaging condition, allowing microscopists to make adjustments on the fly and more efficiently utilize sparse microscope time while samples were pristine. Lastly, the adoption of digital methods of image acquisition opened doors for further advancements in detector technology and data processing software that would not have been possible using film.

One of these much needed advancements was motion correction. It had been well-known for decades that beam-induced motion was a severe resolution-limiting problem for beam-sensitive samples, like proteins, where movements of several angstroms were not uncommon^{54,55}. But until 2012, this was a problem that couldn't be addressed computationally due to the slow speed of CCD detectors. The mid 2000s brought about hybrid and direct electron detectors⁵⁶⁻⁵⁸ which had much smaller point spread functions that allowed for more accurate recording of scattering events, and by extension, made higher resolution information obtainable. These detectors were also much faster, allowing for collection of a series of several hundred millisecond exposures to be summed into one composite image. These individual frames could be aligned and summed to minimize loss of signal and blurring due to sample movement seen in a single long exposure⁵⁹. These two advancements, direct electron detectors and motion correction, were the two greatest advancements in cryo-EM that resulted in the "resolution revolution", where a massive jump in high-resolution structures - including those better than 4Å resolution - began to be reported⁶⁰. For amyloids, the 4.8Å spacing between stacked beta strands was a crucial resolution that needed to be reached in order to correctly align fibril images and trace a chain in three dimensions.

The Rise of Helical Reconstruction Software

Before the mid 1990s, processing of cryo-EM data was restricted to experts capable of writing and implementing their own data processing algorithms in house. This changed in the mid 1990s when data processing packages began to be released for public use, starting with some of the most commonly used packages today including SPIDER⁶¹, IMAGIC⁶² and EMAN⁶³. For helical reconstruction, an approach similar to that taken for single particles was used. EMAN's helixboxer was used for automated segmentation of helical filaments into overlapping boxes such that each view of the helix would be represented. In the 2010s, structural biologists with an interest in amyloid fibril structure began to develop their own packages to tackle problems more commonly associated with amyloids, such as deviations from ideal symmetry and filament bending. In 2014 and 2015, Alexis Rohou and Mattias Schmidt demonstrated that their new software package, FREALIX, could be used to reconstruct fibrils of A β at 5Å resolution - just slightly coarser than the resolution needed to visualize stacked beta strand separation^{64,65}. 2017 brought cryo-EM structures from two amyloids that broke the 4.7Å barrier - two structures at 3.4Å of tau derived from the brains of patients with Alzheimer's disease⁶⁶ processed using RELION⁶⁷ and one at 4.0Å of recombinant A β ⁶⁸ processed using SPARX⁶⁹. After this landmark achievement in amyloid structural biology, many amyloid structures reaching beyond 4.7Å from recombinant and tissue sources started to populate the literature⁷⁰⁻¹⁰⁶.

Single Particle Cryo-EM of Prions: What is the structure of an infectious protein?

After decades of advancements in prion biology, x-ray crystallography, MicroED, and single particle cryo-EM of amyloid filaments, we are now equipped to come back to our central question: what is the structure of an infectious protein?

In addition to the challenges associated with determining single particle cryo-EM structures of amyloids in general, prions pose additional challenges. The infectious, or potentially infectious, nature of prion aggregates introduces the need for additional safety measures and facilities designed for manipulation and storage of prion-containing samples. On top of the sparsity of these types of laboratories and facilities, a vast majority are equipped for experiments using molecular biology and animal models, but not structural biology.

While cryo-EM has become more accessible and more instruments have become available at universities, institutes, national user facilities, and individual laboratories, microscopes typically used for high-resolution cryo-EM data collection can run upwards of 2 million dollars upfront with annual operating costs and service contracts running over \$150,000 per year. The costs, stringent safety requirements, somewhat niche nature of prion research compared to more common neurodegenerative diseases, and the need for a high level of expertise in two fields with steep learning curves, reduces the number of facilities equipped for cryo-EM studies of infectious prion samples to a number that could be tallied on one hand.

There were a number of amyloid fibrils whose high-resolution structures were readily attainable by cryo-EM as soon as the technology caught up^{66,71,74,81,90,94}, but prions were not one of them. Many prion structures had been proposed based on low resolution cryo-EM data^{16,17,107-110} but the many strains of prions and heterogeneity within those strains resulted in many disparate models based on small numbers of fibrils from a large population. These discrepancies even at low resolution alluded to challenges to come in determining high resolution prion structures. Additionally, it is well established that the prion aggregates with high titers of infectivity are not typically long, single fibrils of clearly defined morphology that would be amenable to single particle cryo-EM. Even for samples with tangled nets of long fibrils present, the resistance of prions to typical treatments used to improve the monodispersity of other amyloid fibrils were often futile.

The Cryo-EM Resolution Revolution Reaches Prions

The challenges of isolating monodisperse fibrils of sufficiently homogeneous morphology hindered structure determination of any mammalian prion until 2020. In chapter 3, I discuss the far from trivial task of preparing human prion fibrils for single particle cryo-EM. The resulting 3.5Å structure represented 70% of fibrils present in the sample and was the first prion structure reaching a resolution compatible with unambiguous residue assignment⁹⁶. This human prion structure was composed of two intertwined, hairpin shaped protofilaments made up of residues 106-145 and presented with chaotrope-, protease-, and SDS-resistant properties that are often considered a trademark of prions. These properties could be explained by the compact nature of the structure, where small hydrophobic residues constituted the interface between protofilaments and a pair of interacting polar residues served as a linchpin holding the hairpin structure closed. This structure was the first of a series of prion structures reported in the next year. A comparison and discussion of these cryo-EM structures, including two other polymorphs of human prion fibrils^{97,106}, the familial mutant E196K¹⁰⁴, hamster passaged scrapie 263K prions¹⁰⁵, and anchorless Rocky Mountain Labs (aRML) prions isolated from mice¹⁰⁵, will be presented in the discussion at the end of the thesis. In the final chapter, I detail my efforts towards determining structures of other mammalian prions, mainly preliminary cryo-EM studies of bank vole and elk prion fibrils.

References

1. Bolton, D. C., McKinley, M. P. & Prusiner, S. B. Identification of a protein that purifies with the scrapie prion. *Science* **218**, 1309–1311 (1982).
2. McKinley, M. P., Bolton, D. C. & Prusiner, S. B. A protease-resistant protein is a structural component of the Scrapie prion. *Cell* **35**, 57–62 (1983).
3. Dearmond, S. J. *et al.* Identification of prion amyloid filaments in scrapie-infected brain. *Cell* **41**, 221–235 (1985).
4. Huang, Z. *et al.* Proposed three-dimensional structure for the cellular prion protein. *Proceedings of the National Academy of Sciences* **91**, 7139–7143 (1994).
5. Caughey, B. W. *et al.* Secondary structure analysis of the scrapie-associated protein PrP²⁷⁻³⁰ in water by infrared spectroscopy. *Biochemistry* **30**, 7672–7680 (1991).
6. Mehlhorn, I. *et al.* High-Level Expression and Characterization of a Purified 142-Residue Polypeptide of the Prion Protein. *Biochemistry* **35**, 5528–5537 (1996).
7. Muramoto, T., Scott, M., Cohen, F. E. & Prusiner, S. B. Recombinant scrapie-like prion protein of 106 amino acids is soluble. *Proceedings of the National Academy of Sciences* **93**, 15457–15462 (1996).
8. Zahn, R., von Schroetter, C. & Wüthrich, K. Human prion proteins expressed in *Escherichia coli* and purified by high-affinity column refolding. *FEBS Letters* **417**, 400–404 (1997).
9. Riek, R. *et al.* NMR structure of the mouse prion protein domain PrP(121–231). *Nature* **382**, 180–182 (1996).
10. Billeter, M. *et al.* Prion protein NMR structure and species barrier for prion diseases. *Proceedings of the National Academy of Sciences* **94**, 7281–7285 (1997).
11. Zahn, R. *et al.* NMR solution structure of the human prion protein. *PNAS* **97**, 145–150 (2000).

12. García, F. L., Zahn, R., Riek, R. & Wüthrich, K. NMR structure of the bovine prion protein. *PNAS* **97**, 8334–8339 (2000).
13. Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. & Wuthrich, K. Prion protein NMR structures of elk and of mouse/elk hybrids. *Proceedings of the National Academy of Sciences* **102**, 646–650 (2005).
14. Lysek, D. A. *et al.* Prion protein NMR structures of cats, dogs, pigs, and sheep. *PNAS* **102**, 640–645 (2005).
15. Nguyen, J. T. *et al.* X-ray Diffraction of Scrapie Prion Rods and PrP Peptides. *Journal of Molecular Biology* **252**, 412–422 (1995).
16. Wille, H. *et al.* Structural studies of the scrapie prion protein by electron crystallography. *Proc Natl Acad Sci U S A* **99**, 3563–3568 (2002).
17. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. Evidence for assembly of prions with left-handed β -helices into trimers. *PNAS* **101**, 8342–8347 (2004).
18. Schuler, B., Rachel, R. & Seckler, R. Formation of Fibrous Aggregates from a Non-native Intermediate: The Isolated P22 Tailspike β -Helix Domain *. *Journal of Biological Chemistry* **274**, 18589–18596 (1999).
19. Kreisberg, J. F., Betts, S. D. & King, J. Beta-helix core packing within the triple-stranded oligomerization domain of the P22 tailspike. *Protein Sci* **9**, 2338–2343 (2000).
20. Jenkins, J., Mayans, O. & Pickersgill, R. Structure and Evolution of Parallel β -Helix Proteins. *Journal of Structural Biology* **122**, 236–246 (1998).
21. Balbirnie, M., Grothe, R. & Eisenberg, D. S. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β -sheet structure for amyloid. *PNAS* **98**, 2375–2380 (2001).
22. Perutz, M. F., Staden, R., Moens, L. & De Baere, I. Polar zippers. *Current Biology* **3**, 249–253 (1993).

23. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proceedings of the National Academy of Sciences* **91**, 5355–5358 (1994).
24. Stott, K., Blackburn, J. M., Butler, P. J. & Perutz, M. Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proceedings of the National Academy of Sciences* **92**, 6509–6513 (1995).
25. Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **264**, 670–670 (1994).
26. Diaz-Avalos, R. *et al.* Cross-beta Order and Diversity in Nanocrystals of an Amyloid-forming Peptide. *Journal of Molecular Biology* **330**, 1165–1175 (2003).
27. Cowley, J. M. & Moodie, A. F. The scattering of electrons by atoms and crystals. I. A new theoretical approach. *Acta Crystallographica* **10**, 609–619 (1957).
28. Jap, B. K. & Glaeser, R. M. The scattering of high-energy electrons. II. Quantitative validity domains of the single-scattering approximations for organic crystals. *Acta Cryst A* **36**, 57–67 (1980).
29. Ho, M.-H., Jap, B. K. & Glaeser, R. M. Validity domain of the weak-phase-object approximation for electron diffraction of thin protein crystals. *Acta Cryst A* **44**, 878–884 (1988).
30. Glaeser, R. M. & Ceska, T. A. High-voltage electron diffraction from bacteriorhodopsin (purple membrane) is measurably dynamical. *Acta Cryst A* **45**, 620–628 (1989).
31. Glaeser, R. M. & Downing, K. H. High-resolution electron crystallography of protein molecules. *Ultramicroscopy* **52**, 478–486 (1993).
32. Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin. *Journal of Molecular Biology* **259**, 393–421 (1996).

33. Nelson, R. *et al.* Structure of the cross- β spine of amyloid-like fibrils. *Nature* **435**, 773 (2005).
34. Sawaya, M. R. *et al.* Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453 (2007).
35. Wopfner, F. *et al.* Analysis of 27 mammalian and 9 avian PrPs reveals high conservation of flexible regions of the prion protein 11Edited by A. R. Fersht. *Journal of Molecular Biology* **289**, 1163–1178 (1999).
36. Thompson, M. J. *et al.* The 3D profile method for identifying fibril-forming segments of proteins. *Proceedings of the National Academy of Sciences* **103**, 4074–4078 (2006).
37. Apostol, M. I., Sawaya, M. R., Cascio, D. & Eisenberg, D. Crystallographic Studies of Prion Protein (PrP) Segments Suggest How Structural Changes Encoded by Polymorphism at Residue 129 Modulate Susceptibility to Human Prion Disease*. *Journal of Biological Chemistry* **285**, 29671–29675 (2010).
38. Sievers, S. A. *et al.* Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100 (2011).
39. Rodriguez, J. A., Jiang, L. & Eisenberg, D. S. Toward the Atomic Structure of PrPSc. *Cold Spring Harb Perspect Biol* **9**, a031336 (2017).
40. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, (2013).
41. Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat Meth* **11**, 927–930 (2014).
42. Rodriguez, J. A. *et al.* Structure of the toxic core of α -synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
43. Glynn, C. & Rodriguez, J. A. Data-driven challenges and opportunities in crystallography. *Emerg Top Life Sci* **3**, 423–432 (2019).

44. Martynowycz, M. W. *et al.* MicroED Structures from Micrometer Thick Protein Crystals. *bioRxiv* 152504 (2017) doi:10.1101/152504.
45. Hattne, J. *et al.* Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* **26**, 759-766.e4 (2018).
46. Zee, C. *et al.* Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ. *IUCrJ* **6**, 197–205 (2019).
47. Gallagher-Jones, M. *et al.* Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nature Structural & Molecular Biology* **25**, 131 (2018).
48. Crowther, R. A. Straight and paired helical filaments in Alzheimer disease have a common structural unit. *Proceedings of the National Academy of Sciences* **88**, 2288–2292 (1991).
49. Jiménez, J. L. *et al.* Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *The EMBO Journal* **18**, 815–821 (1999).
50. Jimenez, J. L. *et al.* The protofilament structure of insulin amyloid fibrils. *Proceedings of the National Academy of Sciences* **99**, 9196–9201 (2002).
51. De Ruijter, W. J. Imaging properties and applications of slow-scan charge-coupled device cameras suitable for electron microscopy. *Micron* **26**, 247–275 (1995).
52. Stewart, P. L., Cary, R. B., Peterson, S. R. & Chiu, C. Y. Digitally collected cryo-electron micrographs for single particle reconstruction. *Microscopy Research and Technique* **49**, 224–232 (2000).
53. Zhou, Z. H. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Current Opinion in Structural Biology* **18**, 218–228 (2008).
54. Henderson, R. & Glaeser, R. M. Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals. *Ultramicroscopy* **16**, 139–150 (1985).
55. Glaeser, R. M. & Hall, R. J. Reaching the Information Limit in Cryo-EM of Biological Macromolecules: Experimental Aspects. *Biophysical Journal* **100**, 2331–2337 (2011).

56. Faruqi, A. R., Henderson, R., Pryddetch, M., Allport, P. & Evans, A. Direct single electron detection with a CMOS detector for electron microscopy. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **546**, 170–175 (2005).
57. Faruqi, A. R. & Henderson, R. Electronic detectors for electron microscopy. *Current Opinion in Structural Biology* **17**, 549–555 (2007).
58. Deptuch, G. *et al.* Direct electron imaging in electron microscopy with monolithic active pixel sensors. *Ultramicroscopy* **107**, 674–684 (2007).
59. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* **177**, 630–637 (2012).
60. Orlov, I. *et al.* The integrative role of cryo electron microscopy in molecular and cellular structural biology. *Biology of the Cell* **109**, 81–93 (2017).
61. Frank, J. *et al.* SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. *Journal of Structural Biology* **116**, 190–199 (1996).
62. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A New Generation of the IMAGIC Image Processing System. *Journal of Structural Biology* **116**, 17–24 (1996).
63. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *Journal of Structural Biology* **128**, 82–97 (1999).
64. Rohou, A. & Grigorieff, N. FREALIX: Model-based refinement of helical filament structures from electron micrographs. *J Struct Biol* **186**, 234–244 (2014).
65. Schmidt, M. *et al.* Peptide dimer structure in an A β (1–42) fibril visualized with cryo-EM. *Proc Natl Acad Sci USA* **112**, 11858–11863 (2015).
66. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).

67. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *Journal of Structural Biology* **198**, 163–176 (2017).
68. Gremer, L. *et al.* Fibril structure of amyloid- β (1–42) by cryo–electron microscopy. *Science* **358**, 116–119 (2017).
69. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *Journal of Structural Biology* **157**, 47–55 (2007).
70. Li, B. *et al.* Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications* **9**, (2018).
71. Iadanza, M. G. *et al.* The structure of a β 2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat Commun* **9**, 4517 (2018).
72. Falcon, B. *et al.* Structures of filaments from Pick’s disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
73. Guerrero-Ferreira, R. *et al.* Cryo-EM structure of alpha-synuclein fibrils. *eLife* **7**, e36402 (2018).
74. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
75. Guerrero-Ferreira, R. *et al.* Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife* **8**, e48907 (2019).
76. Ni, X., McGlinchey, R. P., Jiang, J. & Lee, J. C. Structural Insights into α -Synuclein Fibril Polymorphism: Effects of Parkinson’s Disease-Related C-Terminal Truncations. *J Mol Biol* **431**, 3913–3919 (2019).
77. Kollmer, M. *et al.* Cryo-EM structure and polymorphism of A β amyloid fibrils purified from Alzheimer’s brain tissue. *Nat Commun* **10**, 4760 (2019).
78. Rademaker, L. *et al.* Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun* **10**, 1103 (2019).

79. Swuec, P. *et al.* Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nat Commun* **10**, 1269 (2019).
80. Liberta, F. *et al.* Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nat Commun* **10**, 1104 (2019).
81. Röder, C. *et al.* Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy. *Nat Commun* **10**, 3754 (2019).
82. Zhang, W. *et al.* Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *eLife* **8**, e43584 (2019).
83. Schmidt, M. *et al.* Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat Commun* **10**, 5008 (2019).
84. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat Struct Mol Biol* **26**, 619–627 (2019).
85. Boyer, D. R. *et al.* Structures of fibrils formed by α -synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nat Struct Mol Biol* **26**, 1044–1052 (2019).
86. Boyer, D. R. *et al.* The α -synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *PNAS* **117**, 3592–3602 (2020).
87. Sun, Y. *et al.* Cryo-EM structure of full-length α -synuclein amyloid fibril with Parkinson's disease familial A53T mutation. *Cell Res* **30**, 360–362 (2020).
88. Zhao, K. *et al.* Parkinson's disease associated mutation E46K of α -synuclein triggers the formation of a distinct fibril structure. *Nat Commun* **11**, 2643 (2020).
89. Zhao, K. *et al.* Parkinson's disease-related phosphorylation at Tyr39 rearranges α -synuclein amyloid fibril structure revealed by cryo-EM. *PNAS* **117**, 20305–20315 (2020).
90. Schweighauser, M. *et al.* Structures of α -synuclein filaments from multiple system atrophy. *Nature* **585**, 464–469 (2020).
91. Lu, J. *et al.* CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nat Commun* **11**, 4090 (2020).

92. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nat Struct Mol Biol* **27**, 653–659 (2020).
93. Röder, C. *et al.* Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid- β fibrils. *Nat Struct Mol Biol* **27**, 660–667 (2020).
94. Zhang, W. *et al.* Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
95. Gallardo, R. *et al.* Fibril structures of diabetes-related amylin variants reveal a basis for surface-templated assembly. *Nat Struct Mol Biol* **27**, 1048–1056 (2020).
96. Glynn, C. *et al.* Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. *Nature Structural & Molecular Biology* 1–7 (2020) doi:10.1038/s41594-020-0403-y.
97. Wang, L.-Q. *et al.* Cryo-EM structure of an amyloid fibril formed by full-length human prion protein. *Nat Struct Mol Biol* **27**, 598–602 (2020).
98. Hervas, R. *et al.* Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in *Drosophila*. *Science* **367**, 1230–1234 (2020).
99. Lee, M., Ghosh, U., Thurber, K. R., Kato, M. & Tycko, R. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. *Nat Commun* **11**, 5735 (2020).
100. Ghosh, U., Thurber, K. R., Yau, W.-M. & Tycko, R. Molecular structure of a prevalent amyloid- β fibril polymorph from Alzheimer's disease brain tissue. *PNAS* **118**, (2021).
101. Rademaker, L. *et al.* Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. *Nat Commun* **12**, 875 (2021).
102. Bansal, A. *et al.* AA amyloid fibrils from diseased tissue are structurally different from in vitro formed SAA fibrils. *Nat Commun* **12**, 1013 (2021).
103. Li, Q., Babinchak, W. M. & Surewicz, W. K. Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat Commun* **12**, 1620 (2021).

104. Wang, L.-Q. *et al.* Familial prion disease-related mutation E196K displays a novel amyloid fibril structure revealed by cryo-EM. *bioRxiv* 2021.02.18.431846 (2021)
doi:10.1101/2021.02.18.431846.
105. Kraus, A. *et al.* High-resolution structure and strain comparison of infectious mammalian prions. *Molecular Cell* (2021) doi:10.1016/j.molcel.2021.08.011.
106. Li, Q., Jaroniec, C. P. & Surewicz, W. K. *Cryo-EM structure of disease-related prion fibrils provides insights into seeding barriers.* 2021.08.10.455830
<https://www.biorxiv.org/content/10.1101/2021.08.10.455830v1> (2021)
doi:10.1101/2021.08.10.455830.
107. Tattum, M. H. *et al.* Elongated Oligomers Assemble into Mammalian PrP Amyloid Fibrils. *Journal of Molecular Biology* **357**, 975–985 (2006).
108. Terry, C. *et al.* Ex vivo mammalian prions are formed of paired double helical prion protein fibrils. *Open Biology* **6**, 160035.
109. Vázquez-Fernández, E. *et al.* The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy. *PLOS Pathogens* **12**, e1005835 (2016).
110. Kamali-Jamil, R. *et al.* The ultrastructure of infectious L-type bovine spongiform encephalopathy prions constrains molecular models. *PLOS Pathogens* **17**, e1009628 (2021).

Chapter 1

Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp

The work described in this chapter has been reproduced from:

Marcus Gallagher-Jones, Calina Glynn, David R. Boyer, Michael W. Martynowycz, Evelyn Hernandez, Jennifer Miao, Chih-Te Zee, Irina V. Novikova, Lukasz Goldschmidt, Heather T. McFarlane, Gustavo F. Helguera, James E. Evans, Michael R. Sawaya, Duilio Cascio, David S. Eisenberg, Tamir Gonen and Jose A. Rodriguez. "Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp." *Nature Structural and Molecular Biology*, 2018.

Copyright 2018

Marcus Gallagher-Jones, Calina Glynn, David R. Boyer, Michael W. Martynowycz, Evelyn Hernandez, Jennifer Miao, Chih-Te Zee, Irina V. Novikova, Lukasz Goldschmidt, Heather T. McFarlane, Gustavo F. Helguera, James E. Evans, Michael R. Sawaya, Duilio Cascio, David S. Eisenberg, Tamir Gonen and Jose A. Rodriguez.

Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp

Marcus Gallagher-Jones¹, Calina Glynn¹, David R. Boyer², Michael W. Martynowycz³, Evelyn Hernandez¹, Jennifer Miao¹, Chih-Te Zee¹, Irina V. Novikova⁴, Lukasz Goldschmidt², Heather T. McFarlane², Gustavo F. Helguera⁵, James E. Evans⁴, Michael R. Sawaya², Duilio Cascio², David S. Eisenberg^{1,2}, Tamir Gonen³ and Jose A. Rodriguez^{1*}

The atomic structure of the infectious, protease-resistant, β -sheet-rich and fibrillar mammalian prion remains unknown. Through the cryo-EM method MicroED, we reveal the sub-ångström-resolution structure of a protofibril formed by a wild-type segment from the $\beta 2$ - $\alpha 2$ loop of the bank vole prion protein. The structure of this protofibril reveals a stabilizing network of hydrogen bonds that link polar zippers within a sheet, producing motifs we have named 'polar clasps'.

Micro electron diffraction (MicroED) is a cryo-EM method that facilitates the determination of atomic structures from sub-micrometer-thin protein nanocrystals and fragmented crystallites^{1–3}. Several high-resolution amyloid structures have been determined by MicroED, including two structures from the toxic core of the Parkinson's-associated protein α -synuclein⁴, two from the type-2-diabetes-associated protein IAPP⁵, and five ab initio structures^{2,6}. These structures add to those of the fibril cores of α -synuclein⁷, amyloid- β (ref. ⁸), tau⁹, and a fungal prion¹⁰ recently determined by complementary methods. However, an atomic structure of the infectious scrapie form of the mammalian prion protein (PrP^{Sc}) remains unknown¹¹. PrP^{Sc} shares some structural hallmarks of amyloid, appearing as rope-like filaments or rods made of tightly mating β -sheets^{12,13}. However, PrP^{Sc} also differs from other amyloids: it resists proteolysis and denaturation, is infectious, and can spread within and between species to cause disease^{11,14,15}.

To evaluate the source of PrP^{Sc} stability, we investigated segments in mammalian prion proteins (PrP) that might form the core of PrP^{Sc} fibrils. Informed by structure-based prediction of amyloid-prone sequences¹⁶, we identified 168-QYNNQNNFV-176, a segment of the $\beta 2$ - $\alpha 2$ loop of the bank vole (*Myodes glareolus*) PrP (Fig. 1), a universal prion acceptor^{17,18}. This segment lies within the predicted cross- β core of PrP fibrils (Supplementary Fig. 1), shows high conservation in rodents and other mammals (Supplementary Fig. 1), and is rich in asparagines^{19,20}, which may stabilize prion fibrils²¹. At sub-millimolar concentrations, this segment produces highly ordered aggregates (Supplementary Fig. 2) that, when aligned and illuminated by X-rays, produce cross- β diffraction (Supplementary Fig. 2). Like PrP^{Sc}, aggregates formed by this segment are resistant to high concentrations of urea, guanidine, and a range of pH, but are sensitive to sodium hydroxide (Supplementary Fig. 3). Given its shared biophysical

properties with PrP^{Sc}, we labeled our segment proto-PrP^{Sc} and set out to uncover the structural basis for its stability.

From optimized microcrystals (Supplementary Fig. 4), we determined a microfocus X-ray diffraction structure of proto-PrP^{Sc} to 1.1-Å resolution by molecular replacement (Supplementary Fig. 4 and Supplementary Table 1). The conditions in which we observed microcrystals of proto-PrP^{Sc} also produced showers of nanocrystals, evident in electron micrographs (Fig. 1). These nanocrystals diffracted to 0.72 Å by MicroED (Fig. 1). Merging diffraction from multiple crystals, we achieved a high-completeness 0.75-Å-resolution dataset (Supplementary Table 1). From these data, we obtained an ab initio solution that was similar to our microfocal X-ray diffraction structure of proto-PrP^{Sc} (Supplementary Fig. 4) and suitable for atomic refinement (Supplementary Fig. 5). This ultrahigh-resolution MicroED structure of proto-PrP^{Sc} shows features that are invisible in the X-ray structure and critical to our understanding of its stability.

The structure of proto-PrP^{Sc} reveals a prion protofibril with amyloid-like features: β -strands parallel and in register as a class 2 steric zipper¹² in which sheets pair front to back (Fig. 2). Two tightly mating curved sheets make up the proto-PrP^{Sc} fibril (Fig. 2), although side chains in these sheets interdigitate less in proto-PrP^{Sc} than those observed in conventional amyloid structures¹². Although sheets stack in a parallel face-to-back configuration, the convex face of one sheet nestles against the concave face of its neighbor, approximately 10.3 Å away, with a high degree of surface complementarity (S_p , 0.807) (Fig. 2). The interface between these sheets is large, concealing 204.5 Å² per strand, and is entirely devoid of waters at its core (Fig. 1). Atoms in the MicroED structure of proto-PrP^{Sc} are extremely well ordered, with an average B factor of 6.0 Å² overall and 2.8 Å² within its core (Supplementary Fig. 6 and Supplementary Table 1). These values are less than half of the overall B factor in our X-ray diffraction structure of proto-PrP^{Sc} (Supplementary Table 1), thus confirming a greater degree of order in our nanocrystallites compared to that of larger microcrystals of the same segment.

A 3D network of hydrogen bonds stabilizes proto-PrP^{Sc} (Supplementary Table 2); hydrogens at its core participate in intra-residue C5 bonds²² (Fig. 3), and asparagine and glutamine residues stack along its fibril axis (Fig. 2). Glutamines and asparagines

¹Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA, USA. ²Department 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, CA, USA. ³Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. ⁴Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA. ⁵Laboratory of Pharmaceutical Biotechnology, Institute of Biology and Experimental Medicine, Buenos Aires, Argentina. Marcus Gallagher-Jones and Calina Glynn contributed equally to this work. *e-mail: jrodriguez@mbl.ucla.edu

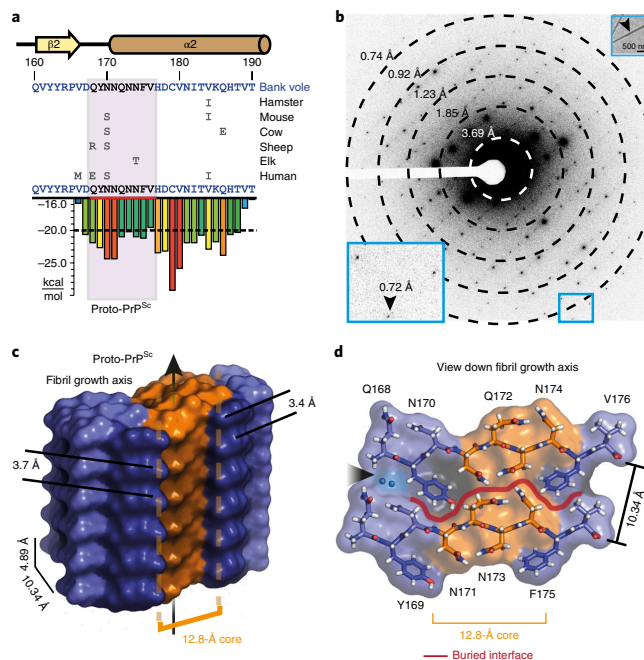


Fig. 1 | Structure of proto-PrP^{Sc} from sub-ångström diffraction of nanoscale crystals by MicroED. **a**, Sequence of the bank vole PrP β 2- α 2 loop, highlighting local secondary structure and variation in sequence compared to hamster, mouse, cow, sheep, elk, and human. The sequence of proto-PrP^{Sc} is highlighted in a plot that indicates propensity for amyloid formation (scored in kcal/mol). The region encoding proto-PrP^{Sc} is boxed in magenta. **b**, Single MicroED image showing strong diffraction to 0.72 Å (inset). A second inset shows a proto-PrP^{Sc} crystal probed by MicroED (arrow). **c, d**, Side (**c**) and top (**d**) views of the proto-PrP^{Sc} molecular surface. Stacked aromatics are colored dark purple, a conformationally constrained core region is colored orange, two waters are noted with a black arrow, and a red line indicates the interface between sheets in **d**.

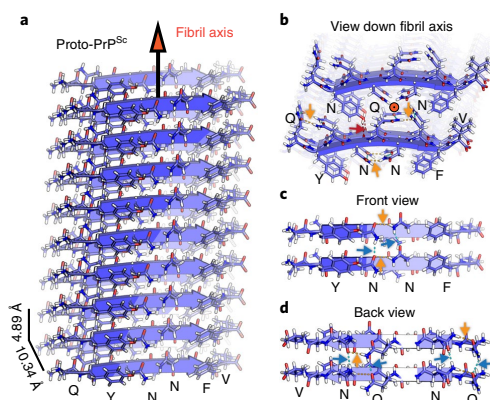


Fig. 2 | Filament structure of proto-PrP^{Sc} and its stabilizing network of hydrogen bonds. **a**, Structure of proto-PrP^{Sc} determined by MicroED shown as an assembly of eighteen strands. Distances between strands and sheets and the fibril growth axis are labeled. **b**, A view down the fibril axis highlighting the interface between adjacent sheets. An orange circle denotes the direction of the fibril axis. **c, d**, Front (**c**) and back (**d**) views of two strands of a sheet. Polar contacts are labeled as orange arrows for intrachain and blue arrows for interchain. A single intersheet hydrogen bond is labeled by a red arrow.

in proto-PrP^{Sc} form networks of hydrogen bonds reminiscent of proton wires²³, polar ladders²⁴, and the polar zippers first proposed by Max Perutz^{25,26} (Supplementary Fig. 7). Neighboring polar ladders in proto-PrP^{Sc} are additionally linked by hydrogen bonds within a strand (Fig. 3), a motif we refer to as a ‘polar clasp’ (Supplementary Fig. 7). Stacks of phenylalanine and tyrosine residues shield clasp at the core of proto-PrP^{Sc} (Figs. 1 and 2). The importance of this aromatic embrace is underscored by a lack of clasps in structures of shorter segments from this region of PrP that lack Tyr169 (refs. 12,27). On this evidence, we hypothesize that polar clasps and stacked aromatic residues act in concert to stabilize proto-PrP^{Sc}, as they might for PrP^{Sc}.

Our hypothesis of proto-PrP^{Sc} stability relies on the locations of hydrogen atoms throughout the structure. Hydrogens in the MicroED structure of proto-PrP^{Sc} are unambiguously assigned, informed by pronounced difference density in ultrahigh-resolution maps (Supplementary Figs. 5, 8, and 9 and Supplementary Table 4). Hydrogens at the core of proto-PrP^{Sc} are as evident as those seen in structures of small organic compounds determined by electron diffraction²⁸, including our own structure of carbamazepine (Supplementary Fig. 10 and Supplementary Table 3). Density in ultrahigh-resolution maps of proto-PrP^{Sc} suggests that hydrogen may occupy positions that deviate from idealized geometry (Supplementary Fig. 5). Improved hydrogen positions indicated by ultrahigh-resolution maps in MicroED could bolster the accuracy of calculations based on observed hydrogen bond networks.

Ultrahigh-resolution maps of proto-PrP^{Sc} mirror features that appear in electron density maps of the highest resolution X-ray

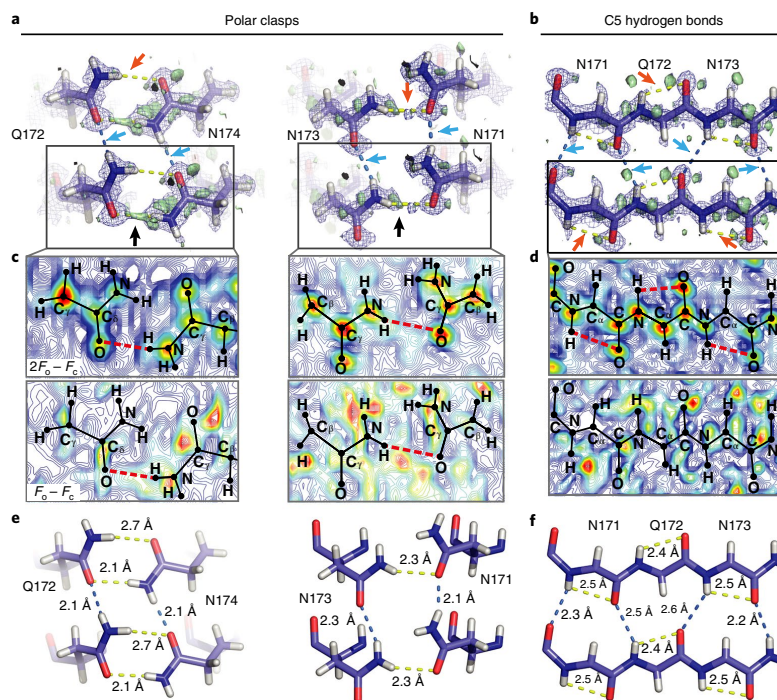


Fig. 3 | Sub-ångström cryo-EM maps show stabilizing hydrogen bonds in proto-PrP^{Sc}. **a**, Polar clasps formed by hydrogen bonds that link Q172 to N174 and N171 to N173. **b**, C5 hydrogen bonds (orange arrows) in the context of backbone hydrogen bonds (blue arrows) at the core of proto-PrP^{Sc}. Bridging density can be seen in both the $2F_o - F_c$ and $F_o - F_c$ maps (purple mesh and green isosurface, respectively) for a subset of the hydrogen bonds (black arrows). **c,d**, Contour maps show $2F_o - F_c$ or $F_o - F_c$ density for polar clasps (**c**) and C5 bonds (**d**). Contours are averaged over a ~ 1.5 -Å-thick volume with high contour values shown in red, low values in blue, and hydrogen bonds as dashed red lines. **e,f**, Distances for both intrachain (yellow lines) and interchain (blue lines) hydrogen bonds are highlighted in polar clasps (**e**) and C5 bonds (**f**).

structures and show similar residual density, which in the latter is assigned to valence electrons²⁹ (Supplementary Fig. 9). In addition, in polar clasps at the core of proto-PrP^{Sc}, density is seen surrounding donor hydrogens and oxygen acceptors (Fig. 3 and Supplementary Figs. 5 and 9). This density is not explicitly accounted for by our atomic model during refinement and is not evident in maps of other small molecules determined by MicroED or in lower-resolution maps of proto-PrP^{Sc} (Supplementary Figs. 8 and 10 and Supplementary Table 4). However, this density is preserved across maps produced by refinement programs with different scattering factor libraries (Supplementary Fig. 11). A full account of the complex features present in ultrahigh-resolution cryo-EM maps may require advances in refinement that incorporate molecular vibrations, crystal bending, non-kinematic scattering and chemical bonding^{29–33}.

Through MicroED, cryo-EM surpasses the 1-Å barrier and yields the structure of proto-PrP^{Sc}, a denaturant-resistant prion protofibril. Cryo-EM maps of proto-PrP^{Sc} at 0.75-Å resolution reveal a 3D network of stabilizing hydrogen bonds that link residues between and within its β -strands through polar clasps.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41594-017-0018-0>.

Received: 9 November 2017; Accepted: 7 December 2017;
Published online: 15 January 2018

References

- Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).
- de la Cruz, M. J. et al. Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 (2017).
- Rodriguez, J. A., Eisenberg, D. S. & Gonen, T. Taking the measure of MicroED. *Curr. Opin. Struct. Biol.* **46**, 79–86 (2017).
- Rodriguez, J. A. et al. Structure of the toxic core of α -synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).
- Krotee, P. et al. Atomic structures of fibrillar segments of hIAPP suggest tightly mated β -sheets are important for cytotoxicity. *eLife* **6**, e19273 (2017).
- Sawaya, M. R. et al. Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl. Acad. Sci. USA* **113**, 11232–11236 (2016).
- Tuttle, M. D. et al. Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein. *Nat. Struct. Mol. Biol.* **23**, 409–415 (2016).
- Schmidt, M. et al. Peptide dimer structure in an $\alpha\beta$ (1–42) fibril visualized with cryo-EM. *Proc. Natl. Acad. Sci. USA* **112**, 11858–11863 (2015).
- Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
- Wasmer, C. et al. Amyloid fibrils of the HET-s(218–289) prion form a β solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526 (2008).
- Rodriguez, J. A., Jiang, L. & Eisenberg, D. S. Toward the atomic structure of PrP^{Sc}. *Cold Spring Harb. Perspect. Biol.* **9**, a031336 (2017).

12. Sawaya, M. R. et al. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
13. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
14. McKinley, M. P., Bolton, D. C. & Prusiner, S. B. A protease-resistant protein is a structural component of the scrapie prion. *Cell* **35**, 57–62 (1983).
15. Kurt, T. D. & Sigurdson, C. J. Cross-species transmission of CWD prions. *Prion* **10**, 83–91 (2016).
16. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* **107**, 3487–3492 (2010).
17. Watts, J. C. et al. Evidence that bank vole PrP is a universal acceptor for prions. *PLoS Pathog.* **10**, e1003990 (2014).
18. Kurt, T. et al. The molecular basis for cross-species prion transmission. *FASEB J.* **30**, 814.7 (2016).
19. Halfmann, R. et al. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol. Cell.* **43**, 72–84 (2011).
20. Zambrano, R. et al. PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res.* **43**, W331–W337 (2015).
21. Kurt, T. D. et al. Asparagine and glutamine ladders promote cross-species prion conversion. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.M117.794107> (2017).
22. Newberry, R. W. & Raines, R. T. A prevalent intraresidue hydrogen bond stabilizes proteins. *Nat. Chem. Biol.* **12**, 1084–1088 (2016).
23. Nagle, J. F. & Morowitz, H. J. Molecular mechanisms for proton transport in membranes. *Proc. Natl. Acad. Sci. USA* **75**, 298–302 (1978).
24. Yoder, M. D., Lietzke, S. E. & Jurnak, F. Unusual structural features in the parallel β -helix in pectate lyases. *Structure* **1**, 241–251 (1993).
25. Perutz, M. F., Staden, R., Moens, L. & De Baere, I. Polar zippers. *Curr. Biol.* **3**, 249–253 (1993).
26. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA* **91**, 5355–5358 (1994).
27. Wiltzius, J. J. W. et al. Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16**, 973–978 (2009).
28. Palatinus, L. et al. Hydrogen positions in single nanocrystals revealed by electron diffraction. *Science* **355**, 166–169 (2017).
29. Hirano, Y., Takeda, K. & Miki, K. Charge-density analysis of an iron-sulfur protein at an ultra-high resolution of 0.48 Å. *Nature* **534**, 281–284 (2016).
30. Glaeser, R. M. & Downing, K. H. High-resolution electron crystallography of protein molecules. *Ultramicroscopy* **52**, 478–486 (1993).
31. Zhong, S., Dadarlat, V. M., Glaeser, R. M., Head-Gordon, T. & Downing, K. H. Modeling chemical bonding effects for protein electron crystallography: the transferable fragmental electrostatic potential (TFESP) method. *Acta Crystallogr. A* **58**, 162–170 (2002).
32. Jelsch, C. et al. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. USA* **97**, 3171–3176 (2000).
33. Zuo, J. M., Kim, M., O’Keeffe, M. & Spence, J. C. H. Direct observation of *d*-orbital holes and Cu–Cu bonding in Cu₂O. *Nature* **401**, 49–52 (1999).

Acknowledgements

We thank C. Ophus (Molecular Foundry-NCM), J. Miao (UCLA), and H. Nelson (UCLA) for helpful discussions. This work is supported by DOE Grant DE-FC02-02ER63421, the EICN in the CNSI at UCLA, the Janelia Research Visitor Program, and the NE-CAT beamline 24-ID-E, funded by NIH-NIGMS P41 GM103403. This work was also supported by STROBE: a National Science Foundation Science and Technology Center under Grant No. DMR 1548924. G.E.H. is supported by the BECAR program, Fundación Bunge y Born, Fundación Williams, and Fundación René Barón, and is a member of CONICET. M.G.J. is supported by a QCB Collaboratory Postdoctoral Fellowship. C.G. is supported by the Cellular and Molecular Biology Training Program. J.A.R. is supported by the Searle Scholar Program and the Beckman Young Investigator Program. T.G. and D.S.E. are supported by the Howard Hughes Medical Institute (HHMI), and J.E.E. and I.V.N. are supported by the DOE-BER Molecules to Mesoscale Bioimaging Project FWP #66382.

Author contributions

J.A.R. directed the work. J.A.R., J.M., E.H., M.W.M., and C.G. grew, evaluated, and optimized crystals. J.A.R., D.R.B., C.G., M.R.S., H.T.M., M.G.-J., C.-T.Z., I.V.N., J.E.E., and D.C. collected data. J.A.R., C.G., J.M., M.G.-J., M.R.S., D.C., M.W.M., G.E.H., E.H., L.G., D.S.E., and T.G. analyzed the data. C.G., M.G.-J., and J.A.R. wrote the article, with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-017-0018-0>.

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

Correspondence and requests for materials should be addressed to J.A.R.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Phylogenetic and sequence analysis of bank vole PrP 27–30. The top 130 most similar wild-type sequences to bank vole PrP 90–231 were generated using NCBI BLAST. Of these 130 sequences, all sequences containing the proto-PrP^{Sc} peptide were selected and aligned.

Characterization of proto-PrP^{Sc} peptide. The synthetic peptide, QYNNQNFFV, corresponding to residues 168–176 of the bank vole prion protein was purchased from GenScript. The peptide used for our experiments was received at >98% purity achieved by reverse-phase HPLC. The peptide was qualified on a Bruker (ultrafleXtreme) MALDI-TOF/TOF mass spectrometer, reported as *m/z* (intensity, arbitrary units). The spectrum has a mass list that includes a [M+H]⁺ peak at 1,141.3 Da (expected 1,140.5 Da), a [M+Na]⁺ peak at 1,163.4 Da (expected 1,162.49 Da), and a [M+K]⁺ peak at 1179.3 Da (1,178.46 Da).

Characterization of carbamazepine. Lyophilized powder of carbamazepine (5H-dibenzo[*b,f*]azepine-5-carboxamide, C₁₅H₁₂N₂O) with purity >99% was purchased from Sigma-Aldrich and crystallized without further purification.

Aggregation of proto-PrP^{Sc}. Proto-PrP^{Sc} peptide was solubilized in ultrapure water at 0.22–1.75 mM. 50 μL of each sample was added to a 96-well clear flat-bottom plate in triplicate and evaluated for aggregate formation by reading absorbance at 350 nm on an Infinite M1000 Pro plate reader (Tecan). Readings were measured immediately after solubilization after 3 h and after 6 h of shaking at 900 r.p.m. at 37 °C. Wells were imaged after 6 h using a Leica M205 C light microscope (Leica Microsystems) and after 3 and 6 h by electron microscopy as described below. Electron microscope images are representative of more than five images taken at each concentration and time point.

Fibril diffraction from proto-PrP^{Sc} aggregates. Solutions containing aggregates of proto-PrP^{Sc} were clarified by centrifugation. Pelleted aggregates were resuspended in a concentrated volume in water, applied between two pulled capillaries and left to dry overnight. Oriented aggregates formed between the capillary ends were resupplied with additional solution containing aggregates and left to dry again. This process was repeated several times to grow the bulk of aligned aggregates on a capillary. Aligned aggregates were then diffracted using 5-min exposures to a FRE+ rotating anode generator with VARIMAX HR confocal optics producing Cu K-α radiation (Rigaku, Tokyo, Japan) and detected using a RIGAKU R-Axis HTC imaging plate detector at a distance of 156 mm from the source.

Chemical denaturation of proto-PrP^{Sc} aggregates. Proto-PrP^{Sc} was solubilized at approximately 3.5 mM and allowed to form aggregates. This solution was diluted to approximately 1:4 before treatment with either 0.5–6.0 M urea, 0.5–4.5 M guanidinium-HCl, 0.75 M HCl, 0.75 M of either MES, pH 2, acetate, pH 4, MES, pH 6, Tris-HCl, pH 8 or 10, or 0.75 M NaOH. Aggregate content was measured by absorbance compared to a control solution consisting of 0.01% (w/v) 1-μm latex spheres in water. Spectra across the visible range (250–700 nm) were collected using a Nanodrop One (Thermo).

Growth of proto-PrP^{Sc} crystals. Peptide powder was weighed and dissolved in ultrapure water at near maximum solubility, 3.5 mM. Crystals were grown at room temperature by the hanging-drop method in a 96-well Wizard screen. Crystals appeared in various conditions and were further optimized in 24-well hanging-drop vapor diffusion experiments. The best crystals of proto-PrP^{Sc} grew within 24 h at a peptide concentration of 1.75 mM in the presence of 0.1 M 2-(*N*-morpholino) ethanesulfonic acid (MES), pH 6.0, and either 10% ethanol or 10% 2-Methyl-2,4-pentanediol (MPD).

Transmission electron microscopy. Approximately 2 μL of aggregated proto-PrP^{Sc} were applied to 300-mesh Formvar-carbon coated grids (Ted Pella Inc.) for 2 min before excess liquid was removed and grids left to dry. Grids were imaged either on a Tecnai T12 or F20 electron microscope (Thermo Fisher, formerly FEI). Samples were imaged at a magnification of 2,100× with a dose rate of <30 e⁻/Å².

Microfocus X-ray data collection. Crystals grown in 0.1 M MES, pH 6.0, and 10% ethanol and mixed with 100% glycerol as cryoprotectant were harvested from 24-well hanging drops using MiTeGen loops and flash frozen in liquid nitrogen. 73 diffraction images were collected, each spanning a 3° wedge, from a single crystal at a temperature of 100 K at the advanced photon source (APS) beamline 24-ID-E, equipped with an ADSC Q315 CCD detector, using a 10-μm beam with a 0.98-Å wavelength.

Microfocus X-ray data processing and structure determination. Diffraction images collected from a single crystal of proto-PrP^{Sc} were indexed and integrated in Denzo, yielding a dataset with 80.75% overall completeness at 1.1-Å resolution in space group *P1*. A suitable molecular replacement solution was obtained from this data using the PHASER program and an idealized β-strand nonapeptide alanine model as a probe. The model was refined using REFMAC against the measured data to a final $R_{\text{work}}/R_{\text{free}}$ of 0.14/0.16.

MicroED sample preparation. Nanoscale needle crystals of proto-PrP^{Sc} were grown in batch in 0.1 M MES, pH 6.0, and 10% ethanol. Crystals were diluted in mother liquor and fragmented by force of pipetting to create an approximately monodisperse solution of crystals. Carbamazepine was crystallized in batches by dilution into neat isopropanol at 100 μg/mL. 2 μL of either solution were placed on a holey carbon grid (1/4, 2/2, 2/4, #300 copper; Ted Pella Inc.) before plunge freezing into liquid ethane and transferring into liquid nitrogen for storage. Grids were held by a liquid-nitrogen-cooled Gatan 626 cryo-holder for transfer into and manipulation within the electron microscope.

MicroED data collection. MicroED data collection from nine sub-micron-thick needle crystals, and a single sub-micron carbamazepine crystal was performed as previously described³⁴. Briefly, crystals of either proto-PrP^{Sc} or carbamazepine lying in a frozen-hydrated state on holey carbon grids were inspected visually in overfocused diffraction mode on either a cryo-cooled FEI Tecnai F20 microscope operated at 200 kV (Janelia Research Campus) or a Titan environmental TEM operated at 300 kV (Environmental Molecular Sciences Lab, PNNL). Diffraction patterns used for structure determination were collected on a TVIPS TemCam-F416 CMOS detector in rolling-shutter mode with 3-s exposures while proto-PrP^{Sc} crystals were unidirectionally rotated at a constant rate of 0.27° s⁻¹ over angular wedges ranging from -55° to +72°. A single carbamazepine crystal was rotated at a speed of 0.2° s⁻¹ over an angular wedge ranging between -45° to +45° with 5-s exposures. Beam intensity was held constant, with an average dose rate of 0.003–0.005 e⁻ Å⁻² sec⁻¹, corresponding to a total dose of ~1–3 e⁻ Å⁻² per dataset. We used a camera length of 520 mm, the equivalent of a sample-to-detector distance of 950 mm in a corresponding lensless system. All diffraction was performed using a circular selected area aperture of ~1 μm² in projection.

MicroED data processing. Diffraction movies were converted to the SMV file format using TVIPS tools as previously described³⁵. Indexing and integration were performed in XDS³⁶. Integrated diffraction intensities from partial datasets of nine different proto-PrP^{Sc} crystals were sorted and merged in XSCALE³⁶. Merged intensities were converted to amplitudes at various resolution cutoffs to produce separate 1.1-Å, 1.0-Å, 0.9-Å, 0.8-Å, and 0.75-Å datasets. Ab initio structure determination was performed on each of these datasets using SHELXD³⁷. Phases obtained from the atomic assembly generated by direct methods were used to produce maps of sufficient quality for subsequent model building in Coot³⁸ and refinement in Phenix³⁹ using electron scattering form factors to produce a final structure in space group *P1* with a final $R_{\text{work}}/R_{\text{free}}$ of 24/25. Refinement in REFMAC was carried out in parallel to a final $R_{\text{work}}/R_{\text{free}}$ of 23/25. The structure refined in Phenix (PDB 6AXZ) was used in all subsequent analysis and is shown in figures. Ab initio structure determination for carbamazepine was performed in SHELXT³⁷ in which a solution was found in space group *P2*₁ with no errors in chemical assignment or atom positions for all carbon, nitrogen, and oxygen atoms. This solution was refined in SHELXL³⁷ using electron form factors to an *R* value of 21.8%. A structure with hydrogen positions refined to best match difference density in the map lowered the *R* value to 19.8%.

Analysis of buried surface area (*S_b*) and surface complementarity (*S_c*) for proto-PrP^{Sc}. The structure of proto-PrP^{Sc} was used to calculate both *S_b* and *S_c* from an assembly consisting of two sheets generated by translational symmetry, each consisting of ten stacked β-strands. *S_b* was computed as an average of the buried surface area per chain in our assembly, calculated as the difference between the sum of the solvent-accessible surface area of the two sheets and the solvent-accessible surface area of the entire complex, divided by the total number of strands in both sheets. *S_c* was calculated using the CCP4 suite for all points at the interface between the two aforementioned sheets.

Comparison of X-ray and MicroED structures. Sequence alignment and structural superposition of both an individual chain and an assembly of two sheets from X-ray and MicroED structures of proto-PrP^{Sc} were performed in PyMOL⁴⁰. This alignment produced an all-atom r.m.s.d. of 0.162 Å, on which structural similarity was assessed.

Analysis of aromatic residues in proto-PrP^{Sc}. Distances were measured between stacked aromatic residues that flank the core of proto-PrP^{Sc}. Aromatics stacked between strands along planes separated by 3.7 or 3.4 Å for Y169 or F175, respectively, in a parallel-displaced configuration⁴¹. We measured the intersheet hydrogen bond created by Y169 to the backbone carbonyl of N171 on an opposing strand. Both Y169 and F175 formed aromatic ladders that channel polar residues into a 12.8-Å-long region at the core.

Analysis of hydrogen bond networks. Informed by the locations of hydrogens, visible in our maps at a 0.7–2.5-σ range, we evaluated hydrogen bond networks determined by the program HBplus (v3.06)⁴² and by manual inspection using PyMOL. We measured asparagine ladders between residues N170, N171, N173, and N174 and the corresponding asparagine residues on the strands above and below. We also measured glutamine ladders between residues Q168 and Q172 and their corresponding residues on strands above and below. We analyzed distances

associated with intrastrand hydrogen bonds formed by three pairs of residues: Q168–N170, N171–N173 and Q172–N174. Polar contacts were also measured between Q168–N170, with a 3.1-Å donor–acceptor (D–A) distance and a 2.3-Å H–O distance, Q172–N174, with a 3.0-Å D–A distance and a 2.1-Å H–O distance and N171–N173, with a 3.1-Å D–A distance and a 2.3-Å H–O distance. In each case, the linked residues faced the same side of the β -strand and bridged residues at positions (N/Q)⁺ and (N)⁺ within the strand. While the HBplus program does not identify C5 hydrogen bonds in our structure, we measured these bonds in residues N171, Q172, and N173, on the basis of the criteria that carbonyl oxygens bond with intraresidue amide protons if their geometry permits, with H–O distances shorter than 2.5 Å (ref. ²⁵).

Amylome profiling. A subset of the predicted amyloyme was analyzed¹⁶, consisting of six-residue segments found to score favorably when threaded onto a template based on the structure of the yeast prion NNQQNY. We chose all segments that scored two s.d. better than the mean score for all peptides (Z score > 2). This subset of six-residue segments represents 95,381 out of 7,900,599 total segments, or 1.2% of all possible segments of this size in the human proteome. For each segment, we searched for a (Y/F) X (N/Q) X (N/Q) X (Y/F) motif across the region of the protein to which the segment belonged; a ten-residue window including two residues upstream and downstream of a profiled segment. QYNNQNNFV satisfies these metrics.

Calculation of contour maps. $2F_o - F_c$ and $F_o - F_c$ density maps were calculated from the final refined MTZ file using the FFT tool in CCP4. Maps were converted to MRC format in Chimera and imported into MATLAB. Contour plots were calculated such that the number of contours spanned the minimum to the maximum values of the maps with intervals of one s.d. (σ) between contour levels.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.




Data availability. Atomic coordinates and structure factors for proto-PrP^{Sc} have been deposited in the EMDB and wwPDB; the structure obtained using MicroED is under accession codes EMD-7017 and PDB 6AXZ; the X-ray crystallography structure is under accession code PDB 6BTK. The structure of carbamazepine using microED has been deposited in the EMDB with accession code EMD-7287. Source data for all figures and files is available from the authors upon reasonable request, please see author contributions for specific datasets.

References

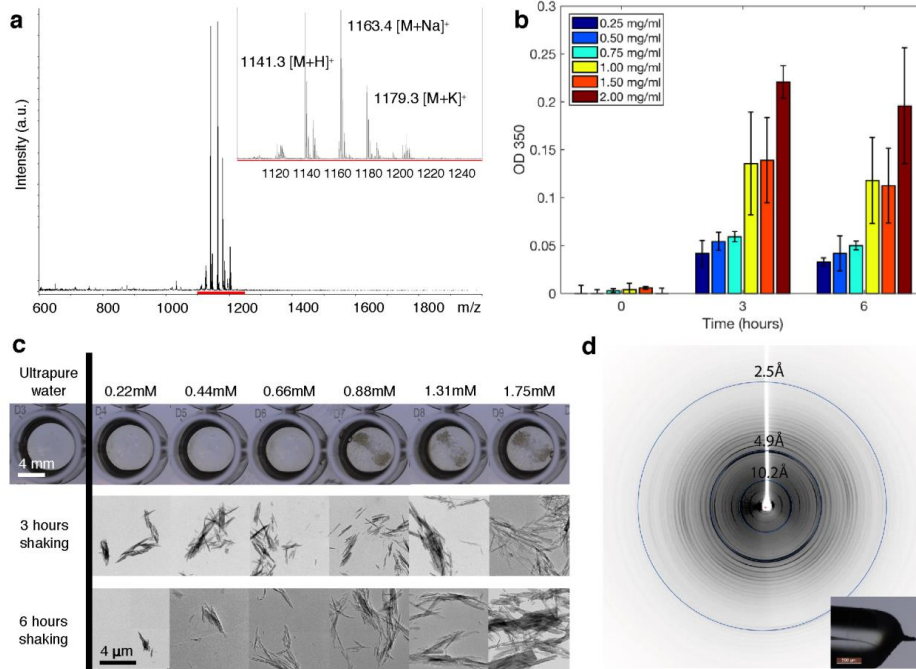
- Shi, D. et al. The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 (2016).
- Hatne, J., Shi, D., de la Cruz, M. J., Reyes, F. E. & Gonen, T. Modeling truncated pixel values of faint reflections in MicroED images. *J. Appl. Crystallogr.* **49**, 1029–1034 (2016).
- Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
- Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
- Delano, W.L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
- McGaughey, G. B., Gagné, M. & Rappé, A. K. π -Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **273**, 15458–15463 (1998).
- McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).

In the format provided by the authors and unedited.

Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp

Marcus Gallagher-Jones¹, Calina Glynn¹, David R. Boyer², Michael W. Martynowycz³, Evelyn Hernandez¹, Jennifer Miao ¹, Chih-Te Zee¹, Irina V. Novikova⁴, Lukasz Goldschmidt², Heather T. McFarlane², Gustavo F. Helguera⁵, James E. Evans⁴, Michael R. Sawaya², Duilio Cascio², David S. Eisenberg ², Tamir Gonen³ and Jose A. Rodriguez ^{1*}

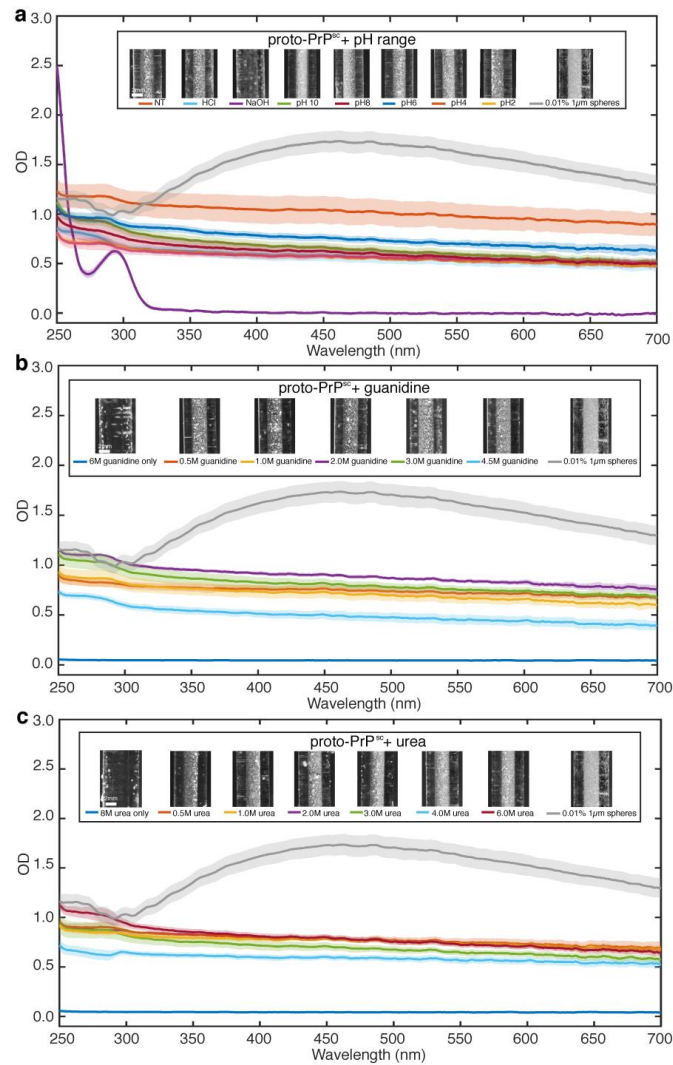
¹Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA, USA. ²Department of Biological Chemistry and Department of Chemistry and Biochemistry, University of California Los Angeles, Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA, USA. ³Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. ⁴Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA. ⁵Laboratory of Pharmaceutical Biotechnology, Institute of Biology and Experimental Medicine, Buenos Aires, Argentina. Marcus Gallagher-Jones and Calina Glynn contributed equally to this work. *e-mail: jrodriguez@mbi.ucla.edu



Supplementary Figure 2

Characterization of proto-PrP^{Sc} aggregates.

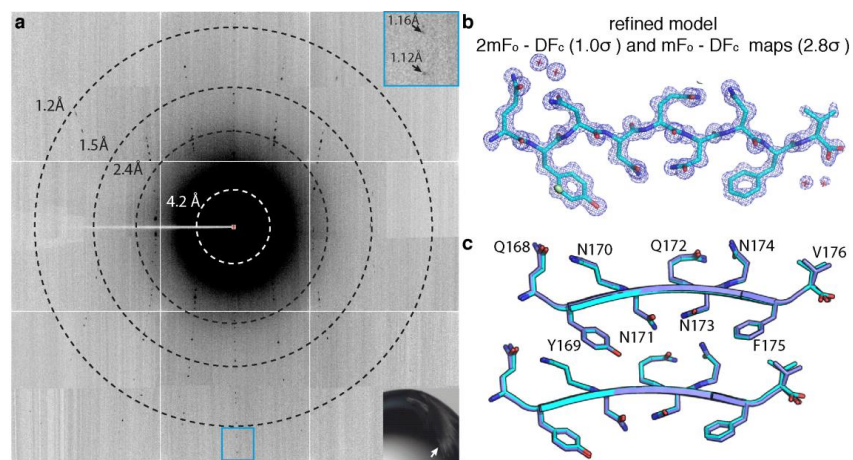
a, MALDI-TOF/TOF mass spectrum of proto-PrP^{Sc} (Exact Mass of 1140.5 Da). **b**, Detection of Proto-PrP^{Sc} aggregates at an optical wavelength of 350 nm. Colors indicate solutions with increasing peptide concentrations with optical density plotted over the course of several hours. Values represent the mean OD₃₅₀ values and error bars represent ± 1 standard deviation ($n = 3$, technical replicates). **c**, Light microscope images of peptide solutions in a 96-well plate and electron micrographs of samples from those wells measured after 3 and 6 hours of incubation. **d**, Fiber diffraction with 10.2Å, 4.9Å, and 2.5Å rings labeled. Inset shows optical microscope image of the fibril from which diffraction was obtained.



Supplementary Figure 3

Physical and chemical denaturation of proto-PrP^{Sc} aggregates.

Visible light spectra (250-700nm) show absorbance from proto-PrP^{Sc} aggregates exposed to: **a**, a range of pH **b**, concentrations of guanidinium-HCl and **c**, concentrations of urea. Insets show each sample in a glass capillary above its corresponding figure legend. A 0.01% (w/v) solution of one μm latex spheres (grey lines) was used as a baseline against which to gauge particle density in various samples. The mean value of three replicates is shown (solid lines) with the shadowed range around each line representing ± 1 standard deviation ($n = 3$, technical replicates).



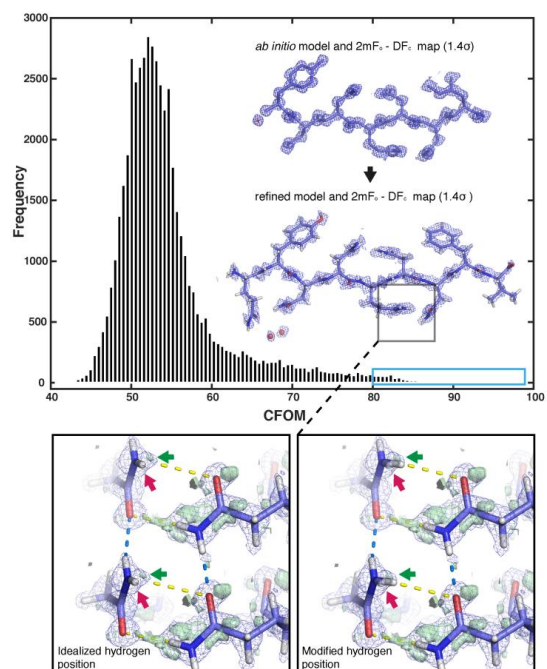
Supplementary Figure 4

Microfocal x-ray structure determination of proto-PrP^{Sc} and comparison to MicroED structure of proto-PrP^{Sc}.

a, Diffraction image collected by microfocal x-ray diffraction. Lower right inset shows hanging drop with a cluster of microcrystals (white arrow). Blue box in upper right shows a magnified view of high-resolution reflections near 1.1 Å (black arrows) from the lower quadrant.

b, Structure of proto-PrP^{Sc} determined by microfocal x-ray crystallography shown as a stick model. Waters are shown as cross marks, 2F_o-F_c density is shown at 1σ (blue mesh) and F_o-F_c at 2.8σ (green isosurface).

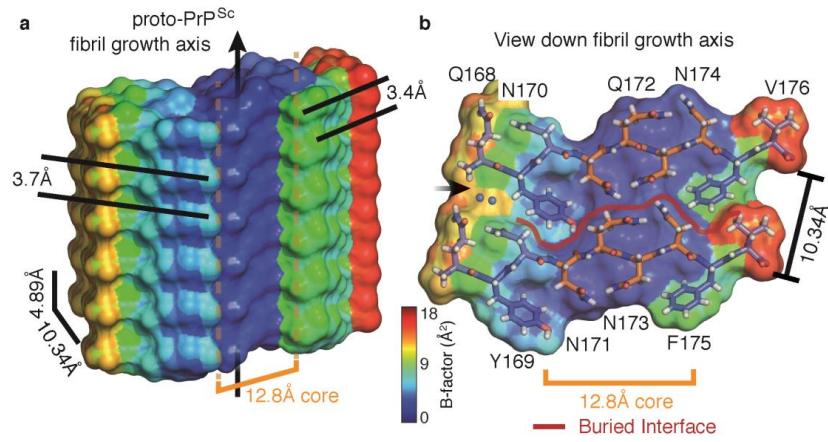
c, Alignment of microfocal x-ray (blue) and MicroED (purple) structures. The peptide backbone is represented as an idealized beta strand and side chains as stick models. Residues are labeled according to their numbering in the sequence of bank vole prion.



Supplementary Figure 5

Ab initio structure determination of proto-PrP^{Sc}.

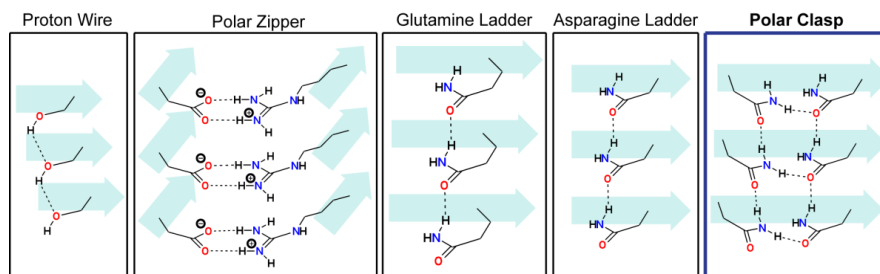
CFOM distribution obtained after 50,000 independent trials of the SHELXD program. The *ab initio* model (top inset) results from the best CFOM in this distribution; the final solution is shown below. The blue box (CFOM > 80) defines regions where correct solutions are expected. The lower insets demonstrate the need to modify hydrogen positions from their idealized placement (magenta arrow) to better match the residual density (green arrow). Hydrogen bonding positions for the final model are shown as dashed lines.



Supplementary Figure 6

B factors in proto-PrP^{Sc}.

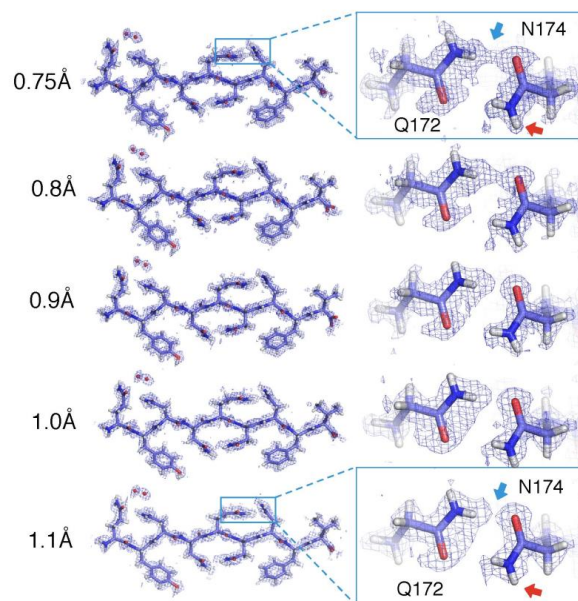
a, Side and **b**, top view of an isosurface model of proto-PrP^{Sc} analogous to those in Figure 1, but now colored by B-factor, where blue indicates low B-factors while red indicates high B-factors. A stick model is shown overlaid onto the isosurface model as a guide. The B-factor scale ranges from 0 to 18.0 Å².



Supplementary Figure 7

Hydrogen bond networks formed by polar side chains in β -sheets.

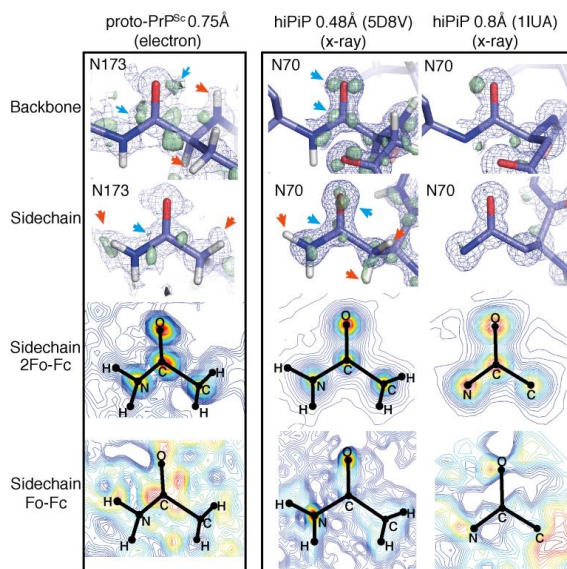
Five examples of hydrogen bonding networks are shown. From left to right, they are the proton wire (Nagle, J. F. & Morowitz, H. J. Molecular mechanisms for proton transport in membranes. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 298–302 (1978).), polar zipper (Perutz, M. F., Staden, R., Moens, L. & De Baere, I. Polar zippers. *Curr. Biol. CB* **3**, 249–253 (1993).), glutamine ladder (Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5355–5358 (1994).), asparagine ladder (Yoder, M. D., Lietzke, S. E. & Jurnak, F. Unusual structural features in the parallel β -helix in pectate lyases. *Structure* **1**, 241–251 (1993).), and the polar clasp (blue outline). For each, three layers are shown, each representing three beta strands stacked along a sheet. Hydrogen bonds are shown as dashed lines.



Supplementary Figure 8

Refinement of proto-PrP^{Sc} as a function of resolution.

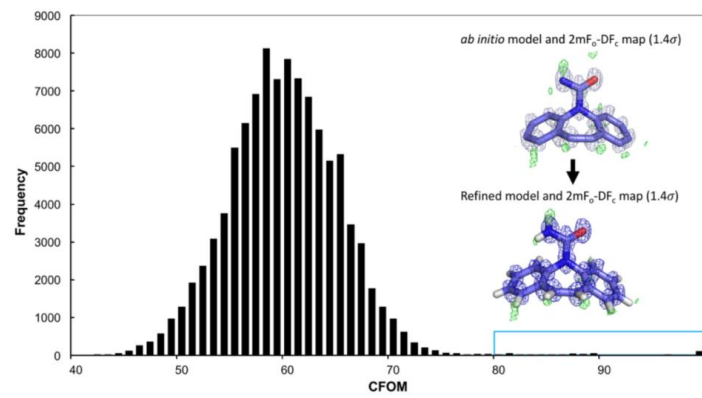
High resolution refinement of proto-PrP^{Sc} allows visualization of density for hydrogen atoms in density maps (orange arrows) at 0.7 σ ; the same atoms cannot be seen at lower resolution. Bridging density for the hydrogen bond between Q172 and N174 (blue arrows) can similarly only be seen at sub-ångstrom resolution.



Supplementary Figure 9

Comparison of ultrahigh-resolution features in maps solved by MicroED maps and X-ray crystallography.

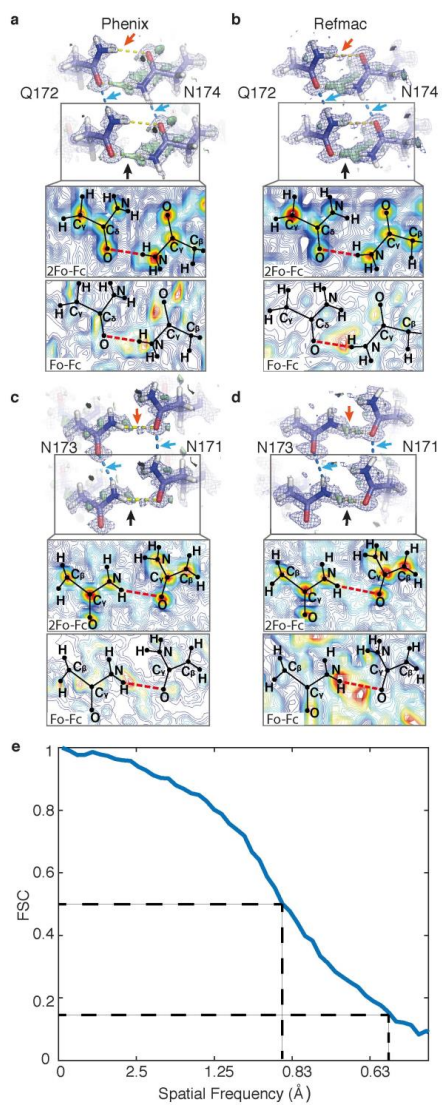
Comparison of ultrahigh resolution features in MicroED maps of proto-PrP^{Sc} to those of 0.48Å and 0.8Å x-ray crystallographic maps of the hiPiP protein (PDB IDs: 1IUA and 5D8V) (Hirano, Y., Takeda, K. & Miki, K. Charge-density analysis of an iron sulfur protein at an ultra-high resolution of 0.48 Å. *Nature* **534**, 281–284 (2016); Liu, L., Nogi, T., Kobayashi, M., Nozawa, T. & Miki, K. Ultrahigh-resolution structure of high-potential iron sulfur protein from *Thermochromatium tepidum*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1085–1091 (2002).). 2mFoDFc maps are rendered as a blue mesh (0.8σ) and mFo-DFc maps are rendered as a solid green isosurface (2.8σ). Density for hydrogen atoms are indicated by red arrows and residual density in the difference map is indicated by blue arrows. The sidechain panel also includes contour maps for 2F_o-F_c and F_o-F_c density.



Supplementary Figure 10

Ab initio structure determination of carbamazepine.

CFOM distribution obtained after 100,000 independent trials of the SHELXD program. The *ab initio* model (top inset) results from the best CFOM in this distribution; the final solution is shown below. 2F_o-F_c maps are shown in blue, F_o-F_c maps in green inform the locations of hydrogens in the model. The blue box (CFOM > 80) defines the region where correct solutions are expected.



Supplementary Figure 11

Comparison of ultra-high-resolution features in MicroED maps of proto-PrP^{Sc} refined in different software packages.

MicroED maps of proto-PrP^{Sc} were refined by either the Phenix (a and c) or Refmac (b and d) software packages. 2mFo-DFc maps are rendered as a blue mesh (0.8 σ) and mFo-DFc maps are rendered as a solid green isosurface (2.8 σ). Arrows indicate bridging hydrogen bonds. **b**, Fourier shell correlation between the maps produced by each refinement package, with a dashed black line indicating FSC cutoffs of 0.5 and 0.143.

Supplementary Tables for “Sub-ångstrom cryo-EM structure of a prion protofibril reveals a polar clasp”

Supplementary Table 1. Crystallographic structure determination of proto-PrP^{Sc}.

proto-PrP ^{Sc}	X-ray	MicroED (Phenix)	MicroED (Refmac)
Data collection			
Space group	P1	P1	
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	4.87 10.11 30.42	4.94 10.34 31.15	
α , β , γ (°)	93.36 91.15 101.66	94.26 92.38 102.20	
Resolution (Å)	30.35 - 1.10 (1.14 - 1.10)	10.34-0.75 (0.78-0.75) ^a	
<i>R</i> _{merge}	0.099 (0.405)	0.232 (0.638)	
<i>R</i> _{meas}	0.142 (0.573)	0.250 (0.729)	
<i>R</i> _{pim}	0.098 (0.405)	0.090 (0.319)	
<i>I</i> / σ (<i>I</i>)	6.20 (1.92)	4.57 (1.77)	
<i>CC</i> _{1/2}	0.986 (0.546)	0.982 (0.209)	
Completeness (%)	80.75 (46.22)	96.95 (96.17)	
No. Reflections	3059 (127)	43252 (2359)	
Redundancy	1.6 (1.2)	5.8 (4.4)	
Refinement			
No. reflections	1866 (110)	7473 (720)	6800 (470)
<i>R</i> _{work}	0.139 (0.29)	0.242 (0.37)	0.233 (0.36)
<i>R</i> _{free}	0.162 (0.23)	0.246 (0.39)	0.254 (0.35)
No. atoms			
Total	85	152	152
Waters	4	2	2
Hydrogens Possible	70	69	69
Hydrogens Modeled ^b	0	51	61
<i>B</i> factors			
Protein	12.27	6.00	5.88
Water	23.09	11.59	11.56
R.m.s. deviations			
Bond lengths (Å)	0.007	0.014	0.034
Bond angles (°)	0.99	1.01	2.21

^a Values in parentheses are for the highest-resolution shell.

^b Hydrogens modelled include only those for which density is observed in 2Fo-Fc maps at 0.70 σ or higher.

Supplementary Table 2. Hydrogen bonds in proto-PrP^{Sc}.

	Donor name ^a	Atom type	Acceptor name ^a	Atom type	DA dist.	DA pair ^b	Gap	CA-CA dist.	DHA angle	H-A dist.	H-AA angle	D-AA angle
1	A0168-GLN	NE2	B0168-GLN	OE1	3.02	SS	-	4.90	136.7	2.21	137.7	150.8
2	A0170-ASN	ND2	A0168-GLN	OE1	3.11	SS	2	6.48	176.4	2.11	108.9	109.4
3	A0169-TYR	N	B0168-GLN	O	3.00	MM	-	6.32	164.6	2.17	163.5	167.8
4	B0170-ASN	N	A0169-TYR	O	2.85	MM	-	6.00	163.4	2.01	166	170.9
5	B0170-ASN	ND2	A0170-ASN	OD1	2.90	SS	-	4.90	159.5	1.94	160.1	167.0
6	A0171-ASN	N	B0170-ASN	O	3.01	MM	-	6.24	144.0	2.27	151.4	161.1
7	B0172-GLN	N	A0171-ASN	O	3.19	MM	-	6.48	142.2	2.46	139.6	149.0
8	A0171-ASN	ND2	B0171-ASN	OD1	2.87	SS	-	4.90	151.1	1.95	153.7	163.4
9	A0173-ASN	ND2	A0171-ASN	OD1	3.09	SS	2	6.93	146.1	2.21	106.4	99.4
10	A0172-GLN	NE2	A0174-ASN	OD1	2.99	SS	2	7.07	173.3	2.00	108.7	110.9
11	A0172-GLN	NE2	B0172-GLN	OE1	2.92	SS	-	4.90	164.1	1.95	162.3	167.7
12	C0171-ASN	ND2	A0172-GLN	OE1	3.04	SS	-	8.49	131.4	2.28	113.5	102.0
13	B0174-ASN	N	A0173-ASN	O	3.06	MM	-	6.48	160.0	2.24	159.1	164.6
14	A0173-ASN	ND2	B0173-ASN	OD1	3.07	SS	-	4.90	156.0	2.13	154.3	161.9
15	A0174-ASN	ND2	D0173-ASN	OD1	3.13	SS	-	8.77	150.1	2.22	120.7	119.4
16	B0174-ASN	ND2	A0174-ASN	OD1	2.98	SS	-	4.90	161.9	2.01	162.2	168.1
17	A0175-PHE	N	B0174-ASN	O	2.93	MM	-	6.16	158.8	2.11	158.6	164.7
18	B0176-VAL	N	A0175-PHE	O	2.87	MM	-	6.08	170.4	2.02	174.9	177.7
19	C0169-TYR	O4	B0171-ASN	O	2.95	SM	-	9.90	142.5	2.23	119.8	128.8
20	A0171-ASN	N	A0171-ASN	O	2.80	MM	0	0	103.8	2.50	76.9	59.7
21	A0172-GLN	N	A0172-GLN	O	2.80	MM	0	0	107.2	2.42	77.2	60.6
22	A0173-ASN	N	A0173-ASN	O	2.81	MM	0	0	103.3	2.48	78.6	61.3

^a Donor and acceptor names are abbreviated as the chain name followed by a four-digit residue number, a dash and the residue three letter code. A four-chain assembly (A-D) was used that encompass all unique hydrogen bonds in proto-PrP^{Sc}; two sheets, with two strands each. Chains A and B make up one sheet, C and D make up the mating sheet.

^b M represents main chain, S represents side chain.

All distances are measured in ångstroms; angles in degrees.

Supplementary Table 3. Crystallographic structure determination of carbamazepine.

Carbamazepine	MicroED
Data collection	
Space group	P 1 21/m 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	7.50, 11.00, 13.70
α , β , γ (°)	90.0, 93.2, 90.0
Resolution (Å)	7.49-0.849 (0.95-0.849)
<i>R</i> _{merge}	0.084 (0.149)
<i>R</i> _{meas}	0.104 (0.182)
<i>R</i> _{pim}	0.059 (0.103)
<i>I</i> / σ (<i>I</i>)	8.19 (4.55)
<i>CC</i> _{1/2}	0.992 (0.961)
Completeness (%)	49.8 (23.3)
No. Reflections	2739 (371)
Redundancy	2.71 (2.01)
Refinement	
No. reflections	2739 (371)
<i>R</i> _{work}	0.20
<i>R</i> _{free}	0.46
No. atoms	
Total	30
Waters	0
Hydrogens Possible	12
Hydrogens Modeled	12

^a Values in parentheses are for the highest-resolution shell.

Supplementary Table 4. Refinement of proto-PrP^{Sc} as a function of resolution.

proto-PrP ^{Sc}	MicroED 0.75Å	MicroED 0.80Å	MicroED 0.90Å	MicroED 1.00Å	MicroED 1.10Å
Data collection					
Space group	P1	P1	P1	P1	P1
Cell dimensions					
<i>a, b, c</i> (Å)	4.94 10.34 31.15	4.94 10.34 31.15	4.94 10.34 31.15	4.94 10.34 31.15	4.94 10.34 31.15
<i>α, β, γ</i> (°)	94.21 92.38	94.26 92.38	94.26 92.38	94.26 92.38	94.26 92.38
	102.20	102.20	102.20	102.20	102.20
Resolution (Å) ^a	10.34-0.75 (0.78-0.75)	10.34-0.80 (0.83-0.80)	10.34-0.90 (0.93-0.90)	10.34-1.0 (1.04-1.0)	10.34-1.10 (1.14-1.10)
<i>R</i> _{merge}	0.232 (0.638)	0.227 (0.524)	0.221 (0.373)	0.217 (0.335)	0.214 (0.265)
<i>R</i> _{meas}	0.250 (0.729)	0.245 (0.585)	0.238 (0.409)	0.233 (0.364)	0.229 (0.285)
<i>R</i> _{pin}	0.090 (0.319)	0.0872 (0.247)	0.0839 (0.159)	0.0823 (0.139)	0.0801 (0.0992)
<i>I</i> / <i>σ</i> (<i>I</i>)	4.57 (1.77)	5.05 (2.52)	5.82 (3.64)	6.30 (4.24)	6.71 (5.65)
<i>CC</i> _{1/2}	0.982 (0.209)	0.981 (0.549)	0.981 (0.847)	0.98 (0.92)	0.979 (0.938)
Completeness (%)	96.95 (96.17)	97.36 (95.54)	97.74 (95.48)	98.08 (97.40)	97.84 (99.24)
No. Reflections	43252 (2359)	36981 (3191)	27687 (2542)	20376 (1848)	15910 (1780)
Redundancy	5.8 (4.4)	6.0 (5.1)	6.3 (5.8)	6.4 (5.9)	6.6 (6.8)
Ab initio phasing					
Atoms Placed ^b	62.4 ± 0.70	67.9 ± 1.37	66.5 ± 0.97	66.8 ± 0.42	66.9 ± 1.52
CFOM	97.0 ± 0.41	98.9 ± 0.41	102.0 ± 0.36	106.1 ± 0.56	99.0 ± 0.89
Refinement					
No. reflections	7473 (720)	6160 (623)	4377 (443)	3166 (314)	2408 (258)
<i>R</i> _{work}	0.243 (0.36)	0.233 (0.35)	0.218 (0.29)	0.205 (0.24)	0.197 (0.23)
<i>R</i> _{free}	0.252 (0.39)	0.247 (0.38)	0.222 (0.31)	0.217 (0.26)	0.215 (0.26)
No. atoms					
Total	152	152	152	152	152
Hydrogens Possible	69	69	69	69	69
Hydrogens Modeled ^c	51	40	40	38	33
<i>B</i> factors					
Protein	6.00	6.51	6.01	5.79	5.45
Water	11.59	16.24	15.12	13.30	13.25
R.m.s. deviations					
Bond lengths (Å)	0.014	0.013	0.013	0.012	0.012
Bond angles (°)	1.01	1.00	1.06	1.06	1.08

^a Values in parentheses are for the highest resolution shell^b Average number of atoms placed by ten independent runs of 10,000 placement tries in SHELX.^c Hydrogens modeled include only those for which density is observed in 2Fo-Fc maps at 0.70σ or higher.

Chapter 2

Structural and biophysical consequences of sequence variation in the $\beta 2\alpha 2$ loop of mammalian prions

Structural and biophysical consequences of sequence variation in the $\beta 2\alpha 2$ loop of mammalian prions

Calina Glynn^{1,*}, Evelyn Hernandez¹, Marcus Gallagher-Jones^{1,2}, Jennifer Miao¹, Jose A. Rodriguez^{1,*}

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

² Present address Rosalind Franklin Institute, Harwell Campus, Harwell, UK

*Correspondence to jrodriguez@mbi.ucla.edu and cglynn@ucla.edu

Abstract

Despite high sequence similarity across mammalian prions, species barriers limit transmission of prion disease between certain pairs of species. Species barriers have been linked to sequence differences in PrP across mammals. In the $\beta 2\alpha 2$ loop of PrP, single residue substitutions at key positions 168, 170, and 174 have been shown to alter interspecies disease transmission and disease susceptibility altogether, yet the biophysical basis for this pattern remains unknown. To better understand this mystery, we interrogate the structural and biophysical consequences of natural and artificial sequence variation in the $\beta 2\alpha 2$ loop of PrP. Using microcrystal electron diffraction, we determine atomic resolution structures of segments encompassing residues 168-176 from the $\beta 2\alpha 2$ loop of PrP with sequences corresponding to human, mouse/cow, bank vole, rabbit/pig and naked mole rat (elk-T174S) $\beta 2\alpha 2$ loops. We also present crystal structures from synthetic $\beta 2\alpha 2$ loop sequences aimed at understanding the contribution of individual amino acids to $\beta 2\alpha 2$ amyloid packing and stability. This collection of structures presents two dominant amyloid

packing polymorphisms – one with an emphasis on interdigitation of sidechains that would be more discriminating in which sequences could serve as a mated strand, and one that is stabilized primarily by polar clasps between residues on the same strand, creating a smooth surface with fewer requirements for the sequence of a mated strand. We find that this distinct packing and array of stabilizing hydrogen bond networks depends most critically on the identity of residue 174. Each of these structures display common and species-specific denaturant resistant behaviors that match patterns observed for full length prion fibrils derived from corresponding species. Incorporation of these $\beta 2\alpha 2$ loop sequences into an 85 residue recombinant segment encoding wild-type bank vole PrP⁹⁴⁻¹⁷⁸ demonstrates that even single residue substitutions can impact fibril morphology as evaluated by negative stain electron microscopy. Our studies support the hypothesis that a diverse array of structures may be accessible to prion proteins and give rise to sequence-specific polymorph preferences influenced by key residues in the $\beta 2\alpha 2$ loop. It is likewise possible that incompatible structural polymorphisms may be responsible for limiting transmissibility across species lines.

Introduction

Discoveries supporting the existence of strains in prion disease were reported as early as 1961¹. In other early studies, it was noted that the inoculum strain and its interaction with the host animal played a role in whether, and how, the resulting prion disease would propagate^{2,3}. Studies of scrapie in mouse models also demonstrated that while the incubation period after inoculating mice was exceptionally long after the first passage, incubation times shortened for subsequent passages. However, in some attempts, mice failed to develop disease at all for unknown reasons³. This effect, whereby the incubation period of prions isolated from a different species is much longer than the incubation period of prions isolated from the same species, was termed the “species barrier”.

Over the course of the past half century, it has been established that each pair of animals, prion strain used, and direction of passage (e.g. mouse to rat vs. rat to mouse) has its own unique species barrier to be characterized³.

In the late 1980s and early 1990s, transgenic animals began to be used to study species barriers in prion diseases. This breakthrough in prion research established that species barriers could be eliminated if the animal had a prion protein (PrP) amino acid sequence matching the inoculum PrP sequence^{4,5}. As such, the aggregated form of the prion protein seen in diseased animals (PrP^{Sc}) was established as playing a major role in controlling disease in a sequence-dependent manner and was not simply a byproduct of infection.

While PrP sequence is highly conserved, residues at several positions along its sequence do vary across mammals. Two regions of relatively high variability are far apart in sequence (residues 164-174 and 215-223) but are near each other in the structure of the native, cellular form of PrP (PrP^C)⁶. Structural studies of PrP^C from various mammals likewise illustrated that the native structure of PrP^C is highly conserved across species, with the only noteworthy distinction being the degree of flexibility seen a loop formed by residues 165-175⁷, herein referred to as the β 2 α 2 loop. In this same study, it was demonstrated that the structural rigidity of the β 2 α 2 loop in PrP^C could be controlled by single amino acid substitutions at two positions within this region – 170 and 174 – without modification to any other residues in the protein.

Transgenic mice with “rigid loop” substitutions S170N and N174T developed prion disease with modest PrP overexpression, below concentrations required to induce disease for wild type mice⁸. A species barrier limited transmission of isolates from these transgenic mice when injected into wild type, “flexible loop” mice, implying that substitution of residues 170 and 174 alone were

sufficient to induce a species barrier. That species barrier is reduced when host and inoculum sequences matched at residues 170 and 174, despite sequence variation at other positions⁹.

From 1967-1979, another disease of clinical and pathological similarity to the spongiform encephalopathies was observed in captive mule deer¹⁰ followed by a similar disease observed in captive Rocky Mountain elk in the same facilities in Colorado¹¹ that was named Chronic Wasting Disease (CWD). In 1999, a strikingly high rate of horizontal transmission of CWD compared to other prion diseases was also noted¹². A molecular explanation for what exactly makes the rate of CWD transmission between cervids so much higher than the rate of transmission between animals with other prion diseases remains a mystery to this day, but the key may lie in the unique rigid loop sequence of the $\beta 2\alpha 2$ loop found in cervid prions.

After the discovery of the influence of sequence on the structural rigidity of the $\beta 2\alpha 2$ loop in PrP^C, the impact of sequence on the structure of the amyloid conformation of the $\beta 2\alpha 2$ loop in PrP^{Sc} became of interest. Segments as short as six residues from the $\beta 2\alpha 2$ loop form amyloids¹³⁻¹⁵, and the structures of human (¹⁷⁰SNQNNF¹⁷⁵)¹³ and elk (¹⁷⁰NNQNTF¹⁷⁵)¹⁴ $\beta 2\alpha 2$ loops pack into their amyloid forms in distinct ways. Likewise, a recently determined microcrystal electron diffraction (MicroED) structure of a nine residue segment of the bank vole $\beta 2\alpha 2$ loop forms a compact amyloid structure¹⁶. Compared to the previously determined 6-mer structures, the latter includes residues 168 and 169, which influence disease transmission and protease resistance¹⁷⁻¹⁹.

In this work, we structurally and biophysically characterize a collection of crystalline amyloid aggregates formed by nine-residue segments of the mammalian prion $\beta 2\alpha 2$ loop spanning residues 168-176 (human numbering) representing species with varying susceptibility to prion disease and differing transmission barriers (Figure 2.1 and 2.2). We found that the identity of residue 174 dictates whether these structures form one of two dominant amyloid packing

polymorphisms, a first held together by polar clasps within the same chain, and another stabilized by interdigitating sidechains reaching across to a mated strand. These packing preferences may have an influence on which regions of PrP are able to serve as a complementary strand packed against the $\beta 2\alpha 2$ loop of each mammalian prion.

Results

MicroED Structures of Amyloids Formed by the $\beta 2\alpha 2$ Loop of Naturally Occurring Mammalian Prions

To better understand the amyloid features adopted by segments of the $\beta 2\alpha 2$ loop of mammalian prions, we set out to determine atomic resolution crystal structures from eleven segments analogous to a previously studied bank vole $\beta 2\alpha 2$ loop structure¹⁶ using microcrystal electron diffraction (microED) (Figure 2.2). These segments consisted of $\beta 2\alpha 2$ loop sequences found in human, mouse/cow, pig/rabbit, and naked mole rat (elk T174S) PrP.

The structures formed can be classified into two structural categories, where peptides derived from bank vole, mouse, and human sequences fall into one category and those derived from elk and naked mole rat fall into a second. These categories arise from distinct, incompatible packing and hydrogen bond network arrangements governed by the identity of residue 174 but not by variation in the identity of residues at other positions within the loop.

The structures falling into category one are derived from amyloid conformations of mouse, human, and bank vole $\beta 2\alpha 2$ loops. These structures are very similar with the only differences being more about relative stability of the structures than the structures themselves (Figure 2.3). All three structures share an architecture rich in unusually complex hydrogen bonding networks, but also

harbor distinct features that may manifest as transmission barriers (Figure 2.4). The charged residue at 168 in the human structure restricts polar ladder – and crystal – formation at pH greater than 4.5, a restriction not associated with crystals from any other naturally occurring segments, including mouse loop crystals. This data, combined with the inability of residue E168 and S170 to form a hydrogen bond at physiological pH in the human structure, suggests that aggregates of human PrP segments are more likely to form in an acidic chemical environment. This behavior is mirrored in full length recombinant human PrP fibrils, where low pH is required for fibrillization²⁰, a restriction not associated with PrP fibrils from other species. With all other hydrogen bond networks in the crystal structures remaining conserved, the bank vole segment is able to form 4 additional hydrogen bonds between residues 168 and 170 in a three-dimensional polar clasp arrangement satisfied by other residues in the same peptide or extending into the peptide layers above and below. The mouse segment is only able to make a single hydrogen bond in the form of a polar ladder for residue 168 only, and the human segment is likewise only able to form a single hydrogen bond in the form of a polar ladder, and only when the surrounding environment is sufficiently acidic.

In the second packing category, hydrogen bond networks formed by N174 are disrupted by replacement with T or S in elk and naked mole rat structures respectively. Neither T174 nor S174 can form a polar clasp with residue Q172 as in packing category 1, and a new network between 170 and 172 is formed that influences the hydrogen bonds formed between side chains throughout the rest of the peptide (Figure 2.5). For the naked mole rat structure, Q168 does not participate in a polar clasp like its packing category one counterparts. Q168 instead forms a polar ladder stabilized by interdigitation with Q172 on an opposing chain. This type of stabilizing interaction – interdigitation between sidechains on mated sheets as opposed to interactions between residues on the chains above and below without interdigitation – is a feature of packing category 2 but not packing category 1.

Rabbits are relatively resistant to cross-species prion disease transmission^{21,22} for reasons that are not completely understood. The rabbit $\beta 2\alpha 2$ loop harbors both S170 and S174, an unusual combination that gives rise to a rigid loop in PrP^C²³. Here we aimed to investigate the structure taken on by the rabbit $\beta 2\alpha 2$ loop in search of structural incompatibilities that may hamper disease transmission. Like the naked mole rat $\beta 2\alpha 2$ loop structure, the rabbit structure displays two unique steric zipper interfaces and interdigitation of polar ladders that are a hallmark of packing category 2 (Figure 2.5). The interface harboring interactions between even numbered residues closely resembles what was seen in the naked mole rat structure. The second interface, which has been conserved across all other 9 residue $\beta 2\alpha 2$ loop structures regardless of which category they belong to, is completely unique to the rabbit structure. N171 and N173 still form polar clasps, but instead of forming between residues on the same strand, each residue reaches across the interface to pair with a residue on the mating strand. This also forces the aromatic residues, which hug inward in all other structures, to rotate and face outward.

MicroED Structures from $\beta 2\alpha 2$ Loop Peptides with Sequence Variations of Known Biological Consequence

To assess the contribution of individual residues to structure and stability of $\beta 2\alpha 2$ loop aggregates, structural and biophysical characterization of several synthetic segments was pursued. The S170N mutation is one of two residues involved in forming the “rigid loop” seen in some mammalian PrP^C structures and has been shown to influence transmission barriers^{9,18}. Mutations to Y169 have been shown to have a dramatic effect on transmission and protease-resistance of the resulting prion aggregates^{17,19}. A Y169G mutation completely hinders disease transmission¹⁷ while a Y169F mutation - the removal of a single oxygen atom - has no impact on infectivity, but results in increased protease sensitivity¹⁹. Here, the impact of these mutations on structure and

stability was assessed for three 9 residue peptides: human S170N and Y169F, and bank vole Y169F. The structure taken on by a 7 residue variant with the Y169G mutation (¹⁶⁹GSNQNNF¹⁷⁵) has been explored in other work^{24,25} and a structure taken on by a synthetic bank vole Q172E mutant - ¹⁶⁸QYNNENN¹⁷⁶FV (Figure 2.2) - will be further described in a future manuscript by Richards and Flores *et al.* Interestingly, none of these substitutions resulted in substantial structural rearrangement. Thus, whatever effects these mutations have that alter protease-resistance and disease transmission could not be captured in these static end-point peptide crystal structures.

Several other nine residue $\beta 2\alpha 2$ loop segments of interest, including those representing chimp (QYSSQNNFV), elk (QYNNQNTFV), and the disease preventing sheep Q168R sequence polymorphism (RYSNQNNFV), were pursued but crystallization was not achieved. For these samples, fiber diffraction was used to assess the amyloid character of the resulting aggregates. All aggregates, regardless of whether they were crystalline or fibrillar in structure, displayed classic features of amyloid steric zippers (Figure 2.7). These included rings in the fiber diffraction patterns at approximately 4.7-4.8Å and 10Å indicative of the interstrand and intersheet spacing respectively.

Influence of $\beta 2\alpha 2$ Loop on Fibril Morphology

To assess the influence of the $\beta 2\alpha 2$ loop on fibril morphology, $\beta 2\alpha 2$ loop substitutions to match bank vole, elk, naked mole rat, human, and mouse / cow sequences were made against a bank vole PrP⁹⁴⁻¹⁷⁸ background. This span of residues was chosen to include the $\beta 2\alpha 2$ loop, all residues previously believed to be part of a beta solenoid prion core²⁶ and all residues in the ordered core of fibrils formed by a Y145Stop truncation capable of inducing disease in mice²⁷ that displayed species-specific ssNMR spectra²⁸. Since all fibrils were formed under identical growth conditions, differences in morphology are not a result of discrepancies in chemical environment and can only

be due to differences imparted by the protein itself. Each sample displayed a distinguishable morphology, with the bank vole loop containing fibrils forming rapidly twisting fibrils, mouse loop containing fibrils forming very short thick fibrils, human loop forming long matted fibrils, elk loop forming short fibrils of mixed thickness, and naked mole rat loop forming the longest rod-shaped filaments (Figure 2.8). This illustrates that changes to the $\beta 2\alpha 2$ loop alone are able to influence overall fibril morphology.

Discussion

Species Barriers: A Product of Variation in PrP^C Structure, Transition States, or End-Point Fibril Structure?

All prion diseases must begin with a conversion of the natively folded PrP^C to the disease associated PrP^{Sc} structure. While there is growing evidence that the end-point fibril structure of PrP^{Sc} differs between mammalian species and strains, the starting structure of PrP^C is highly conserved. The main difference here lies in the rigidity of the $\beta 2\alpha 2$ loop, which has been proposed as a possible explanation for species barriers⁷. The rigid loop hypothesis added to the collective understanding of the conversion path from PrP^C to PrP^{Sc}, where sequence differences give rise to starting points of PrP^C with sequence-dependent susceptibility to misfolding and structural conversion. The work presented here adds to our understanding of the possible endpoint structures reached by different mammalian prions, where two types of structures – one reliant on stabilizing interactions within the same chain and a second reliant on interdigitation of polar residues across an interface. We also presented how these structures may be favored in different chemical environments with different stabilities from one another, meaning the same PrP^{Sc} structures may not be strictly precluded from coexisting in different species, but unfavorable to form in the same environment.

Amyloid structures adopted by $\beta 2\alpha 2$ loop segments

Based on this collection of crystal structures, we have identified two modes of stabilization for amyloid conformations of mammalian prion $\beta 2\alpha 2$ loops. The first (packing category 1) relies completely on interactions between residues on one side of the interface. All hydrogen bonds are satisfied by other residues in the same strand or by layers above and below, creating fewer restrictions on hydrogen bonding capabilities of the sequence of a mated strand. The surface displayed is also smooth (Figure 2.5) and could serve as a complement to many other sequences, not just those with large residues capable of interdigitation. Thus, the interface between this strand and another would not be specific and many different segments within the prion protein could serve as mates for this sheet.

The second mode of stabilization (packing category 2) harbors interactions that more closely resemble those seen in other amyloid peptide structures where polar residues on one strand interdigitate with residues on a mated strand. This new surface for interaction is littered with clefts sized to fit specific residues with specific spacings, and thus this interface would be more discriminating in which sequences could serve as a mated strand than the surface seen in packing category 1.

We have also determined that residue 174 governs the preferred types of stabilizing interactions – interdigitation versus polar clasps - and hydrogen bonds formed by other residues on the same side of the sheet in a way that other polar residues sitting on the same side do not. E versus Q at residue 168, S versus N at residue 170, and Q versus E at residue 172 were not able to influence the hydrogen bonding networks throughout the peptide. In contrast, having an N, as opposed to a T or S, at residue 174 did. Mutations S170N, Y169G, and Y169F also did not have an impact on which packing category a structure fell into, and thus their influence on transmission and

protease-resistance could not be explained by a structural change in the resulting aggregates in this work. For Y169F, this is in agreement with previous work showing that no strain shift occurs as a result of the mutation¹⁹.

For $\beta 2\alpha 2$ loop structures falling into the first packing category (human, mouse, and bank vole), hydrogen bonds were mostly satisfied by interactions between side chains belonging to the same strand with minimal interdigitation or cross-strand interaction. For structures falling into the second packing category (naked mole rat and elk 6mer), two unique steric zipper interfaces were observed, with one identical to the single interface seen in the first packing category, and a second, new interface with slightly more interdigitation resulting from a T or S at residue 174. The rabbit structure more closely resembles the second packing category, due to a similar interface 2 resulting from an S at residue 174. However, the rabbit structure has a completely unique interface that isn't seen in any other $\beta 2\alpha 2$ loop structure presented here. This interface is uniquely formed through polar clasps formed between mated strands. This preferred mode of aggregate stabilization is incompatible with all other $\beta 2\alpha 2$ loop structures, and represents how different preferred hydrogen bond networks could translate into distinct protofilament arrangements and structural incompatibilities that hinder disease transmission.

Stability of $\beta 2\alpha 2$ loop aggregates

Crystals formed by wild type and a Y169F mutant of the bank vole prion $\beta 2\alpha 2$ loop were largely insensitive to pH and chaotropic agents (Figure 2.3). The Y169F mutation has been shown to increase proteinase K sensitivity and decrease overall density of prion aggregates without creating a strain shift¹⁹. This mutation resulted in no changes to the structure taken on by $\beta 2\alpha 2$ loop aggregates studied here, including no changes to the number of ordered water molecules observed in the structure. Likewise, the crystalline aggregates of both the wild type bank vole

$\beta 2\alpha 2$ loop and the Y169F mutant were relatively insensitive to changes in pH and chaotropic agent concentrations compared to other crystalline $\beta 2\alpha 2$ loop sequences. This mutation leaves all polar ladders - including the three polar clasps - in the structure intact. The location of the oxygen atom removed by the change from tyrosine to phenylalanine is buried in both structures - past tightly packed backbones spaced 4.7Å apart and past stable polar clasps held together by 4 interlocking hydrogen bonds per layer and an aromatic stack. When present, the oxygen atom only makes a single weak hydrogen bond to the backbone of the mated beta sheet and its removal is inconsequential for these crystalline aggregates. In a recently determined structure of an infectious prion from the 263K scrapie strain²⁹, Y169 is solvent facing. Thus, removal of an oxygen atom here may have an influence on prion fibril structure and interaction partners that cannot be captured in the peptides studied here. Another possibility is that this residue is critical for a transition state, which could neither be assessed here nor by any end-state fibril structure.

Crystals formed by the wild type mouse/cow $\beta 2\alpha 2$ loop dissociate on exposure to basic environments and are somewhat more sensitive to chaotropic agents than wild type and Y169F mutant bank vole counterparts. Conformations of all residues in the mouse/cow peptide structure were essentially identical to those in the bank vole peptide structure with the exception of residue 170, which is an asparagine in the bank vole sequence and a serine in the mouse/cow sequence. The serine residue is unable to form a polar ladder due to its small size and lack of a nitrogen atom, and thus a polar clasp also cannot form between residues 168 and 170. This reduces the strength of the three-dimensional polar clasp consisting of four hydrogen bonds per pair of stacked sheets, to the single linear hydrogen bond formed along the Q168 polar ladder. This small change results in a much less stable structure and greater sensitivity to aggregate denaturation by chaotropic agents guanidine and urea.

As might be anticipated, the weakest aggregates harbored a negatively charged residue at position 168, which prevents the formation of aggregates at pH above the pKa of glutamate (~4.5) and readily dissolves existing aggregates. This sensitivity to basic environments cannot be mitigated by introduction of an asparagine at position 170 in place of serine, which allows for formation of a polar ladder regardless of pH, and formation of a full polar clasp at pHs below the pKa of glutamate. The human $\beta 2\alpha 2$ loop crystalline aggregates as well as loop mutants Y169F and S170N - which leave E168 unchanged - were also more sensitive to denaturation by urea and guanidine than all $\beta 2\alpha 2$ loop aggregates with a glutamine at residue 168 except for rabbit/pig $\beta 2\alpha 2$ loop aggregates. This can be explained by the distinct packing arrangement seen in the rabbit/pig $\beta 2\alpha 2$ loop crystals. This peptide crystallized as a class 1 (face-to-face)¹³ steric zipper with two unique interfaces between mated sheets, whereas all other peptides investigated in stability experiments crystallized as class 2 (face-to-back) steric zippers with a single unique interface between mated sheets. The absence of steric zippers and polar clasps on the outer edge of human (including mutants that do not mutate E168) and rabbit/pig $\beta 2\alpha 2$ loop crystals are a likely explanation for their increased sensitivity to chaotropic agents compared to Q168 containing $\beta 2\alpha 2$ loop crystalline aggregates that pack as class 2 steric zippers.

Lastly, these changes in peptide structure and environment response translated to changes in fibril morphology when loop mutations were incorporated into a bank vole prion background. Each construct displayed malleable morphological preference when subjected to different chemical growth environments (data not shown), illustrating the known influence of environment on fibril morphology formed. When growth conditions were held constant, fibrils containing different $\beta 2\alpha 2$ loop sequences – just 3 residue changes maximum – gave rise to readily distinguishable polymorphs that were sequence dependent. This illustrates that just a few residues in the $\beta 2\alpha 2$ loop are sufficient to influence fibril structure, at least to some degree.

Materials and Methods

Crystallization

All peptides were synthesized and purified by HPLC to 98% purity by Genscript. All peptides were dissolved in water at the concentration listed in Table 1 with the exception of naked mole rat and rabbit / pig 168-176, which were dissolved in 1% and 4% DMSO respectively. Peptides were then either added directly to 96-well hanging-drop broad screen crystallization trays or filtered using a 0.22 micron filter according to Table 1. Some conditions produced crystals that could be readily identified by light microscopy while others produced aggregates that could be either crystalline or fibrillar. In the absence of clear crystalline hits, conditions with potential crystals were screened by transmission electron microscopy (TEM). Aggregates with crystalline features were also screened for diffraction quality by TEM. Initial hits containing either crystals or near-crystalline aggregates were then optimized via 24 well hanging-drop vapor diffusion experiments until the final crystallization conditions that yielded the largest monomorphic crystals with the best diffraction patterns observed by TEM were obtained (Table 1). It is worth noting that optimizing crystallization conditions for many of these peptides was not trivial. Subtle changes in peptide preparation and growth conditions impacted crystal morphology and led to incremental improvements in diffraction quality. For some peptides, many rounds of tedious optimization were needed to reach the final condition used for structure determination.

Expression and Purification BvPrP⁹⁴⁻¹⁷⁸ with β 2 α 2 loop substitutions

As will be described in chapter 3 on rHuPrP⁹⁴⁻¹⁷⁸ ³⁰ and in chapter 4 on rBvPrP⁹⁴⁻¹⁷⁸, a gene encoding the equivalent bank vole PrP residues, 94-178, with M109 genotype was purchased from IDT. Four additional constructs, which were identical in sequence except for β 2 α 2 loop

substitutions to match elk, naked mole rat, human, and mouse $\beta 2\alpha 2$ loops, were also purchased. Each construct was cloned into a pET24a+ derivative vector lacking purification or solubility tags using Gibson assembly cloning between NcoI and XhoI restriction sites. The sequence of each construct was verified by DNA sequencing (Genewiz). A summary of all protein constructs used in this chapter and subsequent chapters is included in appendix Table 1. Each construct was transformed into BL21 GOLD (DE3) cells and grown to an optical density at 600nm (OD_{600}) of 0.6-0.8 at 37 degrees C while shaking at 180-230rpm. Overexpression was induced by addition of IPTG to 1mM and cells were allowed to continue shaking under the same condition for 4-6 hours before being harvested via centrifugation at 8,000g for 10 minutes. Cell pellets were stored at -80 degrees C until subsequent purification steps were performed.

Purification was carried out as previously described^{30,31} and reiterated here. Cell pellets were resuspended in 25 mM Tris-HCl pH 8 and 5 mM EDTA (Buffer A) with HALT protease-inhibitor cocktail (Sigma Aldrich). Cells were then lysed using an EmulsiFlex-C3 High Pressure Homogenizer (Avestin) and pelleted via centrifugation at 30,000g for 1 hr (rBvPrP⁹⁴⁻¹⁷⁸ with bank vole, elk, naked mole rat, and human loops) or at 10,000g for 40 minutes (rBvPrP⁹⁴⁻¹⁷⁸ with mouse / cow loop) at 4 degrees C. The supernatant was removed before the pellet was resuspended again in Buffer A and centrifuged under the same conditions to remove any remaining soluble material. This pellet was either immediately processed or stored at -80 degrees C until purification continued the next day. The pellet was next solubilized in freshly made 8 M Guanidine-HCl, 25 mM Tris-HCl pH 8 and 100 mM DTT (Buffer B) before centrifugation for 20 minutes at 20,000g (rBvPrP⁹⁴⁻¹⁷⁸ with bank vole, elk, and naked mole rat loops) or 35,000g (rBvPrP⁹⁴⁻¹⁷⁸ with human and mouse / cow loops) based on how readily soluble and insoluble material separated. The supernatant containing solubilized inclusion bodies was filtered using a 0.45 micron filter. 2 μ l per injection of filtered sample was injected into an NGC chromatography system (Biorad) and flowed

over an ENrich SEC 650 10 × 300 column (Biorad) equilibrated with freshly made 6 M Guanidine-HCl, 12.5 mM Tris-HCl pH 8, 5 mM DTT and 1 mM EDTA (Buffer C).

All proteins eluted partially in the void volume, indicative of multimers or incomplete solubilization, but a majority eluted as a monomer. Monomeric prion protein-containing fractions were pooled and buffer exchanged into 8M urea using a Duoflow chromatography system (Biorad) and a HiTrap Desalting column (GE Healthcare) either the same day or the next day following flash freezing and storage at -80 degrees C. After desalting, protein containing fractions were pooled and concentrated to 3.2 - 8 mg/ml before being flash frozen using liquid nitrogen and stored at -80 degrees C until use.

Fibrillization of rBvPrP⁹⁴⁻¹⁷⁸ with β 2 α 2 Loop Substitutions

In order to control for differences in fibril morphology induced by differences in growth conditions, all proteins were fibrillized under the same conditions. All fibrils were formed at 1mg/ml in 1M urea, 200mM NaCl, and 50mM NaCitrate pH 4 via acoustic resonance mixing at 38Hz for a few days.

Transmission electron microscopy

Crystalline samples were prepared for Transmission electron microscopy (TEM) screening for crystal and diffraction quality as previously described¹⁶. In brief, 2-3 μ L drops from 24 well crystallization screens were either pipetted and applied directly to a 300 mesh formvar carbon (F/C) grid (Ted Pella) and allowed to incubate for approximately 2 minutes before excess liquid was wicked away with filter paper. For drops where crystals stuck to the cover slip or aggregates were difficult to pipet, an additional 2-3 μ L of well solution was added to the drop to aid in sample

removal before application to a grid. Grids were screened on a Thermo Fisher Tecnai 12 microscope for crystal presence and quality before being screened for dry diffraction quality on a Thermo Fisher Tecnai F20. Samples with the highest quality diffraction were used for microED data collection. Fibrillar samples were prepared similarly and as previously described³⁰.

MicroED Sample Preparation

Each peptide crystal solution was pipetted to aid in creating a monodisperse solution for grid application. All holey carbon grids (Quantifoil R 2/4, 1/4, 2/1, or 2/2, 200 or 300 mesh Copper, Electron Microscopy Sciences) were glow discharged using a PELCO easiGlow and plunge frozen into liquid ethane using an FEI Vitrobot Mark IV set to 0% humidity after sample application. 1.5 μ L (EYNNQNNFV, QYSNQNSFV), 1.8 μ L (QYSNQNNFV, EFSNQNNFV, QFNNQNNFV), 2 μ L (YSNQNNF, EYSNQNNFV, QYNNENNFV, QYNNQNSFV), or 2.5 μ L (QYSNQNSFV) was applied to the carbon side only (QYSNQNNFV, EYNNQNNFV, EFSNQNNFV, QFNNQNNFV, QYSNQNNFV) or both sides (YSNQNNF, EYSNQNNFV, QYSNQNSFV, QYNNENNFV, QYNNQNSFV) of the grid. A blot force of 22 was used for all samples except EYNNQNNFV, YSNQNNF and EYSNQNNFV, which used a blot force of 20 (EYNNQNNFV) or 6-10 (YSNQNNF and EYSNQNNFV). Blot times of 15-20 (YSNQNNF, EYSNQNNFV), 18.5 (QYSNQNNFV, EFSNQNNFV, QFNNQNNFV), 20 (EYNNQNNFV), 22 (QYNNENNFV), 24 (QYSNQNSFV), 25 (QYNNQNSFV), or 26 (QYSNQNSFV) were used depending on whether peptide crystals diffracted better more wet or more dry.

MicroED Data Collection

For the peptides YSNQNNF, EYSNQNNFV, QYSNQNNFV, EYNNQNNFV, EFSNQNNFV, QFNNQNNFV, QYSNQNSFV, QYNNENNFV and GSNQNNF, diffraction patterns were collected

under cryogenic conditions using an FEI Tecnai F20 transmission electron microscope operating at 200keV equipped with a bottom mount TemCam-F416 CMOS camera (TVIPS). Diffraction patterns were collected at a detector distance of 730mm or 520mm (QYNNENNFV and GSNQNNF only) with 2 second, 3 second (GSNQNNF only), or 5 second (QYSNQNSFV and GSNQNNF only) exposures while continuously rotating at 0.3 degrees per second, 0.25 degrees per second (QYSNQNSFV and QYNNENNFV only), or 0.2 degrees per second (GSNQNNF only). For QYNNQNSFV, diffraction patterns were also collected under cryogenic conditions, but using a Thermo Fisher Talos Arctica operating at 200keV equipped with a Thermo Fisher CetaD CMOS detector. Diffraction patterns were collected at a detector distance of 750mm with 3 second exposures while the stage was continuously rotated at 0.3 degrees per second.

MicroED Data Processing

Diffraction images for all peptides were indexed and integrated in XDS³² with the best datasets contributing to the final solutions merged in XSCALE³². All datasets were of sufficient quality at high-resolution to yield unambiguous direct methods solutions and ranged in resolution from 0.85 - 1.05Å. Atomic models were built into Coulomb potential maps in Coot³³ and refined in either PHENIX³⁴ (EYSNQNNFV, QYSNQNNFV, EYNNQNNFV, QYSNQNSFV, QYNNQNSFV), REFMAC³⁵ (EFSNQNNFV, QYNNENNFV, QFNNQNNFV), or a combination of PHENIX, REFMAC, and Buster³⁶ (GSNQNNF, YSNQNNF). Refinement statistics for all structures are listed in Table 2.

Fibril diffraction of mammalian β 2 α 2 loop peptide crystals or fibrils

Crystalline (QYNNQNNFV, QYSNQNNFV, EYSNQNNFV, QYSNQNSFV, QYNNQNSFV, QFNNQNNFV, EFSNQNNFV) or fibrillar (QYNNQNTFV, RYSNQNNFV, QYSSQNNFV)

aggregates were centrifuged at low speed, had growth buffer removed via pipetting, and subjected to addition of a smaller volume of water repeatedly in order to remove excess salts that would compromise fibril diffraction patterns. The concentrated aggregate solution was then applied between two blunted capillaries and allowed to dry. Additional aggregate solution was applied repeatedly until a sufficiently sized aggregate bundle was achieved. All samples with the exception of QYSNQNSFV were allowed to dry before being diffracted with the cryo-stream pointed away from the sample using a 5 minute exposure to a FRE+ rotating anode generator with VARIMAX HR confocal optics producing Cu K- α radiation (Rigaku, Tokyo, Japan). Diffraction patterns were collected using a RIGAKU R-AXIS HTC imaging plate detector at a distance of 156 mm from the x-ray source. For QYSNQNSFV, small sample volumes prevented complete removal of all excess salts. To minimize salt rings in diffraction patterns, QYSNQNSFV aggregates were diffracted while wet with 50% glycerol applied to the aggregate immediately before diffraction with the cryo-stream directed at the sample.

Chemical Denaturation of Mammalian $\beta 2\alpha 2$ loop crystals

Each peptide was allowed to form crystalline aggregates according to the conditions given in Table 1. These solutions were treated similarly to the previously described bank vole $\beta 2\alpha 2$ loop crystals¹⁶. Using a high concentration of crystalline material, all crystals were diluted 4x into a final concentration solution of 0.5-4.5M guanidinium-HCl, 0.5-6M urea, 0.75M HCl, MES pH 2.5, NaAc pH 4.5, MES pH 6, Tris-HCl pH 8.5 or 10, MES pH 11.5, NaOH, or water. After mixing with buffers of varying pH, each solution was tested with pH paper to ensure the desired pH had been reached. For QYSNQNSFV crystals, the crystallization buffer reacted with low pH solutions, so crystals were centrifuged at low speed to pellet crystalline material. The supernatant containing the crystallization buffer was removed and replaced with an equal volume of water, which did not cause the crystals to dissolve. Absorbance was used as a proxy for aggregate content, and an

absorbance spectra from 250-700nm was recorded using a Nanodrop One (Thermo) within a few minutes of mixing crystals with each solution. Each measurement was performed in triplicate with the exception of QYSNQNNFV, which had one measurement recorded per condition. A solution consisting of 0.01% w/v latex spheres suspended in water was used as an absorbance control. Absorbance readings for each crystal solution were scaled by setting the untreated crystals to a relative optical density at 600nm of 1, thus readings shown in Figure 2.1 are absorbance level relative to untreated crystals of the same type.

References

1. Pattison, I. H. & Millson, G. C. Scrapie Produced Experimentally in Goats With Special Reference To the Clinical Syndrome. *Journal of Comparative Pathology and Therapeutics* **71**, 101-IN10 (1961).
2. Bruce, M. E., Dickinson, A. G. & Fraser, H. Cerebral Amyloidosis in Scrapie in the Mouse: Effect of Agent Strain and Mouse Genotype. *Neuropathology and Applied Neurobiology* **2**, 471–478 (1976).
3. Kimberlin, R. H., Cole, S. & Walker, C. A. Y. 1987. Temporary and Permanent Modifications to a Single Strain of Mouse Scrapie on Transmission to Rats and Hamsters. *Journal of General Virology* **68**, 1875–1881.
4. Scott, M. *et al.* Transgenic mice expressing hamster prion protein produce species-specific scrapie infectivity and amyloid plaques. *Cell* **59**, 847–857 (1989).
5. Prusiner, S. B. *et al.* Transgenic studies implicate interactions between homologous PrP isoforms in scrapie prion replication. *Cell* **63**, 673–686 (1990).
6. Billeter, M. *et al.* Prion protein NMR structure and species barrier for prion diseases. *Proceedings of the National Academy of Sciences* **94**, 7281–7285 (1997).
7. Gossert, A. D., Bonjour, S., Lysek, D. A., Fiorito, F. & Wuthrich, K. Prion protein NMR structures of elk and of mouse/elk hybrids. *Proceedings of the National Academy of Sciences* **102**, 646–650 (2005).
8. Sigurdson, C. J. *et al.* De novo generation of a transmissible spongiform encephalopathy by mouse transgenesis. *PNAS* **106**, 304–309 (2009).
9. Sigurdson, C. J. *et al.* A molecular switch controls interspecies prion disease transmission in mice. *J Clin Invest* **120**, 2590–2599 (2010).

10. WILLIAMS, E. S. & YOUNG, S. CHRONIC WASTING DISEASE OF CAPTIVE MULE DEER: A SPONGIFORM ENCEPHALOPATHY¹. *Journal of Wildlife Diseases* **16**, 89–98 (1980).
11. WILLIAMS, E. S. & YOUNG, S. SPONGIFORM ENCEPHALOPATHY OF ROCKY MOUNTAIN ELK¹. *Journal of Wildlife Diseases* **18**, 465–471 (1982).
12. Sigurdson, C. J. *et al.* Oral transmission and early lymphoid tropism of chronic wasting disease PrPres in mule deer fawns (*Odocoileus hemionus*). *Journal of General Virology* **80**, 2757–2764.
13. Sawaya, M. R. *et al.* Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453 (2007).
14. Wiltzius, J. J. W. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nat Struct Mol Biol* **16**, 973–978 (2009).
15. Rodriguez, J. A., Jiang, L. & Eisenberg, D. S. Toward the Atomic Structure of PrPSc. *Cold Spring Harb Perspect Biol* **9**, a031336 (2017).
16. Gallagher-Jones, M. *et al.* Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nature Structural & Molecular Biology* **25**, 131 (2018).
17. Kurt, T. D. *et al.* Prion Transmission Prevented by Modifying the β 2- α 2 Loop Structure of Host PrPC. *J. Neurosci.* **34**, 1022–1027 (2014).
18. Kurt, T. D. *et al.* Human prion protein sequence elements impede cross-species chronic wasting disease transmission. *The Journal of Clinical Investigation* **125**, 1485 (2015).
19. Leske, H. *et al.* Protease resistance of infectious prions is suppressed by removal of a single atom in the cellular prion protein. *PLOS ONE* **12**, e0170503 (2017).
20. Swietnicki, W., Morillas, M., Chen, S. G., Gambetti, P. & Surewicz, W. K. Aggregation and Fibrillization of the Recombinant Human Prion Protein huPrP90–231. *Biochemistry* **39**, 424–431 (2000).

21. Vorberg, I., Groschup, M. H., Pfaff, E. & Priola, S. A. Multiple Amino Acid Residues within the Rabbit Prion Protein Inhibit Formation of Its Abnormal Isoform. *J. Virol.* **77**, 2003–2009 (2003).
22. Eraña, H. *et al.* In Vitro Approach To Identify Key Amino Acids in Low Susceptibility of Rabbit Prion Protein to Misfolding. *Journal of Virology* **91**, e01543-17.
23. Wen, Y. *et al.* Unique Structural Characteristics of the Rabbit Prion Protein. *J Biol Chem* **285**, 31682–31693 (2010).
24. Martynowycz, M. W. *et al.* MicroED Structures from Micrometer Thick Protein Crystals. *bioRxiv* 152504 (2017) doi:10.1101/152504.
25. Hattne, J. *et al.* Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* **26**, 759-766.e4 (2018).
26. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. Evidence for assembly of prions with left-handed β -helices into trimers. *PNAS* **101**, 8342–8347 (2004).
27. Choi, J.-K. *et al.* Amyloid fibrils from the N-terminal prion protein fragment are infectious. *PNAS* **113**, 13851–13856 (2016).
28. Theint, T. *et al.* Species-dependent structural polymorphism of Y145Stop prion protein amyloid revealed by solid-state NMR spectroscopy. *Nature Communications* **8**, 1–10 (2017).
29. Kraus, A. *et al.* Structure of an infectious mammalian prion. *bioRxiv* 2021.02.14.431014 (2021) doi:10.1101/2021.02.14.431014.
30. Glynn, C. *et al.* Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. *Nature Structural & Molecular Biology* 1–7 (2020) doi:10.1038/s41594-020-0403-y.
31. Wan, W. *et al.* Structural Studies of Truncated Forms of the Prion Protein PrP. *Biophysical Journal* **108**, 1548–1554 (2015).
32. Kabsch, W. XDS. *Acta Cryst D* **66**, 125–132 (2010).

33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst D* **66**, 486–501 (2010).
34. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**, 213–221 (2010).
35. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D* **67**, 355–367 (2011).
36. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER–TNT. *Acta Cryst D* **60**, 2210–2221 (2004).
37. Zahn, R. *et al.* NMR solution structure of the human prion protein. *PNAS* **97**, 145–150 (2000).

Acknowledgements

We thank David S. Eisenberg, Duilio Cascio, Michael Sawaya, and Christina Sigurdson for helpful discussions. We thank Drs. Dan Shi and Tamir Gonen for technical assistance and facilitating TEM use. This work is supported by the HHMI Janelia Visiting Researcher Program and the STROBE National Science Foundation Science and Technology Center under grant number DMR1548924. C.G. was funded by the Ruth L. Kirschstein NRSA GM007185 (NIH T32 Cellular and Molecular Biology Training Grant at UCLA) and is now funded by the NIAID Ruth L. Kirschstein Predoctoral Individual NRSA, 1F31AI143368 and the Fowler Fellowship in Protein Science (UCLA). M.G.J. was supported by a QCB Collaboratory Postdoctoral Fellowship (UCLA) at the time the work was carried out. J.A.R. is supported by the NIGMS MIRA R35GM128867.

Data Availability

Atomic coordinates and structure factors for all microED structures have been deposited in wwPDB under accession codes 7RVC (EYSNQNNFV), 7RVD (QYSNQNNFV), 7RVE (EYNNQNNFV), 7RVF (QFNNQNNFV), 7RVG (QYSNQNSFV), 7RVH (QYNNENNFV), 7RVI (QYNNQNSFV), 7RVJ (YSNQNNF), 7RVK (GSNQNNF), and 7RVL (EFSNQNNFV). Other structures that are described in this work but previously published can be found under PDB accession codes 2OL9 (SNQNNF)¹³, 6AXZ (QYNNQNNFV)¹⁶, and 3FVA (NNQNTF)¹⁴. Source data for all figures and files is available from the authors upon reasonable request.

Contributions

J.A.R. directed the work. J.A.R., E.H., J.M., and C.G. grew, evaluated, and optimized crystals.

J.A.R., C.G., and E.H. collected data. J.A.R., C.G., E.H., M.G.J., and J.M. analyzed the data.

C.G. and J.A.R. wrote the article with input from all authors.

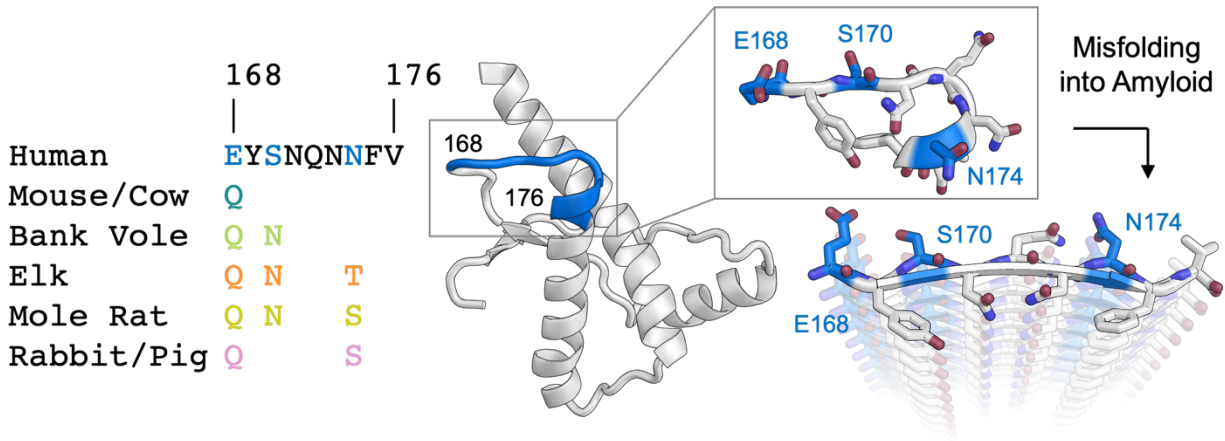


Figure 2.1. The $\beta 2\alpha 2$ loop misfolds into amyloid. The $\beta 2\alpha 2$ loop of mammalian prions (left) takes on a loop structure in PrP^C (center and top right) but misfolds into a beta sheet amyloid conformation (bottom right). Residues that vary between mammals (168, 170 and 174) are colored. Human PrP^C structure with PDB ID 1QLX³⁷ is shown.

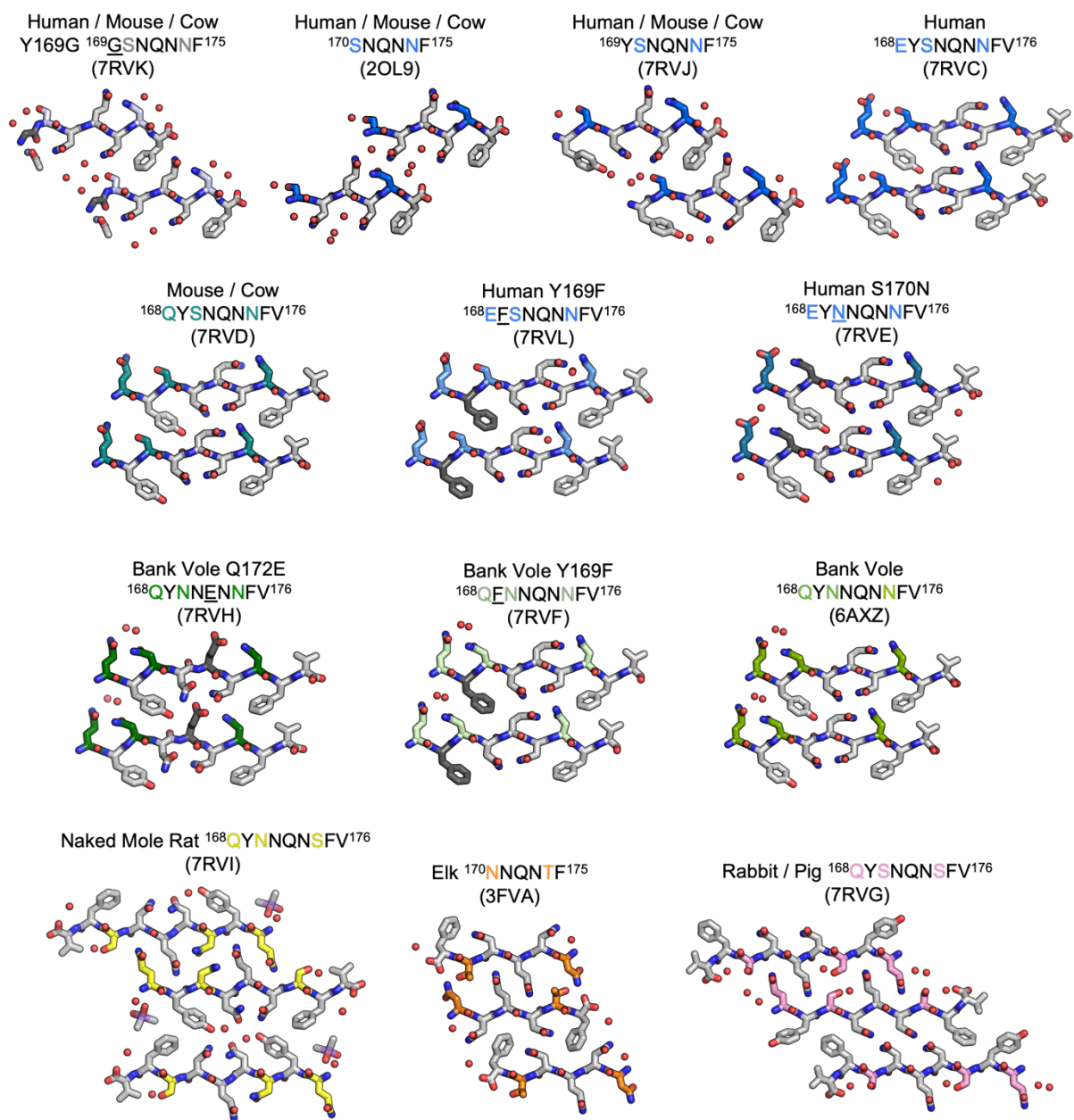


Figure 2.2. Structures of crystalline mammalian prion $\beta 2\alpha 2$ loop aggregates. All crystal structures resolved in this chapter are presented here. Additionally, three published structures from peptides SNQNNF¹³, NNQNTF¹⁴, and QYNNQNNFV¹⁶ are included for reference. All structures except for the bottom row are class 2 (face-to-back) steric zippers that fall into packing category 1. The bottom row is composed of class 1 (face-to-face) steric zippers that fall into packing category 2.

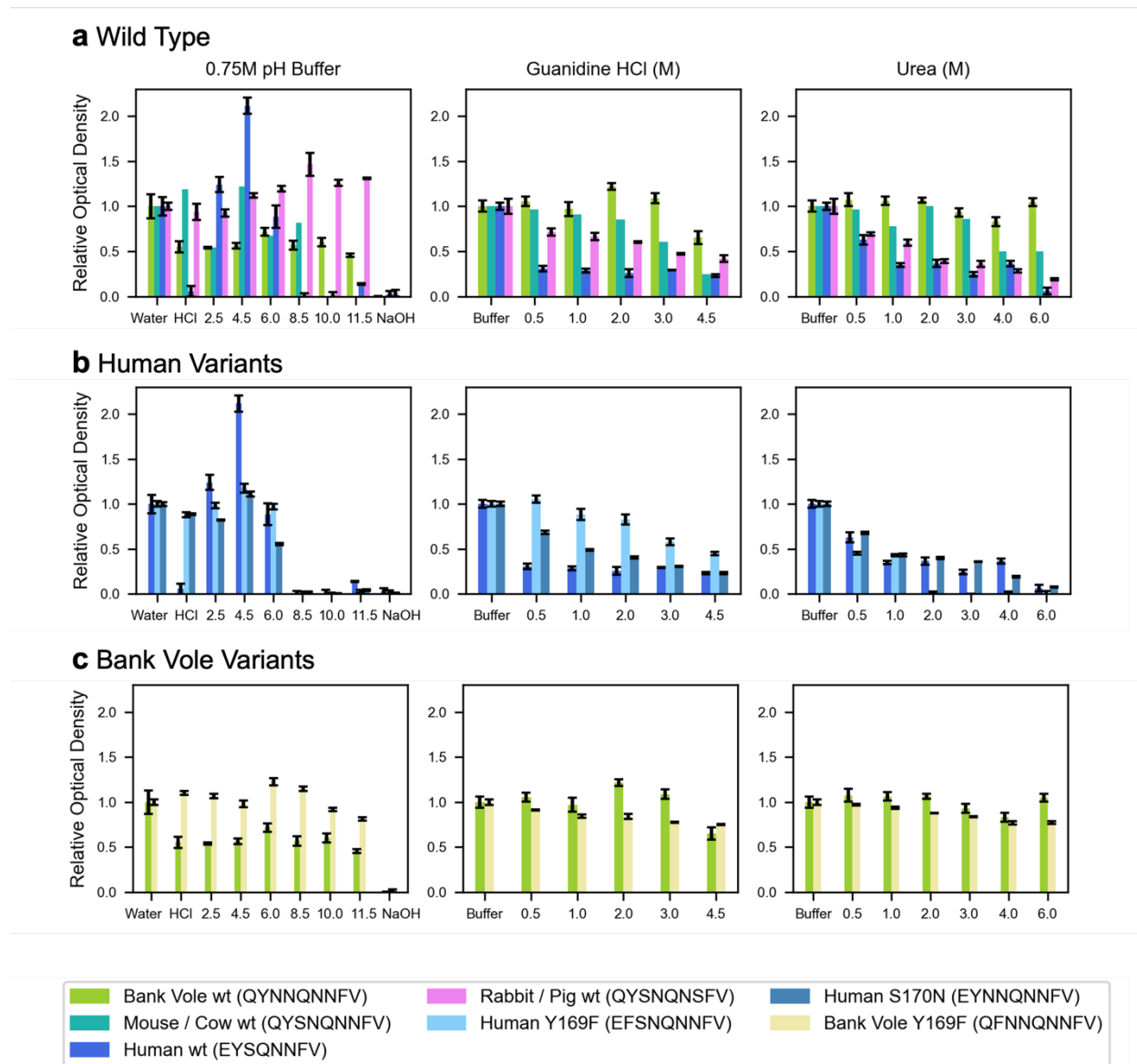


Figure 2.3. Denaturation of $\beta 2\alpha 2$ loop crystals. (a) Wild type peptide crystals were subjected to 0.75M buffers at the specified pH (left column), increasing concentrations of guanidine HCl (center column), and increasing concentrations of urea (right column). (b) Human and (c) bank vole variants, including the wild type crystals shown in (a), plotted on the same axis as their wild type counterparts for direct comparison

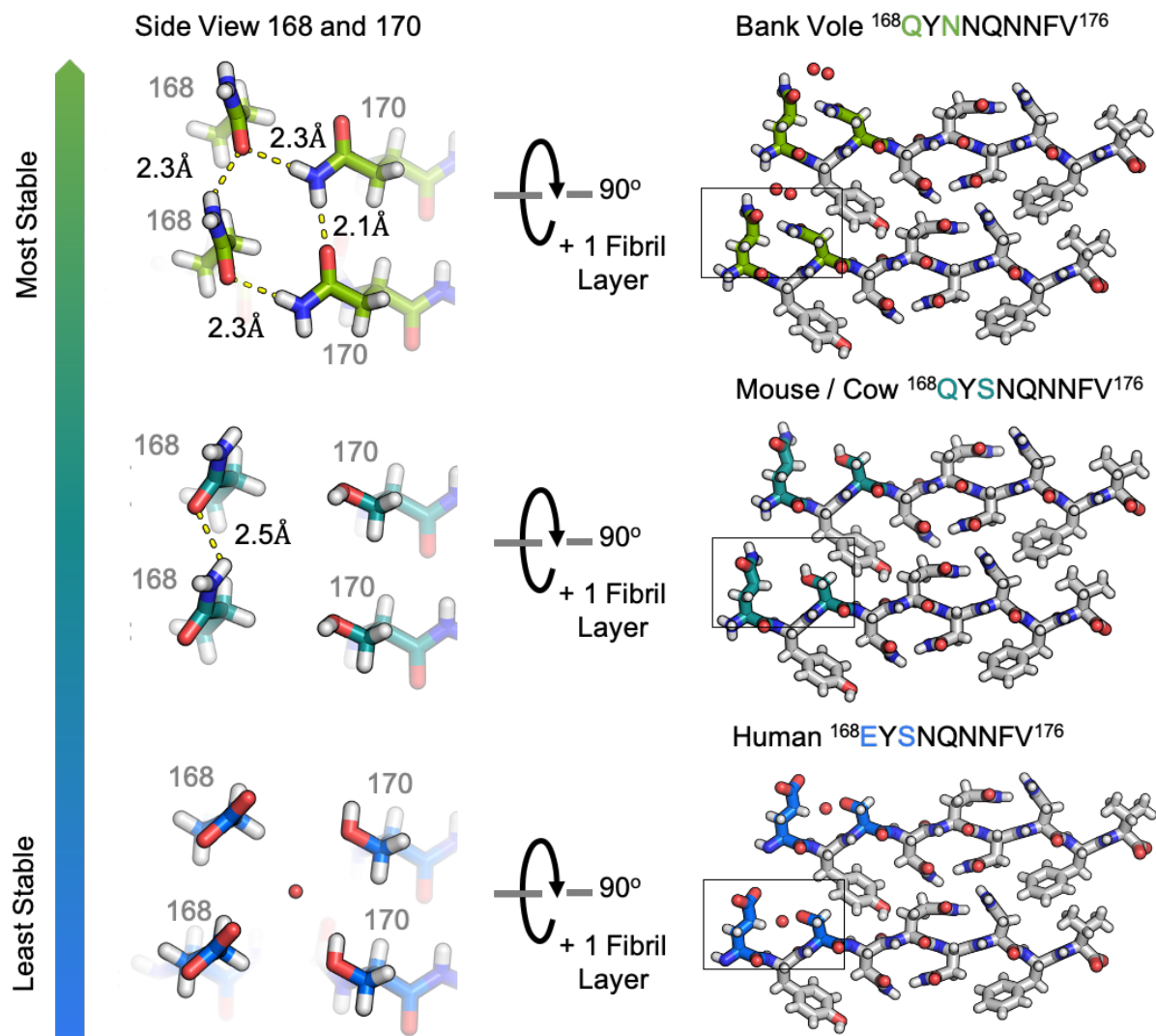
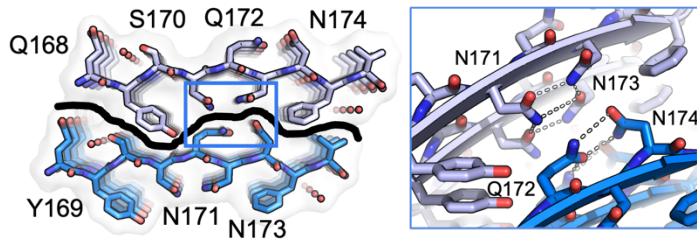


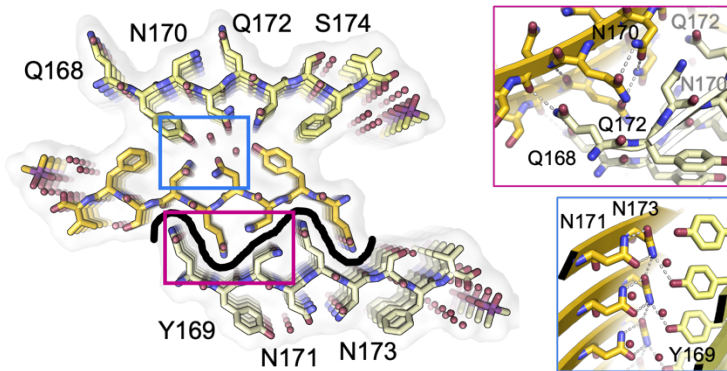
Figure 2.4. Structural Conservation in the $\beta 2\alpha 2$ Loop. Structures from the amyloid state of bank vole (top) mouse (center) and human (bottom) $\beta 2\alpha 2$ loops show a conserved interface and sets of hydrogen bond networks (right). Species-specific differences in the hydrogen bonding networks between residues 168 and 170 give rise to differing stabilities for the three crystalline aggregates (left).

Packing 1: Human ¹⁶⁸EYSNQNNFV¹⁷⁶



Species of Origin	$\beta 2\alpha 2$ loop 168-176
bank vole	QYNNQNNFV
mouse	QYSNQNNFV
human	EYSNQNNFV

Packing 2: Naked Mole Rat ¹⁶⁸QYNNQNSFV¹⁷⁶



Species of Origin	$\beta 2\alpha 2$ loop 168-176
mole rat	QYNNQNSFV
elk (6mer)	NNQNIF
rabbit	QYSNQNSFV*

Figure 2.5. $\beta 2\alpha 2$ Loop Structures Fall Into Two Structural Categories. Packing arrangement for mouse, human, and bank vole (top) and mole rat, elk, and rabbit (bottom) $\beta 2\alpha 2$ loops with black line highlighting the promiscuous (top) and selective (bottom) interfaces formed in each packing category. Insets show a conserved hydrogen bonding network across species between N171 and N173 (center column, blue boxes). For bank vole, mouse, and human loops (top) N171 and N174 form a clasp (center column, top blue box) while elk and naked mole rat structures (bottom) form a clasp between N170 and N172 (center column, center pink box), leaving Q168 free to form a polar ladder that can interdigitate with a complementary mated sheet. Residues that vary between species are colored in text and residue 174 is shown in color and underlined. *The rabbit loop structure falls into packing category 2, but has additional unique features that distinguish it from both categories.

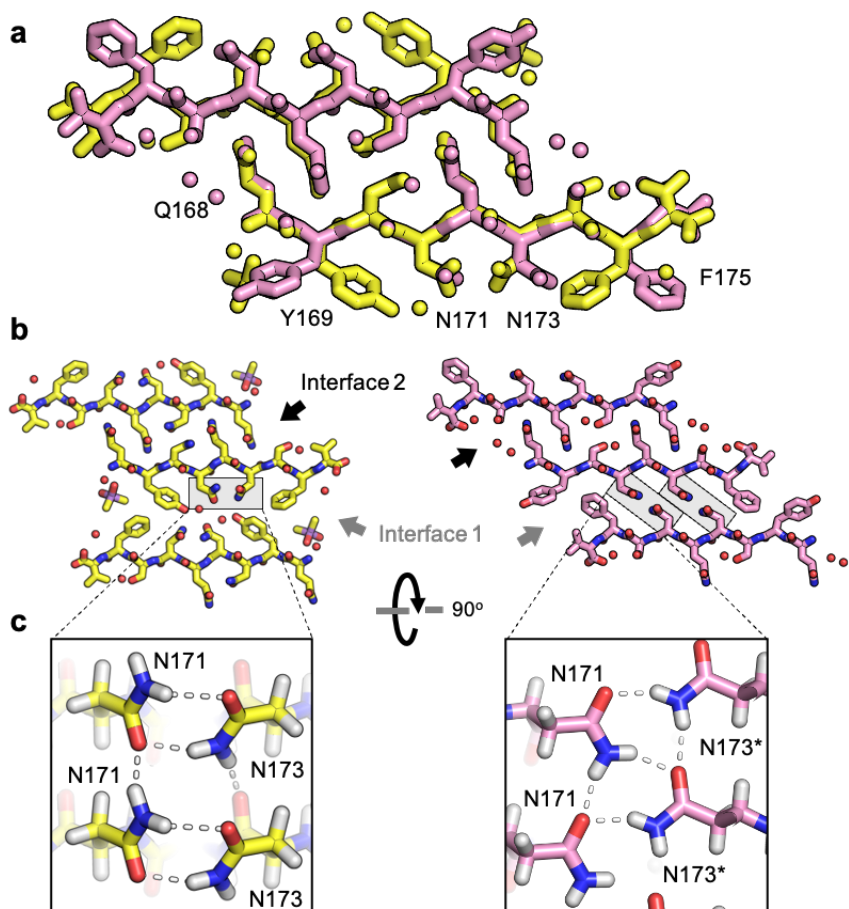


Figure 2.6. Comparison of naked mole rat and rabbit loop structures. (a) Naked mole rat (yellow) and rabbit (pink) loop structures share a common interface seen in packing category 2. (b) Interface 1 is composed of inward facing aromatic residues and a polar clasp between N171 and N173 on the same strand for all structures regardless of packing category with the exception of the rabbit loop structure (right). (c) A polar clasp is formed in both non-rabbit (left) and rabbit (right) structures, however in the rabbit structure the clasp is formed by N171 and N173 on mated sheets rather than via interactions within a single sheet.

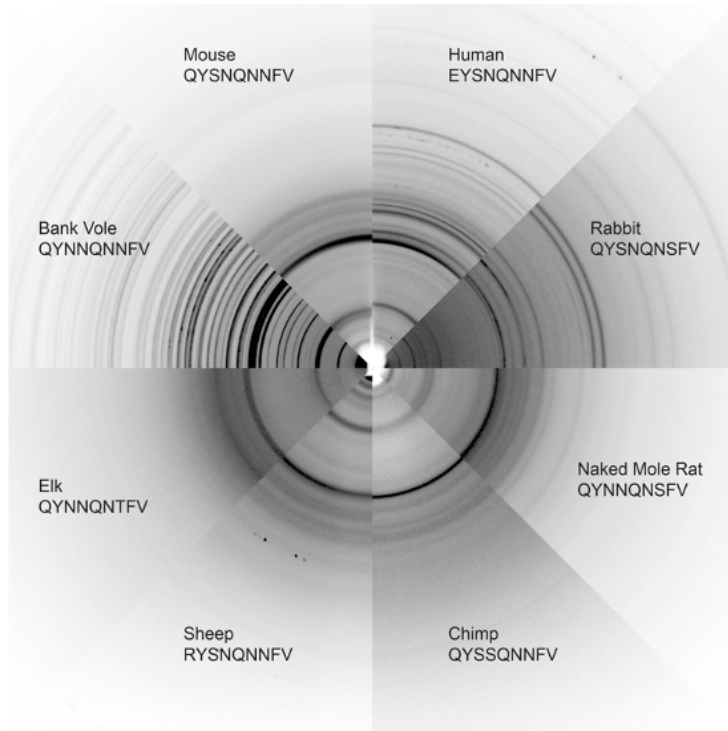
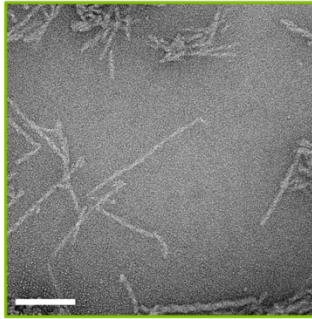
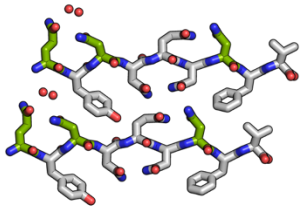
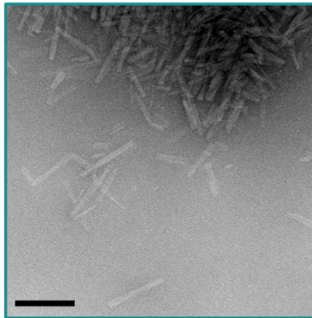
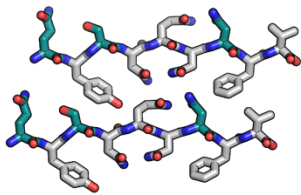


Figure 2.7. Fiber Diffraction of $\beta 2\alpha 2$ loop aggregates. Fiber diffraction patterns were collected for crystalline (bank vole, mouse, human, rabbit, and naked mole rat) and fibrillar (chimp, sheep R168 sequence polymorph, and elk) aggregates. All aggregates share ~ 4.8 and $\sim 10\text{\AA}$ reflections indicative of underlying amyloid structure.

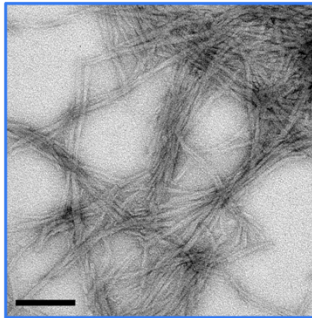
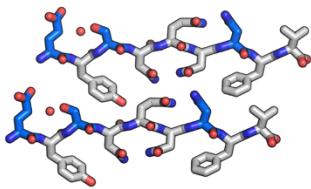
Bank Vole ¹⁶⁸QYNNQNNFV¹⁷⁶



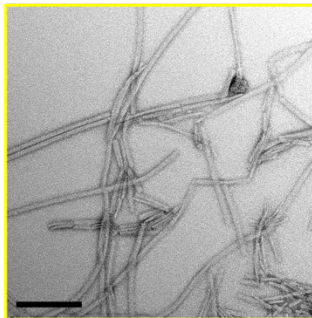
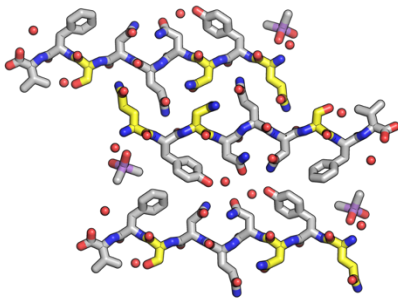
Mouse / Cow ¹⁶⁸QYSNQNNFV¹⁷⁶



Human ¹⁶⁸EYSNQNNFV¹⁷⁶



Naked Mole Rat ¹⁶⁸QYNNQNSFV¹⁷⁶



Elk ¹⁷⁰NNQNTFV¹⁷⁵

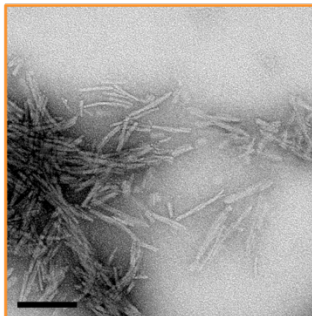
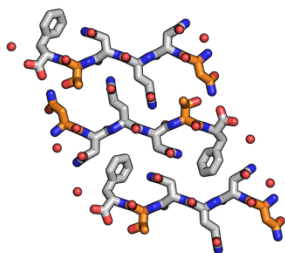


Figure 2.8. Influence of the $\beta 2\alpha 2$ loop on fibril structure. Crystal structures for bank vole, mouse / cow, human, naked mole rat, and elk $\beta 2\alpha 2$ loops are shown along with fibrils formed by rBvPrP⁹⁴⁻¹⁷⁸ harboring the loop substitutions in residues 168-176 belonging to each species. Scalebars = 200nm.

Table 1: Crystallization Conditions for β 2 α 2 Loop Peptides

Species (Mutation)	Sequence	Concentration (mg/ml)	Solute	Filtered	Crystallization Condition
Human	EYSNQNNFV	4	Water	No	20% EtOH 0.1M NaAc pH 4.5 0.1M Li ₂ SO ₄
Mouse / Cow	QYSNQNNFV	10	Water	No	15% EtOH 0.2M MES pH 6.0 0.5M ZnOAc
Naked mole rat	QYNNQNSFV	4	1% DMSO	Yes	5% isopropanol 0.1M NaCacodylate pH 6.5 0.1M ZnAc
Rabbit / Pig	QYSNQNSFV	2	4% DMSO	Yes	0.05M CHES pH 9 0.2M Li ₂ SO ₄ 1.3M Na/K Tartrate
Human (S170N)	EYNNQNNFV	2	Water	Yes	0.1M NaOAc pH 4.5 1M NaCl 0.1M Li ₂ SO ₄
Human (Y169F)	EFSNQNNFV	2	Water	Yes	Water
Bank Vole (Y169F)	QFNNQNNFV	2	Water	No	10% MPD 0.2M MES pH 6
Bank Vole (Q172E)	QYNNENNfV	1	Water	Yes	0.1M NaAc pH 4.5 2.5M NaCl 0.1M Li ₂ SO ₄
Human / Mouse (Y169G)	GSNQNNF	10	Water	Yes	10% (w/v) PEG-8000 0.1M MES pH 6 0.2M Zn(OAc) ₂
Human / Mouse / Cow	YSNQNNF	10	Water	Yes	10% (w/v) PEG-3000; 0.1M Na/K phosphate pH 6.2

Table 2. Data collection and refinement statistics

7	EYSNQNNFV (7RVC)	QYSNQNNFV (7RVD)	QYNNQNSFV (7RVI)	QYSNQNSFV (7RVG)	EYNNQNNFV (7RVE)	EFSNQNNFV (7RVL)	QFNNQNNFV (7RVF)	QYNNENNfV (7RVH)	GSNQNNF (7RVK)	YSNQNNF (7RVJ)
Data collection										
No. Crystals Merged	6	8	5	4	3	8	9	6	7	7
Space group	P1	P1	C2	P2 ₁	P1	P1	P1	P1	P1	P1
Cell dimensions										
<i>a</i> , <i>b</i> , <i>c</i> (Å)	10.02 4.89	4.87 10.17	62.76 4.85	23.57 4.86	4.93 10.14	4.90 10.38	4.88 10.56	4.87 10.06	14.14 4.87	24.39 4.95
α , β , γ (°)	31.33	31.29	21.52	27.70	31.56	30.26	29.98	30.66	18.15	20.86
<i>a</i> , <i>b</i> , <i>c</i> (°)	90.99 91.43	94.65 90.73	90.00 109.14	90.00 111.21	94.13 90.59	90.91 90.82	93.89 92.38	94.85 90.26	93.21 91.02	86.52 77.16
<i>a</i> , <i>b</i> , <i>c</i> (°)	102.18	101.15	90.00	90.00	102.74	102.25	103.29	99.99	101.83	85.68
Resolution (Å) ^a	10.44 - 1.00	10.39 - 1.00	10.76 - 1.05	7.18 - 1.00	9.86 - 0.85	10.14 - 1.00	10.24 - 1.00	7.98 - 0.90	13.84 - 1.00	11.87 - 1.00
<i>a</i> , <i>b</i> , <i>c</i> (°)	(1.03 - 1.00)	(1.03 - 1.00)	(1.08 - 1.05)	(1.03 - 1.00)	(0.87 - 0.85)	(1.00 - 1.03)	(1.03 - 1.00)	(0.93 - 0.90)	(1.03 - 1.00)	(1.03 - 1.00)
<i>R</i> _{merge}	0.187 (0.432)	0.167 (0.514)	0.287 (0.589)	0.204 (0.698)	0.185 (0.505)	0.178 (0.363)	0.229 (0.636)	0.197 (0.475)	0.238 (0.430)	0.283 (0.478)
<i>I</i> / σ (<i>I</i>)	5.56 (2.11)	3.99 (1.33)	5.00 (2.47)	5.72 (2.45)	3.81 (1.22)	5.99 (3.66)	5.33 (2.28)	5.21 (2.56)	5.05 (2.25)	4.49 (1.72)
<i>CC</i> _{1/2}	0.988 (0.840)	0.988 (0.742)	0.981 (0.931)	0.994 (0.699)	0.985 (0.617)	0.985 (0.940)	0.980 (0.711)	0.988 (0.783)	0.980 (0.862)	0.957 (0.336)
Completeness (%)	97.49 (77.00)	93.74 (94.70)	80.00 (83.50)	74.90 (71.70)	89.80 (43.20)	87.10 (69.70)	94.80 (54.90)	81.00 (82.30)	80.30 (64.70)	85.80 (56.5)
Redundancy	5.2 (4.1)	3.0 (2.4)	8.4 (8.6)	7.4 (6.6)	3.5 (2.4)	5.4 (5.0)	6.2 (5.2)	5.4 (4.9)	6.2 (4.3)	5.5 (2.3)
Refinement										
Resolution (Å)	1.00 (1.26 - 1.00)	1.00 (1.26 - 1.00)	1.05 (1.32 - 1.05)	1.00 (1.26 - 1.00)	0.85 (0.97 - 0.85)	1.00 (1.26 - 1.00)	1.00 (1.03 - 1.00)	0.90 (0.92 - 0.90)	1.00 (1.12 - 1.00)	1.00 (1.12 - 1.00)
No. reflections	3031 (167)	2977 (215)	2612 (187)	2669 (181)	4694 (161)	2737 (163)	2938 (123)	2076 (123)	4309 (204)	
<i>R</i> _{work}	0.163 (0.174)	0.201 (0.199)	0.227 (0.212)	0.218 (0.259)	0.218 (0.274)	0.192 (0.196)	0.221 (0.336)	0.222 (0.302)	0.202 (0.208)	0.226 (0.222)
<i>R</i> _{free}	0.168 (0.188)	0.250 (0.257)	0.265 (0.264)	0.220 (0.277)	0.266 (0.331)	0.211 (0.213)	0.251 (0.252)	0.235 (0.254)	0.238 (0.222)	0.233 (0.236)
No. atoms										
Protein	144	147	79	144	148	143	148	148	97	224
Ligand/ion	0	0	6	0	0	0	0	0	8	0
(specify/describe)			(NaCacodylate)						(ZnAcetate)	
Water	1	0	3	4	2	1	2	2	6	8
<i>B</i> factors										
Protein	9.09	8.49	3.33	5.41	23.60	7.95	6.25	4.48	4.95	3.57
Ligand/ion		-	9.47	-	-	-	-	-	8.71	-
Water	20.04	-	9.01	11.46	22.27	24.20	21.24	12.64	14.52	4.40
R.m.s. deviations										
Bond lengths (Å)	0.017	0.008	0.011	0.011	0.013	0.010	0.008	0.009	0.010	0.010
Bond angles (°)	1.607	0.920	1.086	0.844	1.106	1.019	1.662	1.306	0.940	0.910

^a Values in parentheses are for highest-resolution shell

Chapter 3

Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core

The work described in this chapter has been reproduced from:

Calina Glynn, Michael R. Sawaya, Peng Ge, Marcus Gallagher-Jones, Connor W. Short, Ronquiah Bowman, Marcin Apostol, Z. Hong Zhou, David S. Eisenberg and Jose A. Rodriguez. "Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core." *Nature Structural and Molecular Biology*, 2020.

Copyright 2020

Calina Glynn, Michael R. Sawaya, Peng Ge, Marcus Gallagher-Jones, Connor W. Short, Ronquiah Bowman, Marcin Apostol, Z. Hong Zhou, David S. Eisenberg and Jose A. Rodriguez.



Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core

Calina Glynn¹, Michael R. Sawaya², Peng Ge³, Marcus Gallagher-Jones¹, Connor W. Short¹, Ronquiah Bowman¹, Marcin Apostol^{2,5}, Z. Hong Zhou^{3,4}, David S. Eisenberg² and Jose A. Rodriguez¹✉

Self-templating assemblies of the human prion protein are clinically associated with transmissible spongiform encephalopathies. Here we present the cryo-EM structure of a denaturant- and protease-resistant fibril formed in vitro spontaneously by a 9.7-kDa unglycosylated fragment of the human prion protein. This human prion fibril contains two protofilaments intertwined with screw symmetry and linked by a tightly packed hydrophobic interface. Each protofilament consists of an extended beta arch formed by residues 106 to 145 of the prion protein, a hydrophobic and highly fibrillogenic disease-associated segment. Such structures of prion polymorphs serve as blueprints on which to evaluate the potential impact of sequence variants on prion disease.

Prions¹, from benign assemblies to pathogenic proteinaceous infectious particles, undergo self-templated propagation and are found across the domains of life^{2–4}. The protein-only hypothesis for prion proteopathy rests on the existence of a polypeptide quaternary structure that encodes infectious and toxic properties⁵. The mammalian prion protein can adopt both a native globular fold (PrP^C) and a self-templating infectious fold (PrP^{Sc}) that can spread within or between certain species¹. Despite the availability of native PrP^C (ref. 7) structures, atomic structures of PrP^{Sc} are lacking^{6–10} and its mechanism for self-templating and lateral transmission remains a mystery^{11–13}.

The enigma of prion propagation is deepened by the capacity of a single prion sequence to present variable pathologies^{1,12}. Prion-induced disorders, such as transmissible spongiform encephalopathies (TSEs)¹³, can have both genetic and sporadic origins^{12,13}. Those caused by coding variants of the human prion protein include Creutzfeldt–Jakob disease (CJD), fatal familial insomnia (FFI) and Gertsmann–Sträussler–Scheinker (GSS), while sporadic TSEs include cases of Kuru, CJD, sporadic fatal insomnia (sFI), iatrogenic CJD and the new variant form of CJD (vCJD) associated with bovine encephalopathy^{1,13}.

A molecular basis for prion disease has been established in part through the characterization of PrP^{Sc} isolates¹. PrP^{Sc} from infected tissues can be resistant to proteolysis and denaturation^{14,15} and associated with fibrillar deposits¹⁶. Assemblies of infectious protease-resistant prions (PrP^{Res}) can be associated with disease, fibrillar and variable in size and infectivity^{17,18}. Electron micrographs of one protease-resistant infectious prion fragment, known as PrP 27–30 (27–30 kDa)¹⁹, reveal polymorphic deposits that contain beta-sheet-rich fibrils or rods^{20,21}, potentially comprised of two intertwined protofilaments²².

Progress in prion studies has been accelerated by laboratory-generated constructs. Atomic structures of peptides fewer than 10 residues long reveal partial clues to the molecular basis for

prion assemblies^{23,24}, but lack context. PrP segments as short as 21 residues in length are fibrillogenic and toxic, but incapable of propagating disease²⁵, while recombinant fibrils of an 89-residue segment of bank vole PrP (PrP 90–178) are likewise fibrillogenic and show biophysical characteristics that mimic brain-derived PrP^{Sc} (ref. 25). Recombinantly generated^{26,27}, C-terminally truncated²⁸ and unglycosylated²⁹ prions have now been shown to spontaneously form PrP^{Sc}-like aggregates that under certain conditions can retain infectivity and recapitulate transmission barriers in mouse models^{27,30,31}. For example, recombinant, C-terminally-truncated prion amyloid fibrils with parallel, in-register strands—whose most ordered region includes residues 112–139 of the mouse prion sequence^{32,33} (moPrP23-144)—are infectious to mice³⁴. Thus, recombinant forms of mammalian PrP can create synthetic pathogenic prions²⁸. However, structural variation between in vitro-generated and tissue-isolated pathogenic prions can exist, as illustrated by comparisons of recombinant PrP fibrils generated in vitro with fibrils from full-length ex vivo infectious mouse prion isolates^{22,34,35}. These differences underscore the need for atomic structures of both tissue-isolated and in vitro-generated prion assemblies.

Results

Cryo-EM of recombinant, protease-resistant human prion fibrils. To gain atomic insights into the biophysical properties of prion assemblies, we interrogated the structure of prion fibrils formed by an 85-residue, 9.7-kDa fibrillogenic peptide encoding residues 94–178 of the human prion protein (rPrP 94–178) (Extended Data Fig. 1). This recombinant, unglycosylated segment encompasses a substantial portion of the protease-resistant core of PrP^{Sc} (ref. 16,36,37) and the entire hydrophobic region of PrP (Fig. 1a)^{38–41}. Within these regions are major portions of fibrillogenic segments previously shown to form infectious species with biophysical properties that mimic those of PrP^{Sc} (ref. 25,28,32). Continuous mixing

¹Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles, Los Angeles, CA, USA. ²Department of Biological Chemistry and Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA, USA. ³California NanoSystems Institute, University of California Los Angeles, Los Angeles, CA, USA. ⁴Department of Microbiology Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, CA, USA. ⁵Present address: ADRx, Thousand Oaks, CA, USA. ✉e-mail: jrodriguez@mbi.ucla.edu

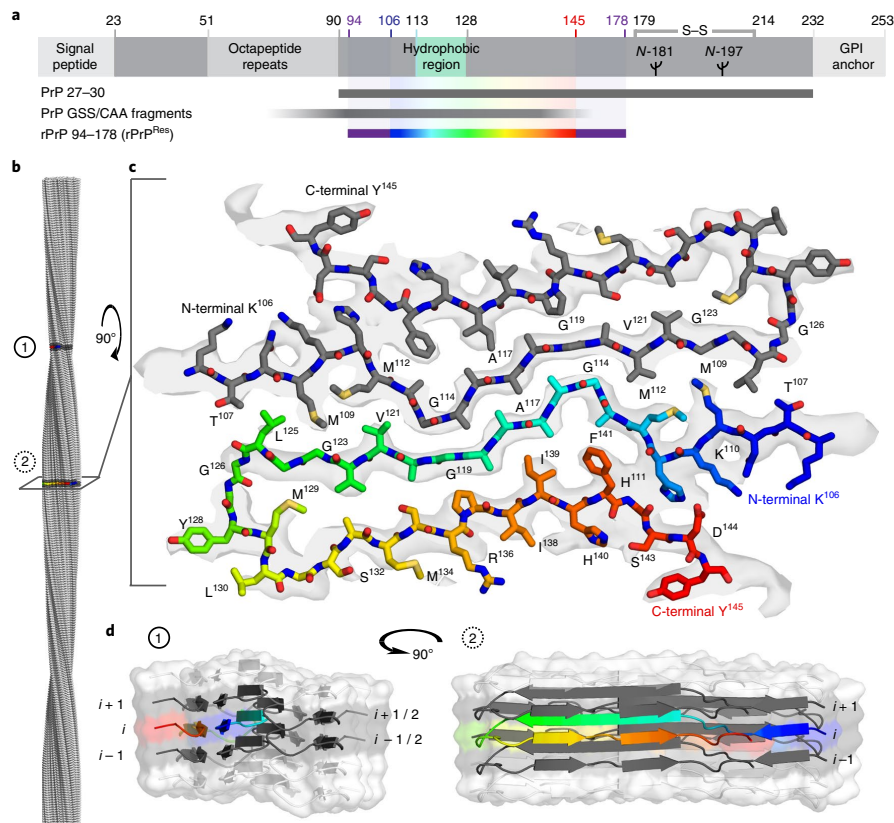


Fig. 1 PrP domain architecture and structure of the rPrP^{Res} fibril, the protease-resistant core of rPrP 94–178. **a**, Schematic representation of prion protein features and segments of interest. The sequence of PrP 27–30 (ref. 19), a protease-resistant disease-associated fragment^{54,55} is shown. The region of the prion protein approximately corresponding to PrP 27–30 (ref. 19), a protease-resistant scrapie prion fragment, is shown as a solid gray bar. Fragments approximately 7- to 8-kDa in size observed in people with GSS and CAA^{54,55} are indicated by a translucent gray bar. A purple bar and numbers indicate the rPrP 94–178 fragment, and, within it, the segment that forms rPrP^{Res} is shown in rainbow, starting with residue 106 (blue) and ending with residue 145 (red). The disulfide bond between residues 179 and 214 in the native protein and the two possible N-linked glycosylation sites at residues 181 and 197 are noted. **b**, Surface representation of the fibril model indicating its full 1,260-Å pitch with regions of interest colored in dark gray and labeled 1 and 2. **c**, A cross-section of the fibril with a stick representation of the model built into density is shown with termini, and select residues are labeled. **d**, A space-filled and cartoon model of ten chains in a rPrP^{Res} filament. The trajectory of a single rainbow-colored strand (i) and the four (gray) chains it directly contacts are shown. The two orthogonal views correspond to regions 1 and 2 of the fibril shown in **b**.

of recombinantly purified rPrP^{94–178} in acidified 4M urea produced a heterogeneous population of aggregates that included amorphous species, fibrils and rods (Extended Data Fig. 1)²⁵. Treatment with proteinase-K clarified the rPrP 94–178 aggregates by removing some species while sparing others (Extended Data Fig. 2). The remnants of protease digestion were preserved in a solution of 2% SDS in a frozen-hydrated state for high-resolution cryo-electron microscopy (cryo-EM) (Extended Data Figs. 1 and 2).

Views of rPrP 94–178 fibrils were sorted into two-dimensional (2D) classes that revealed one major, and several less abundant, morphologies. The most abundant morphology included unbranched fibrils approximately 8–10 nm wide with a 126-nm helical pitch; this type of fibril represented approximately 71% of selected particles. The other morphologies, broadly classified as rods or ribbons of thickness varying from 7.2–18.8 nm, accounted for the remaining 29% of selected particles (Extended Data Figs. 3 and 4).

The three-dimensional structures of the less abundant species were not determined in this work owing to their apparent absence of a regular helical twist. For this reason, we could not discern whether 2D classes of rod or ribbon fibrils represented different views of the same polymorph.

Architecture of a recombinant, protease-resistant human prion fibril polymorph. A 3.5-Å-resolution 3D map of the rPrP 94–178 fibril core (Fig. 1c) was obtained by helical reconstruction with 2_1 screw symmetry from 76,246 square windows with a length of 428 Å, selected along the length of rPrP 94–178 fibrils (Extended Data Fig. 3 and Table 1). Approximately half of the 85 residues in rPrP 94–178 could be placed into the cryo-EM map and modeled as part of two symmetric protofilaments that compose rPrP 94–178 fibrils. These represent a protease-resistant structure we refer to as rPrP^{Res} (Fig. 1 and Extended Data Fig. 5).

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

	rPrP ^{Res} (EMD-20900, PDB 6UUR)
Data collection	
Magnification	×130,000
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	36
Defocus range (μm)	1.1–5.1
Pixel size (Å) ^a	1.07
Symmetry imposed	C1 (2, screw)
Helical rise (Å)	2.4
Helical twist (°)	179.3
Initial particles images (no.)	217,988
Final particle images (no.)	76,246
Map Resolution (Å)	3.54 (3.19)
FSC threshold	0.5 (0.143)
Map resolution range (Å)	200–3.54
Refinement	
Model resolution (Å)	3.60 (3.19)
FSC threshold	0.5 (0.143)
Model resolution range (Å)	200–3.60
Map sharpening B factor (Å ²)	−139.98
Model composition	
Nonhydrogen atoms	2,770 (10 chains)
Protein residues	400 (10 chains)
B factors (Å ²)	
Protein	23.7
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.752
Validation	
MolProbity score	1.81
Clashscore	8.15
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	94.74
Allowed (%)	5.26
Disallowed (%)	0.00

^a0.535 Å for super-resolution mode

Secondary structure in rPrP^{Res} fibrils differs from that in PrP^C; segments that in the globular structure of PrP^C form loops or helices, make four beta strands linked by turns in rPrP^{Res} (Fig. 2a). Strands in each rPrP^{Res} protofilament stack parallel, in-register, spaced 4.8 Å apart (Extended Data Fig. 5). The overall structure of rPrP^{Res} is reminiscent of, but distinct from, predicted models of truncated PrP fibrils^{32,42}, and segments of rPrP^{Res} coincide with atomic structures of prion hexapeptides to varying degrees. Some hexapeptide structures (Table 2) are predictive of local secondary structure in rPrP^{Res} and mimic conformations suited for close amyloid-like packing in the context of the fibril (Extended Data Fig. 6 and Supplementary Table 1).

Stabilizing features of rPrP^{Res} fibrils. The rPrP^{Res} fibril is a compact, low-energy structure with canonical amyloid features (Fig. 3).

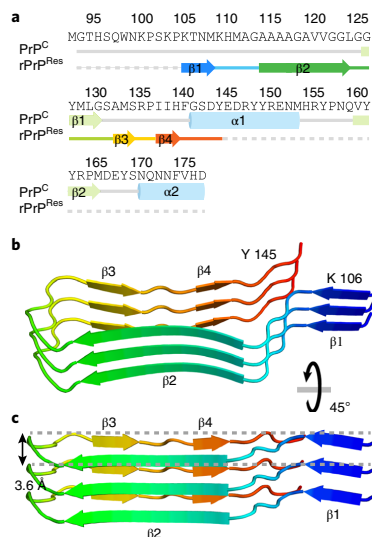


Fig. 2 | Secondary structure of rPrP^{Res} in contrast to that of PrP^C and topology of the rPrP^{Res} protofilament. **a**, Sequence and secondary structure alignment of residues 94–178 of natively folded human PrP^C and an amyloid conformation of rPrP^{94–178} observed here (rPrP^{Res}). Dotted lines represent residues included in the construct but not built into the model. A single nonnative methionine was added as a start codon for rPrP 94–178. **b**, Spatial arrangement of three stacked beta strands shown in **a**, making up one-half of the filament. **c**, A side view of the same side of the filament highlighting the flat nature of the structure, which only spans 3.6 Å from the highest to lowest backbone position along the fiber axis in a single chain.

Each chain in rPrP^{Res} forms a beta arch whose ordered core starts at K106 and ends with Y145 in the human PrP sequence. Within a protofilament, the amino- and carboxy-terminal halves of each beta arch meet at a tight, dry interface, one of two unique interfaces in rPrP^{Res} (Fig. 3).

The two protofilaments meet at the second tight interface, with backbone atoms separated by a mere 3.5 Å at closest atomic contact (Fig. 3 and Extended Data Fig. 5). Each layer of rPrP^{Res} buries 44 apolar residues across two unique interfaces (Fig. 3). The interface between protofilaments is large; each molecule contributes 10 side chains to the interface and buries 460 Å². The core of each protofilament buries an additional 848 Å² per chain and is sealed from solvent by a pair of polar residues: H111 and D144 (Fig. 3 and Extended Data Figs. 5 and 7). Each rPrP^{Res} protofilament is also stabilized by a network of backbone hydrogen bonds that link strands along the growing axis of the fibril (Extended Data Fig. 7).

Compatibility of rPrP^{Res} with human sequence variants and patterns of disease progression. The proteopathic relevance of rPrP^{Res} is supported by its apparent incompatibility with sequence variants that may alter both PrP^{Sc} structure and the course of prion disease. A familial mutant (A117V)⁴³ produces a protease-sensitive form of PrP^{Sc} that appears to induce clashes at the protofilament interface in rPrP^{Res} (Fig. 4). Likewise, a natural variant (G114V)⁴⁴ associated with early onset but atypically long disease duration is poorly accommodated by the quaternary structure of rPrP^{Res} (Fig. 4). In contrast, a valine associated with the protective heterozygous

Table 2 | X-ray data collection and refinement statistics

	¹¹³ AGAAA ¹¹⁸ (PDB 6PQ5)	¹¹⁹ GAVVGG ¹²⁴ (PDB 6PQA)
Data collection^a		
Space group	P2 ₁	P2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	18.5, 9.5, 17.6	4.78, 12.8, 20.8
α , β , γ (°)	90.0, 120.5, 90.0	90, 90, 90
Resolution (Å)	1.5 (1.55–1.50) ^b	1.45 (1.50–1.45)
<i>R</i> _{merge}	0.18 (0.50)	0.15 (0.31)
<i>I</i> / σ (<i>I</i>)	5.42 (1.59)	7.21 (1.92)
Completeness (%)	91.5 (64.9)	93.6 (93.2)
Redundancy	3.4 (2.4)	5.4 (2.8)
Refinement		
Resolution (Å)	1.5	1.45
No. reflections	2,830	2,451
<i>R</i> _{work} / <i>R</i> _{free}	0.237 / 0.271	0.187 / 0.238
No. atoms		
Protein	60	32
Ligand (malonate)	14 (malonate)	0
Water	0	3
B factors		
Protein	14.4	5.0
Ligand (malonate)	28.8	0
Water	0	31.5
R.m.s. deviations		
Bond lengths (Å)	0.006	0.007
Bond angles (°)	1.1	1.1

^aData are from one crystal. ^bValues in parentheses are for highest-resolution shell.

condition (M/V129)⁴⁵ is not structurally precluded when modeled in place of methionine at position 129. However, V129 creates a small void within the core of each protofibril (Fig. 4). Thus, rPrP^{Res} leaves ambiguity in the origin of resistance to certain types of prion diseases conferred by 129 heterozygosity. The notion that sequence variants G114V, A117V and M129V might engender polymorphs with similar core geometry to rPrP^{Res} is supported by the adoption of energetically favorable folds by rPrP^{Res} models containing these mutations after minimal backbone rearrangement and side-chain repacking (Extended Data Fig. 8 and Supplementary Table 2). This raises the possibility that disease-relevant PrP sequence variants may represent closely related but distinct polymorphs, as observed for other amyloid assemblies^{46,47}.

A natural sequence variant of PrP (G127V)⁴⁸ completely prevents prion disease and is expected to induce clashes between protein layers in the structure of rPrP^{Res} (Fig. 4). This is in agreement with observations indicating this sequence variant hinders the formation of stable PrP fibrils by destabilizing beta sheets and disrupting hydrogen bonding^{49,50}. Unlike models based on rPrP^{Res} that contain variants G114V, A117V and M129V, those that contain G127V predict folds that are less energetically favorable than wild-type rPrP^{Res} despite backbone rearrangement and side-chain repacking (Extended Data Fig. 8d and Supplementary Table 2). These models offer a potential structural explanation for why mouse prion fibrils of the G127V variant require higher monomer concentrations to achieve nucleation and a lengthened lag phase of fibril growth⁵¹.

Compatibility of rPrP^{Res} with mammalian PrP sequences and species barriers. Insights into cross-seeding of PrP fibrils between mammals can be inferred from an analysis of sequence variability in PrP 106–145 (Extended Data Fig. 6). Only 7 of the 40 residues in the structure display any sequence variability across closely related mammals, and in most cases, physical properties are conserved. All seven varying residues (108, 109, 112, 138, 139, 143 and 145) are confined to the same region of the structure near the H111–D144 bridge. The positioning of these residues in rPrP^{Res} agrees with conclusions about cross-species conversion drawn from nuclear magnetic resonance (NMR) experiments on a similar fragment with known infectivity³². In those studies, a methionine variant of human residue I139 acts as a switch, changing the conformation of PrP^{Sc} from one resembling a human prion structure to one resembling that of a hamster. In rPrP^{Res}, this residue faces the protofilament core, and its methionine variant is disfavored (Extended Data Fig. 6). In contrast, a second methionine variant at I138, which is solvent-facing in rPrP^{Res}, does not alter the overall fibril fold of PrP^{Sc} (ref. ³²). Likewise, an M112V variant, found in mice, would create an interface vacancy that could disfavor, if not prevent, the formation of the rPrP^{Res} polymorph.

Discussion

The structural characterization of rPrP^{Res} presents a framework for evaluation and comparison of properties associated with protease-resistant prion assemblies. The major fibril polymorph of rPrP^{Res} forms a parallel, in-register structure with low solvation free energy, set apart from most other amyloid fibril structures (Fig. 3 and Supplementary Table 3), which is consistent with its resistance to denaturants, protease digestion and detergents (Extended Data Figs. 1 and 2). Assignment of residues within the core of rPrP^{Res} reveals a large number of apolar side chains, tightly packed to facilitate formation of its two beta arch protofilaments. The tight geometry of rPrP^{Res} allows for an evaluation of the potential impact of prion sequence variants on its overall fold. Collectively, these variants reflect the constraints on sequence variation imposed by the close packing of residues in rPrP^{Res}, in line with their impacts on the structure and function of PrP^{Sc}.

While rPrP^{Res} offers atomic insight into its structural accessibility to a variety of PrP sequence variants and homologs, this single polymorph represents only a small portion of the total structural space available to PrP, as evidenced by the diverse range of prion assemblies reported in the literature^{6–11,22,25,34}. The high degree of variation in prion strain properties, including growth kinetics, size distribution, proteinase K resistance, post-translational modification, cofactor binding, tissue distribution, infectivity and mortality, may be reflective of structural differences among strains⁵². Structural heterogeneity in fibrils isolated from prion strains contributes to the challenges facing an atomic definition of PrP^{Sc}. Uncovering the atomic structures of both recombinant and tissue-isolated prion assemblies may therefore yield clues to strain specific features.

Since in vitro-generated prions can, under certain conditions, offer greater control over fibril polymorph distributions, they may more readily yield structural insights into prion assemblies^{25,31,42}. However, the apparent structural differences between tissue-isolated infectious prion assemblies and those generated in vitro highlight the need for structural comparisons at the atomic scale between rPrP^{Res} and yet-to-be-determined structures of other prion polymorphs. For example, reconstructions of prion fibrils from various sources show a variety of structural repeats, ranging from ~19 Å (ref. ²²) to 60 Å (ref. ³³) that are not recapitulated by rPrP^{Res}. However, other molecular features of rPrP^{Res} are shared with amyloid-like fibrils formed by truncated Y145Stop prions that can transmit prion disease in mice and model a heritable, disease-causing human variant⁴². Like these disease-causing fibrils, the structure of rPrP^{Res} demonstrates that a protease-resistant human prion fibril polymorph can exhibit canonical amyloid features: parallel,

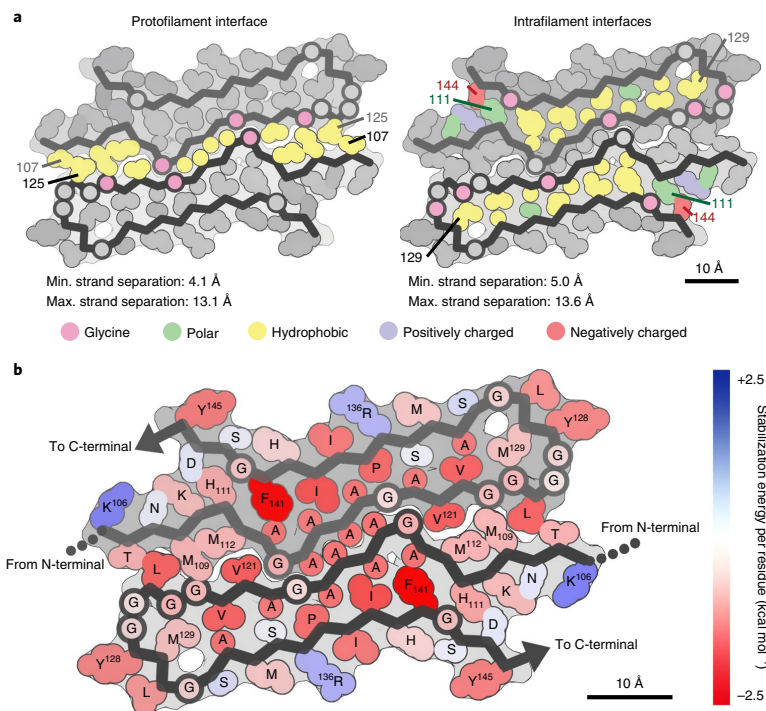


Fig. 3 | Stability of the core of rPrP^{Res}. **a**, Interfaces formed by each protofilament. The interface at the center of the fibril (left) is created by apolar residues between protofilaments where 460 Å² of surface area is buried. A second set of identical interfaces, one at the heart of each protofilament (right), buries 848 Å² of surface area per chain. Residues are colored according to their type, as shown in the color key. H111 is colored to be polar and could be protonated on the basis of this chemical environment. Maximum and minimum strand separation between backbones for each interface are also noted. **b**, Stabilization energy per residue based on calculations of energy of solvation. In the diagram, a single color is assigned to each residue, rather than each atom. The color corresponds to the sum of free energy values of each of the atoms in the residue. Across the structure, the average ΔG^0 per residue is -0.74 kcal mol⁻¹, or -29.6 kcal mol⁻¹ per chain. Stabilization energy is shown as a gradient from positive (blue) to negative (red) in units of kcal mol⁻¹; red is more stabilizing.

in-register strands spaced 4.8 Å apart, held together by the dry interfaces of a tightly packed core.

The relevance of rPrP^{Res} in the context of other prion assemblies is underscored by the similarity of its protease-resistant regions⁵⁴ and fibril dimensions to those found in patient-isolated samples⁵⁵ and synthetic, infectious prions³². For example, people with PrP cerebral amyloid angiopathy (CAA)⁵⁴ show a truncation of the prion protein at residue 145, whose soluble fragment forms 8–10 nm fibrils⁵⁴. The dimensions of these fibrils are similar in width to those of rPrP^{Res} (Fig. 1a and Extended Data Fig. 3). PrP fragments in CAA have been shown to include residues 90 to 145; most of which are clearly resolved in rPrP^{Res} fibrils⁵⁴ (Fig. 1a). Mutations that drive prion disease can also result in the formation of fibrils with similar residues at their core. For example, patients with GSS-inducing mutations accumulate short, ~7- to 8-kDa fragments that start near residue 90 of human PrP⁵⁵—akin to the segment that forms rPrP^{Res} (ref. 35). Overall, the structure of rPrP^{Res} is consistent with core residues, fragment size, filament dimensions, protease resistance and denaturant resistance of disease-associated prions^{54,55}. Nonetheless, alternative atomic arrangements at the core of prion assemblies may share features with rPrP^{Res} but ultimately vary in disease-relevant features, such as number of ordered residues, post-translational modifications⁵⁶ or interaction with co-factors⁵⁷.

Like the discovery of the prion⁵⁸, the unveiling of near-atomic structures of prion fibrils, including rPrP^{Res}, opens new opportunities to unravel the mysteries of protein-based infection. With it, we look forward to the structures of prion strains that will collectively provide a blueprint for atomically defined proteinaceous infectious agents.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-020-0403-y>.

Received: 12 November 2019; Accepted: 28 February 2020;
Published online: 13 April 2020

References

1. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
2. Wasmer, C. et al. Amyloid fibrils of the HET-s(218–289) prion form a β solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526 (2008).
3. Yuan, A. H. & Hochschild, A. A bacterial global regulator forms a prion. *Science* **355**, 198–201 (2017).

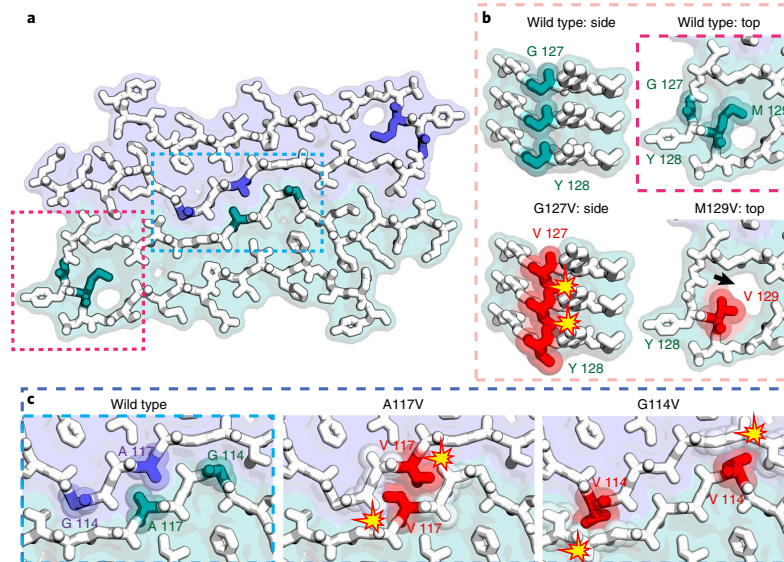


Fig. 4 | Compatibility of sequence variants at key residues in rPrP^{Res}. **a**, A space-filled model overlaid onto a stick representation of rPrP^{Res} in which one protofilament surface is shown in purple and the other in green. **b**, Magnified views of the turn region containing residues G127 and M129, where mutant G127V is protective against disease. A side view (left) shows that G127 fits well into the model, but V127 cannot, owing to steric clashes along the fibril axis (red). A view down the fibril axis (right) shows that both M129 and V129 polymorphisms can be accommodated by the model, but V129 introduces a small void (arrow) in the turn region of the rPrP^{Res} arch. **c**, Magnified view of the interface between protofilaments shows two GSS-associated mutants, A117V and G114V, that delay disease onset or produce protease-sensitive aggregates, respectively. Both are poorly accommodated in rPrP^{Res} owing to steric clashes with the opposing protofilament (red).

- Cobb, N. J. & Surewicz, W. K. Prion diseases and their biochemical mechanisms. *Biochemistry* **48**, 2574–2585 (2009).
- Riek, R. et al. NMR structure of the mouse prion protein domain PrP(121–231). *Nature* **382**, 180–182 (1996).
- Diaz-Espinoza, R. & Soto, C. High-resolution structure of infectious prion protein: the final frontier. *Nat. Struct. Mol. Biol.* **19**, 370–377 (2012).
- Wille, H. & Requena, J. R. The structure of PrPSc prions. *Pathogens* **7**, E20 (2018).
- Surewicz, W. K. & Apostol, M. I. in *Prion Proteins* (ed. Tatzelt, J.) 135–167 (Springer Berlin Heidelberg, 2011).
- Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
- Rodriguez, J. A., Jiang, L. & Eisenberg, D. S. Toward the atomic structure of PrPSc. *Cold Spring Harb. Perspect. Biol.* **9**, a031336 (2017).
- Wille, H. et al. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl Acad. Sci. USA* **99**, 3563–3568 (2002).
- Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. Prion protein biology. *Cell* **93**, 337–348 (1998).
- Collins, S. J., Lawson, V. A. & Masters, C. L. Transmissible spongiform encephalopathies. *Lancet* **363**, 51–61 (2004).
- Prusiner, S. B., Groth, D., Serban, A., Stahl, N. & Gabizon, R. Attempts to restore scrapie prion infectivity after exposure to protein denaturants. *Proc. Natl Acad. Sci. USA* **90**, 2793–2797 (1993).
- Caughey, B., Raymond, G. J., Kocisko, D. A. & Lansbury, P. T. Scrapie infectivity correlates with converting activity, protease resistance, and aggregation of scrapie-associated prion protein in guanidine denaturation studies. *J. Virol.* **71**, 4107–4110 (1997).
- McKinley, M. P., Bolton, D. C. & Prusiner, S. B. A protease-resistant protein is a structural component of the scrapie prion. *Cell* **35**, 57–62 (1983).
- Rubenstein, R. et al. Detection of scrapie-associated fibril (SAF) proteins using anti-SAF antibody in non-purified tissue preparations. *J. Gen. Virol.* **67**(Pt 4), 671–681 (1986).
- Silveira, J. R. et al. The most infectious prion protein particles. *Nature* **437**, 257–261 (2005).
- Bolton, D. C., Meyer, R. K. & Prusiner, S. B. Scrapie PrP 27–30 is a sialoglycoprotein. *J. Virol.* **53**, 596–606 (1985).
- Pan, K. M. et al. Conversion of α -helices into β -sheets features in the formation of the scrapie prion proteins. *Proc. Natl Acad. Sci.* **90**, 10962–10966 (1993).
- Nguyen, J. T. et al. X-ray diffraction of scrapie prion rods and PrP peptides. *J. Mol. Biol.* **252**, 412–422 (1995).
- Vázquez-Fernández, E. et al. The structural architecture of an infectious mammalian prion using electron cryomicroscopy. *PLoS Pathog.* **12**, e1005835 (2016).
- Gallagher-Jones, M. et al. Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat. Struct. Mol. Biol.* **25**, 131–134 (2018).
- Sawaya, M. R. et al. Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
- Wan, W. et al. Structural studies of truncated forms of the prion protein PrP. *Biophys. J.* **108**, 1548–1554 (2015).
- Kocisko, D. A. et al. Cell-free formation of protease-resistant prion protein. *Nature* **370**, 471 (1994).
- Legname, G. et al. Synthetic mammalian prions. *Science* **305**, 673–676 (2004).
- Choi, J.-K. et al. Amyloid fibrils from the N-terminal prion protein fragment are infectious. *Proc. Natl Acad. Sci. USA* **113**, 13851–13856 (2016).
- Taraboulos, A. et al. Acquisition of protease resistance by prion proteins in scrapie-infected cells does not require asparagine-linked glycosylation. *Proc. Natl Acad. Sci. USA* **87**, 8262–8266 (1990).
- Zhang, Z. et al. De novo generation of infectious prions with bacterially expressed recombinant prion protein. *FASEB J.* **27**, 4768–4775 (2013).
- Wille, H. et al. Natural and synthetic prion structure from X-ray fiber diffraction. *Proc. Natl Acad. Sci. USA* **106**, 16990–16995 (2009).
- Theint, T. et al. Species-dependent structural polymorphism of Y145Stop prion protein amyloid revealed by solid-state NMR spectroscopy. *Nat. Commun.* **8**, 753 (2017).
- Theint, T. et al. Structural studies of amyloid fibrils by paramagnetic solid-state nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **140**, 13161–13166 (2018).
- Terry, C. et al. Structural features distinguishing infectious ex vivo mammalian prions from non-infectious fibrillar assemblies generated in vitro. *Sci. Rep.* **9**, 376 (2019).
- Terry, C. et al. Ex vivo mammalian prions are formed of paired double helical prion protein fibrils. *Open Biol.* **6**, 160035 (2016).

36. Prusiner, S. B., Groth, D. F., Bolton, D. C., Kent, S. B. & Hood, L. E. Purification and structural studies of a major scrapie prion protein. *Cell* **38**, 127–134 (1984).
37. Stahl, N. et al. Structural studies of the scrapie prion protein using mass spectrometry and amino acid sequencing. *Biochemistry* **32**, 1991–2002 (1993).
38. Walsh, P., Simonetti, K. & Sharpe, S. Core structure of amyloid fibrils formed by residues 106–126 of the human prion protein. *Structure* **17**, 417–426 (2009).
39. Jobling, M. F. et al. The hydrophobic core sequence modulates the neurotoxic and secondary structure properties of the prion peptide 106–126. *J. Neurochem.* **73**, 1557–1565 (1999).
40. Biasini, E. et al. The hydrophobic core region governs mutant prion protein aggregation and intracellular retention. *Biochem. J.* **430**, 477–486 (2010).
41. Norstrom, E. M. & Mastroianni, J. A. The AGAAAAGA palindrome in PrP is required to generate a productive PrP^{Sc}-PrP^C complex that leads to prion propagation. *J. Biol. Chem.* **280**, 27236–27243 (2005).
42. Aucoin, D. et al. Protein-solvent interfaces in human Y145Stop prion protein amyloid fibrils probed by paramagnetic solid-state NMR spectroscopy. *J. Struct. Biol.* **206**, 36–42 (2019).
43. Asante, E. A. et al. Inherited prion disease A117V is not simply a proteinopathy but produces prions transmissible to transgenic mice expressing homologous prion protein. *PLoS Pathog.* **9**, e1003643 (2013).
44. Rodriguez, M.-M. et al. A novel mutation (G114V) in the prion protein gene in a family with inherited prion disease. *Neurology* **64**, 1455–1457 (2005).
45. Collinge, J., Palmer, M. S. & Dryden, A. J. Genetic predisposition to iatrogenic Creutzfeldt–Jakob disease. *Lancet* **337**, 1441–1442 (1991).
46. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
47. Falcon, B. et al. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
48. Asante, E. A. et al. A naturally occurring variant of the human prion protein completely prevents prion disease. *Nature* **522**, 478–481 (2015).
49. Zheng, Z. et al. Structural basis for the complete resistance of the human prion protein mutant G127V to prion disease. *Sci. Rep.* **8**, 13211 (2018).
50. Zhou, S., Shi, D., Liu, X., Liu, H. & Yao, X. Protective V127 prion variant prevents prion disease by interrupting the formation of dimer and fibril from molecular dynamics simulations. *Sci. Rep.* **6**, 21804 (2016).
51. Sabareesan, A. T. & Udgaonkar, J. B. The G126V mutation in the mouse prion protein hinders nucleation-dependent fibril formation by slowing initial fibril growth and by increasing the critical concentration. *Biochemistry* **56**, 5931–5942 (2017).
52. Morales, R. Prion strains in mammals: different conformations leading to disease. *PLoS Pathog.* **13**, e1006323 (2017).
53. Tattum, M. H. et al. Elongated oligomers assemble into mammalian PrP amyloid fibrils. *J. Mol. Biol.* **357**, 975–985 (2006).
54. Ghetti, B. et al. Vascular variant of prion protein cerebral amyloidosis with tau-positive neurofibrillary tangles: the phenotype of the stop codon 145 mutation in PRNP. *Proc. Natl Acad. Sci. USA* **93**, 744–748 (1996).
55. Piccardo, P. et al. Prion proteins with different conformations accumulate in Gerstmann-Sträussler-Scheinker disease caused by A117V and F198S mutations. *Am. J. Pathol.* **158**, 2201–2207 (2001).
56. Sim, V. L. & Caughey, B. Ultrastructures and strain comparison of underglycosylated scrapie prion fibrils. *Neurobiol. Aging* **30**, 2031–2042 (2009).
57. Li, Q. et al. Structural attributes of mammalian prion infectivity: insights from studies with synthetic prions. *J. Biol. Chem.* **293**, 18494–18503 (2018).
58. Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144 (1982).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Recombinant expression, purification and characterization of rPrP 94–178.

A gene encoding human PrP 94–178 (IDT) with an additional antecedent methionine start codon was cloned by Gibson Assembly into a derivative of a pET28 vector lacking purification or solubility tags and verified via DNA sequencing (GeneWiz). Recombinant human PrP 94–178 (rPrP 94–178) was expressed in *Escherichia coli* BL21 GOLD cells (Agilent). Briefly, cells were grown at 37°C shaking at 180–230 r.p.m. until optical density at 600 nm (OD_{600}) reached approximately 0.6. Protein overexpression was induced by addition of 1 mM IPTG, followed by 4–6 h of growth under the same conditions. Cells were then harvested and stored at –80°C until protein purification.

rPrP 94–178 was purified using previously described protocols as a guide³⁵. Specifically, cell pellets were resuspended in 25 mM Tris-HCl pH 8 and 5 mM EDTA (Buffer A) with 100× HALT protease inhibitor cocktail and lysed using an EmulsiFlex-C3 High Pressure Homogenizer (Avestin). Lysed cells were centrifuged at 30,000g for 1 h, and the supernatant was removed. The pellet was resuspended in an equal volume of Buffer A, vortexed to resuspend the pellet, and centrifuged again at the same speed and duration to remove any remaining soluble material. The supernatant was discarded and replaced with half the volume of 8 M Guanidine-HCl, 25 mM Tris-HCl pH 8 and 100 mM DTT (Buffer B), vortexed to resuspend, and centrifuged at 14,000g for 20 min. The supernatant containing any previously insoluble proteins was filtered using a 0.45- μ m filter. One- or 2-mL injections were loaded into an NGC chromatography system (Biorad) and eluted through an ENrich SEC 650 10×300 column (Biorad) equilibrated in 6 M Guanidine-HCl, 12.5 mM Tris-HCl pH 8, 5 mM DTT and 1 mM EDTA (Buffer C). Under these conditions, rPrP 94–178 eluted in both monomeric and multimeric forms; monomeric species eluted as a major peak (Extended Data Fig. 1). Fractions containing monomeric rPrP 94–178 were pooled and buffer exchanged into 8 M urea using a Duoflow chromatography system (Biorad) and a HiTrap Desalting column (GE Healthcare). Protein was concentrated to 1.6–2.4 mg mL⁻¹ and flash frozen for storage at –80°C until fibril formation.

Mass spectrometry of rPrP 94–178. Purified rPrP 94–178 in 8 M urea was diluted to 20 μ M in 50% methanol and 0.1% formic acid. Mass spectrometry was carried out using a Q-Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo). The most abundant peaks correspond to ions of rPrP 94–178 with an extracted exact mass of 9665.49 Da, within approximately 1 p.p.m. of the expected exact mass of 9,665.48 Da (Extended Data Fig. 1).

Growth of rPrP 94–178 fibrils. Aliquots of monomeric rPrP 94–178 were diluted to 0.8–1.2 mg mL⁻¹ in 1–4 M urea and 50 mM Sodium Citrate pH 4 (growth buffer) and added to a 96-well plate. Samples were agitated for 1–3 d by acoustic resonance mixing³⁶ at 37°C using a custom-built 96-well plate shaker set to 38 Hz. No change in fibril morphology was observed as a function of incubation time. However, each fibril preparation was highly polymorphic, and preparations varied in polymorph yield and distribution. Once fibrils formed, they appeared stable for months with no perceivable dissolution or evolution by negative stain electron microscopy. Fibril suspensions were collected and centrifuged at 16,160g for 5–7 min to pellet aggregated material and washed with water twice before resuspending in a volume of water that allowed ideal concentration for cryo-EM.

Crystallization, data collection and structure determination of PrP 113–118

and PrP 119–124. Both segments were purchased at a purity of 99% from CS BIO. Screening for crystallization conditions was achieved by the hanging drop method using ‘Crystal Screen’ and ‘Index Screen’ from Hampton Research. Idealized crystal growth conditions were then optimized. Crystals of PrP 113–118 were grown from 1.92 M sodium malonate pH 4.0 at a concentration of 35 mg mL⁻¹. Crystals of PrP 119–124 were grown from a condition of 2.4 M sodium malonate pH 7.0 at a concentration of 10 mg mL⁻¹.

Diffraction data were collected at the microdiffraction beamline ID13 at the European Synchrotron Radiation Facility (ESRF), where crystals were mounted directly onto the ends of pulled glass capillaries. Because of the tight packing between segments in the crystals, cryoprotectant was unnecessary. Crystals were cooled to –180°C during data collection. In both cases, data were collected using 5° wedges.

In most cases, diffraction data represented patterns from epitaxially twinned crystals that could not be separated during mounting. Despite the multiple diffraction patterns found on each frame, the program XDS³⁷, which uses an automated algorithm for picking spots from multiple frames, was successfully used to index the images. The program SCALEPACK³⁸ was used for scaling, merging and reducing data from multiple crystals. The programs ‘xds2mos’ written by Richard Kahn (Grenoble; <http://www.ccp4.ac.uk/newsletters/newsletter40.pdf>), and the utility ROTGEN, distributed with the software package MOSFLM³⁹, were used to convert the crystal orientation and cell parameters from XDS into DENZO format. Re-indexing using the XDS parameters and subsequent data reduction was performed in DENZO and SCALEPACK from the HKL suite⁴⁰. The merged and scaled data were imported into the CCP4 format with programs from the CCP4 program suite.

Structures were phased by molecular replacement using the program PHASER⁴¹ with an idealized poly-alanine beta strand as a probe. The program

Coot⁴² was used for model building, and refinement was carried out using REFMAC⁴³. Data collection and refinement statistics can be found in Table 2 and structures shown in Extended Data Fig. 6.

Negative stain electron microscopy of rPrP 94–178 fibrils. Imaging of fibrils embedded in stain was performed by applying 3 μ L of washed sample to glow-discharged formvar/carbon 200 mesh copper grids (Ted Pella) and incubated at room temperature for 2 min, followed by the addition of 3 μ L of 2% uranyl acetate and additional 2-min incubation before liquid was wicked away and the grid was left to dry for 2–3 min. Images were recorded on a Tecnai 12 electron microscope (Extended Data Fig. 1) (FEI).

Partial protease digestion and nephelometry monitoring of rPrP 94–178 fibrils.

Fibrils in growth buffer were incubated with a 1:10 molar ratio of freshly dissolved proteinase K (GoldBio) to rPrP 94–178 monomer for 24 h at room temperature without agitation⁴⁴. After the incubation period, any remaining insoluble material was pelleted by centrifugation at 16,160g for 5–7 min and the supernatant was removed. The pellet was resuspended in water (for negative-stain imaging) or 2% SDS (for cryo-EM).

For a measure of insoluble character in response to proteinase K treatment, dissolution experiments were monitored by nephelometry (Extended Data Fig. 2). A slurry of 100 μ L of rPrP 94–178 fibrils in growth buffer with the addition of 1:10 molar ratio of proteinase K: rPrP 94–178 monomer (10 μ M ProK: 100 μ M rPrP 94–178 monomer), growth buffer with the same amount of proteinase K, growth buffer only or water were added to a sealed 96-well plate and loaded into a NEPHELOstar Plus (BMG Labtech). Readings in nephelometry units (NTUs) were recorded at 37°C for 0.1 s every 10 min for 24 h with a 2.5-mm beam at 12% intensity. Before each reading, the plate was shaken at 300 r.p.m. for 20 s.

Preparation of frozen-hydrated rPrP 94–178 fibrils for high-resolution

imaging. Following treatment of fibrils with proteinase K, insoluble material was resuspended in a solution one-fifth the original volume and containing 2% wt/vol SDS and sonicated in an ultrasonic bath (Fisher Scientific) for 10 min. Sonication for up to 20 min showed no observable effect on individual fibril morphology or size. However, sonication for intervals greater than or equal to 3 min modestly reduced fibril clumping; 1.8 μ L of proteinase K-treated rPrP 94–178 fibrils in 2% SDS were applied to each side of glow-discharged holey carbon grids (Quantifoil, R1.2/1.3 200 mesh Cu, Electron Microscopy Sciences) after 10 min of both sonication (Extended Data Fig. 1). Blotting and plunge freezing into liquid ethane was performed on a FEI Vitrobot Mark IV using a blot force of 1 and a blot time of 7 s with no wait or drain time.

High-resolution imaging of frozen-hydrated rPrP 94–178 fibrils. High-resolution images were collected on a Titan Krios microscope (FED) with an accelerating voltage of 300 kV and \times 130,000 magnification using a Gatan K2 Summit direct electron detector attached to a Gatan image filter (GIF) in super-resolution counting mode with a GIF slit of 20 eV. We collected 30 movie frames of 200 ms each for a total exposure time of 6 s and a total accumulated dose of 36 e⁻ μ m² at a calibrated pixel size of 1.07 Å (0.535 Å for super-resolution). Automated data collection was carried out using the Legion software package⁴⁵ with a target defocus value of –2.0 μ m, to record a total of 4,341 micrographs.

Pre-processing of rPrP 94–178 fibril cryo-EM images. Patchwise motion correction using 5×5 patches per micrograph, and frame alignment were carried out using MotionCorr2 (ref. 69⁷⁰) integrated in RELION 3.0 (ref. 71). The resulting dose-weighted micrographs were used to determine defocus values for each image using Gctf⁷². Micrographs that (1) did not have fibrils in the field of view, (2) had clumps of fibrils with too much overlap to select even small segments, (3) were too far from focus or (4) had ice that would limit extraction of high-resolution information were not processed further. Particles were manually picked as helical segments using RELION 3.0 from 2,542 out of 4,341 micrographs.

Two-dimensional classification of rPrP 94–178 fibril cryo-EM images. All analysis was carried out with extracted particles that were down-sampled by a factor of two, such that the pixel size during 2D classification through the reported map was 1.07 Å. There were 217,988 segments that were initially ctf (contrast transfer function) phase-flipped and extracted using a box size of 267.5 Å (250 pixels) and an interbox distance of 28.8 Å or 10.8% of the box size. In RELION, this interbox distance was achieved by specifying 6 asymmetric units where each asymmetric unit was 4.8 Å. Several rounds of 2D classification using a 250-pixel, or a larger 400-pixel, box were carried out to remove particles in less defined classes (Extended Data Fig. 3). The larger box size allowed for more of the helical twist to be visible within a single box, which helped prevent misalignment based on only the 4.8 Å feature. After all rounds of 2D classification, 113,320 particles with a 400-pixel box were used for 3D classification.

Fibril crossover distance was estimated by stitching together 2D classes with a box size of 400 pixels (428 Å) where there was noticeable overlap between class averages. A box size nearly the length of the estimated crossover distance (1,000 0.535-Å pixels rescaled to 250 pixels, or 535 Å) was used to extract and classify

the same particles, where agreement in morphology and the estimated crossover distance was achieved between both methods (Extended Data Fig. 3).

The images also contained filaments belonging to minor populations that did not have an apparent regular helical twist (Extended Data Fig. 4). These filaments were manually picked, extracted and classified separately using a 1,000-pixel box rescaled to 250 pixels (535 Å) and a 475-Å mask to accommodate thicker filament morphologies. There were 75,860 extracted segments that fell into 2D classes of visibly defined morphology (Extended Data Fig. 4c), of which 6,007 segments had a thick ribbon appearance. The remaining narrower particles underwent additional 2D classification using a smaller box size of 500 pixels rescaled to 250 pixels (267.5 Å). Poorly defined particles were removed, and additional rounds of 2D classification were carried out until the resolution in the best-defined classes ceased to improve. This resulted in 47,479 particles falling into the best classes assessed by a combination of RELION's estimated resolution and visual assessment (Extended Data Fig. 4d)⁷¹. These classes were not deemed sufficient for 3D structure determination as part of this work.

3D helical reconstruction of rPrP 94–178 fibrils. A Gaussian ellipsoid was generated as an initial reference model for 3D classification of the main morphology using 3 classes and a helical z value set to 30% of the box size. An initial estimate of the helical rise observed in 2D classes of 4.8 Å was used and helical symmetry in C_1 was enforced. An initial estimate of the helical twist was calculated from the estimated crossover distance of ~630 Å. All 3 classes were of identical ultrastructure, but only 76,246 particles in the highest-resolution class were moved forward for optimization of the twist and rise during a subsequent 3D classification run. The twist and rise were each refined individually, starting with the twist, then together until convergence. A staggering of strands at the interface between two identical protofilaments was observed, and C_1 symmetry with a z_1 screw was enforced. Final values of 2.4 Å helical rise and 179.3° helical twist for a left-handed filament were used for subsequent 3D auto-refinement where two half-maps were independently refined.

To prevent a bimodal tilt distribution during 3D classification and auto-refinement (maxima near 75° and 105°), rather than the expected mono-modal tilt distribution about 90° for filaments preserved in a thin layer of vitreous ice⁷², a Gaussian ellipsoid was used as a reference in the first round of 3D auto-refinement. The flags `-helical_keep_tilt_prior_fixed`, `-limit_tilt` 70, and `sigma_tilt` and `psi` values of 1 were used with a starting `healpix` order of 5—or 0.9375 degrees. A subsequent round of 3D auto-refinement was carried out using the previous auto-refinement reconstruction filtered to 7 Å as a reference. The orientations of the particles from the previous run were also used as the starting point for particle orientations. This 3D auto-refinement was carried out without the flag `-limit_tilt` but still kept the helical tilt prior fixed at 90° and started with an angular sampling using `healpix` order 5. After a few iterations, the `sigma_tilt` and `psi` were expanded to 3, and a smooth distribution about 90° was achieved with only very few particles with tilt angles past 75° or 105°. It is worth noting that reconstruction from particles with tilts distributed bimodally resulted in a reconstruction that appeared similar in cross-section but had substantial differences in the perceived staggering of beta strands within each protofilament.

After final 3D auto-refinement, helical symmetry in real space was imposed using the central 10% of the box to ensure uniform resolution throughout the length of the fiber. Masking and post-processing were carried out in RELION 3.0, and sharpening was performed in PHENIX using `phenix.auto_sharpen` to 3.4 Å and a B factor of -139.98. Overall resolution estimates were calculated on the basis of the Fourier shell correlation (FSC) coefficient of both 0.143 and 0.5 between the independently refined half-maps (Extended Data Fig. 5)⁷¹.

Helical rise multiples of 4.8 Å up to 4 repeats (19.2 Å) and their corresponding helical twist values for C_1 symmetry were also used to test for repeating structures encompassing multiple rungs, but no repeating patterns were detected.

Model building and refinement for rPrP 94–178 fibrils. A cross-section of the map normal to the fibril axis was used for initial sequence assignment, where large side chains and the tight protofilament interface were used as guides for building the initial model in Coot⁷³. Using that initial model, five chains per protofilament, for a total of ten identical chains, were generated and placed in a central region of the cryo-EM density map. Manual refinement was performed in Coot, where changes were made to a single chain and propagated to all ten chains to ensure model compatibility with neighboring chains. Iterations of manual adjustments in Coot and refinement in PHENIX were employed to arrive at a final model that optimally balanced geometric constraints, Ramachandran angles, map fit and clashes. In PHENIX⁷⁴, all ten chains were also refined identically as in Coot.

The density for residue 126 left an ambiguity in the map where a choice to build the backbone up the fiber axis or down the fiber axis would have to be made. In order to address this ambiguity, we built and refined both models against the map and found the chosen direction to be substantially more favorable for Ramachandran angles, clashes and fit to the density.

Model building and refinement were carried out independently on maps with either right-handed⁷⁵ or left-handed beta sheet twists to validate the assignment of handedness. Each model was subjected to identical refinement procedures and resulted in very different profiles for Ramachandran angles with the left-handed

reconstruction being substantially more favorable. All further analysis was carried out using the left-handed model.

Analysis of buried surface area (S_b) in rPrP^{94–178} interfaces. S_b was computed as an average of the buried surface area per chain in a 20-chain assembly representing the fibril using PyMol⁷⁶. It was calculated as the difference between the sum of the solvent accessible surface area of the two sheets and the solvent accessible surface area of the entire complex, divided by the total number of strands in both sheets.

Comparison of rPrP 94–178 paired helical filament structure to prion hexapeptide zipper structures. Model building for rPrP 94–178 was performed without previous knowledge or consideration for the described peptide crystal structures. Sequence alignment and structural superpositions of both an individual chain and the crystalline arrangement of six-residue prion peptide structures were performed in PyMOL⁷⁶ (Extended Data Fig. 6). This alignment produced an all-atom root-mean-square deviation (RMSD) between a single chain in the peptide crystal structures and the corresponding residues in the fibril structure, which we used to assess similarity between the structures (Supplementary Table 1). For peptide structures with alternate conformations, the conformation for comparison was chosen by `pdcbcr` (CCP4) based on the highest occupancy.

Threading of sequences onto rPrP^{94–178} and their energetic evaluation using PyRosetta. PyRosetta⁷⁷ was used to evaluate the fit of a specific sequence onto the backbone built into density for the structure of rPrP^{94–178} for a series of 40-residue sequences. To prepare a structure for threading, an ensemble representing 20 chains of the structure of rPrP^{94–178} (10 chains per protofilament) was fast-relaxed in PyRosetta while keeping the 3 chains at the top and bottom of the assembly fixed. The resulting model yielded an energetically favorable overall score (Supplementary Table 1) reported in Rosetta energy units (REU).

To assess the impact of sequence variants on the energetic favorability of the rPrP^{94–178} polymorph, each relevant site was independently mutated and followed by side chain repacking without backbone movement. A second trial was performed by first using a fast-relax protocol in PyRosetta (Extended Data Fig. 8 and Supplementary Table 1) and followed by side chain repacking. This allowed the assessment of both the compatibility of mutations with our density-informed model and which small movements may still lead to a favorable structure. The energy of the resulting models was evaluated as the average of six independent trials (Supplementary Table 1, repack only) or three independent trials (Supplementary Table 1, fast-relax followed by repacking). This procedure did not enforce the symmetry present in rPrP^{94–178} but was restricted to small backbone movements. It was further constrained by allowing the fast-relax movements to only the middle strands of the fibril model. These restrictions were put in place to reduce the influence fibril ends might have on the overall structure. RMSD calculations were performed in PyMol over ten chains to obtain an average RMSD over the whole assembly generated by PyRosetta. Comparisons were made between fast-relaxed wild-type and mutant sequences and the density-informed model before fast relaxing (Supplementary Table 1).

Calculation of solvation free energy for rPrP^{94–178}. Our calculation of free energy is an adaptation of the solvation free energy described previously^{78,79}, in which the energy is calculated as the sum of products of the area buried of each atom and its corresponding atomic solvation parameter (ASP). ASPs were taken from previous work⁷⁸. Area buried was calculated as the difference in solvent accessible surface area (SASA) of the reference state (that is unfolded state) and the folded state. The SASA of residue i of the unfolded state was approximated as the SASA of residue i in the folded structure after removing all other atoms except the main chain atoms of residues $i - 1$ and $i + 1$.

The SASA of the folded state was measured for each atom in the context of all amyloid fibril atoms. Fibril coordinates were symmetrically extended by three to five chains on either side of the molecule to ensure the energetic calculations were representative of the majority of molecules in a fibril, rather than a fibril end. To account for energetic stabilization of main chain hydrogen bonds, the ASP for backbone N/O elements was reassigned from -9 to 0 cal per mol per Å² if they participated in a hydrogen bond. Similarly, if an asparagine or glutamine side chain participated in a polar ladder (two hydrogen bonds per amide) and was shielded from solvent (SASAfolded < 5 Å²), the ASPs of the side chain N and O elements were reassigned from -9 to 0. Lastly, the ASP of ionizable atoms (for example Asp, Glu, Lys, His, Arg, N-terminal amine or C-terminal carboxylate) were assigned the charged value (-37 or -38) unless the atoms participated in a buried ion pair, defined as a pair of complementary ionizable atoms within 4.5 Å distance of each other, each with SASAfolded < 50 Å². In that case, the ASP of the ion pair was reassigned to -9 × (distance - 2.8 Å) / 2.8 Å². Side-chain conformational entropy terms adapted from Koehl and Delarue⁸⁰ were added to the energy values obtained above⁸⁰. The entropy terms were scaled by the percentage of side-chain surface area buried in the assembly. Based on these calculations, the average ΔG° per residue for rPrP^{94–178} is -0.74 kcal mol⁻¹, indicating an overall favorable energy for assembly (Fig. 3).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

EM maps and atomic coordinates for rPrP 106–145 have been deposited in the EMDB and wwPDB as EMD-20900 and PDB 6UUR, respectively; atomic coordinates and structure factors have been deposited in the wwPDB for PrP 113–118 (PDB 6PQ5) and PrP^{119–124} (PDB 6PQA).

References

59. Gao, Y., Tran, P., Petkovic-Duran, K., Swallow, T. & Zhu, Y. Acoustic micromixing increases antibody-antigen binding in immunoassays. *Biomed. Microdevices* **17**, 79 (2015).
60. Nagapudi, K., Umanzor, E. Y. & Masui, C. High-throughput screening and scale-up of cocrystals using resonant acoustic mixing. *Int. J. Pharm.* **521**, 337–345 (2017).
61. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
62. Otwinowski, Z., Minor, W., Borek, D. & Cymborowski, M. in *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules* 2nd edn. (eds. Rossman, M. G. et al.) Ch. 11.4 (Wiley, 2012).
63. Leslie, A. G. W. & Powell, H. R. in *Evolving Methods for Macromolecular Crystallography* (eds. Read, R. J. & Sussman, J. L.) 41–51 (Springer Netherlands, 2007).
64. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
66. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
67. Hilz, H., Wiegiers, U. & Adamietz, P. Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of 'masked' proteins. *Eur. J. Biochem.* **56**, 103–108 (1975).
68. Carragher, B. et al. Legikon: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* **132**, 33–45 (2000).
69. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
70. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
71. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
72. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
73. He, S. *Helical Reconstruction in RELION*. Doctoral thesis, Univ. of Cambridge (2018).
74. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
75. Liberta, F. et al. Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nat. Commun.* **10**, 1104 (2019).
76. Delano, W. *The PyMOL Molecular Graphics System* (Schrödinger LLC).
77. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinform. Oxf. Engl* **26**, 689–691 (2010).
78. Eisenberg, D. & McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).
79. Eisenberg, D., Wesson, M. & Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chemica Scr.* **29A**, 217–221 (1989).
80. 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).

Acknowledgements

We thank D. Cascio (UCLA), H. McFarlane (UCLA) and C. Sigurdson (UCSD). This work is supported by National Science Foundation (NSF) Grants DMR-1548924 and DBI-1338135, DOE Grant DE-FC02-02ER63421 and National Institutes of Health (NIH) grants R35 GM128867, AG054022 and 1U24GM116792, as well as NIH instrumentation grants 1S10OD016387-01, 1S10RR23057 and 1S10OD018111, which support our use of instruments at the Electron Imaging Center for NanoMachines and CNSI at UCLA. C.G. was funded by the Ruth L. Kirschstein NRSA GM007185 (NIH T32 Cellular and Molecular Biology Training Grant, UCLA) and is now funded by the Ruth L. Kirschstein Predoctoral Individual NRSA, 1F31 AI143368. J.A.R. is supported as a Searle Scholar, a Pew Scholar and a Beckman Young Investigator. D.S.E. is supported by the Howard Hughes Medical Institute.

Author contributions

C.G. and R.B. produced, evaluated and optimized recombinant PrP fibril preparations. C.G. and P.G. performed electron microscopy. M.A. and M.R.S. performed X-ray structure determination of PrP segments. C.G. and C.W.S. selected particles for analysis. C.G., P.G. and M.G.J. performed fibril reconstruction. C.G. and M.R.S. built the fibril model. C.G., P.G., M.R.S., M.G.J., C.W.S., R.B., M.A., Z.H.Z., D.S.E. and J.A.R. critically analyzed and provided feedback on data. C.G. and J.A.R. wrote the manuscript, with input from all authors.

Competing interests

D.S.E. is SAB chair and an equity holder in ADRx.

Additional information

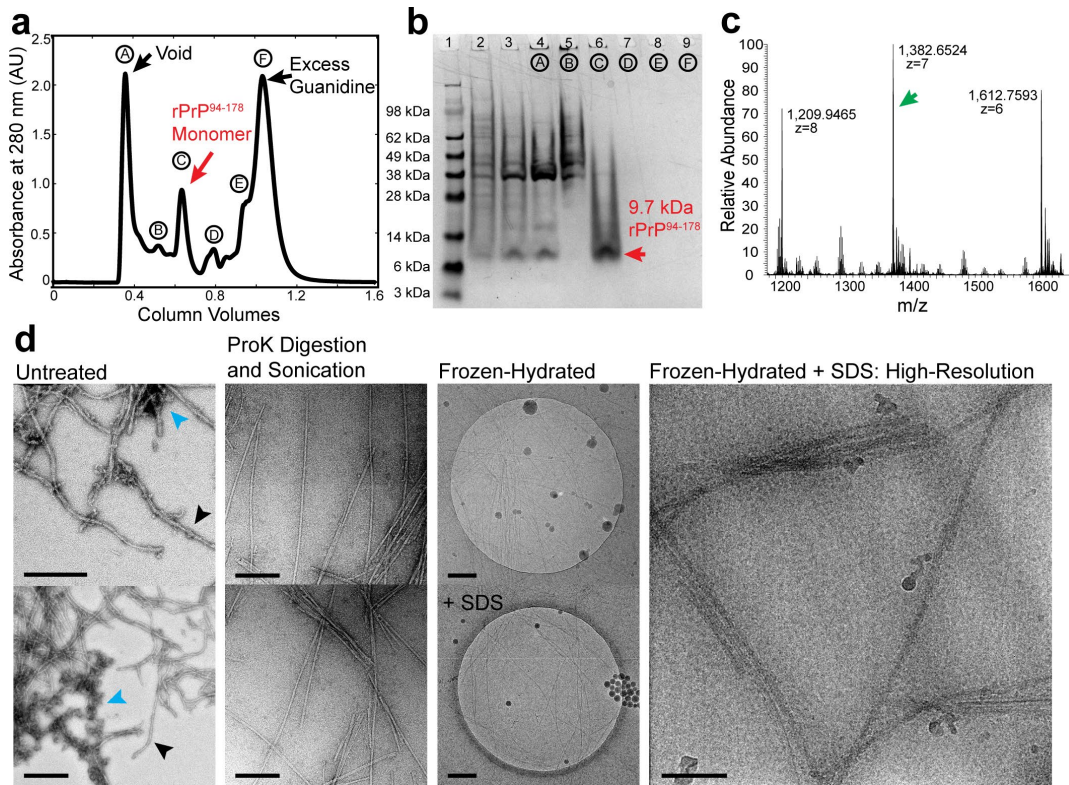
Extended data is available for this paper at <https://doi.org/10.1038/s41594-020-0403-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-020-0403-y>.

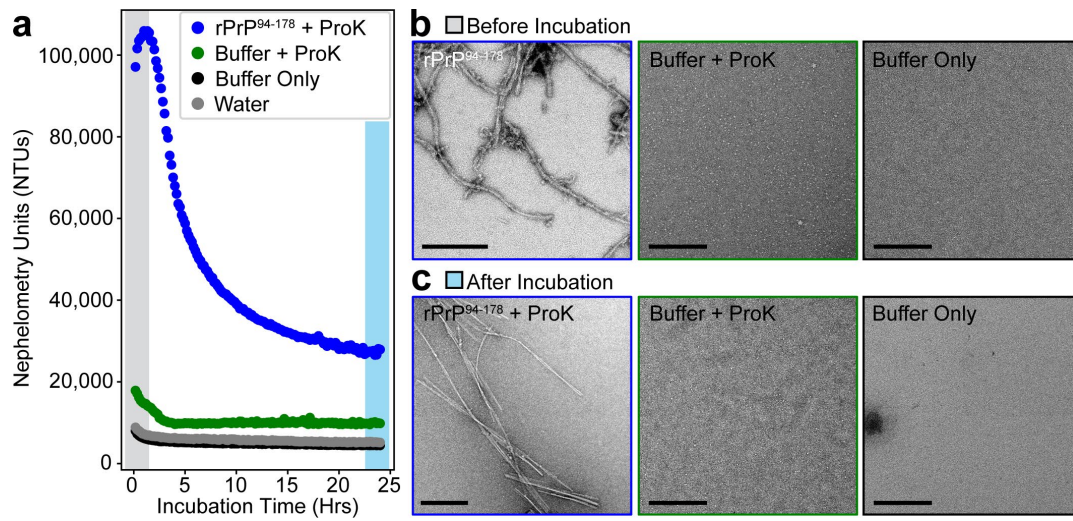
Correspondence and requests for materials should be addressed to J.A.R.

Peer review information Inés Chen was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

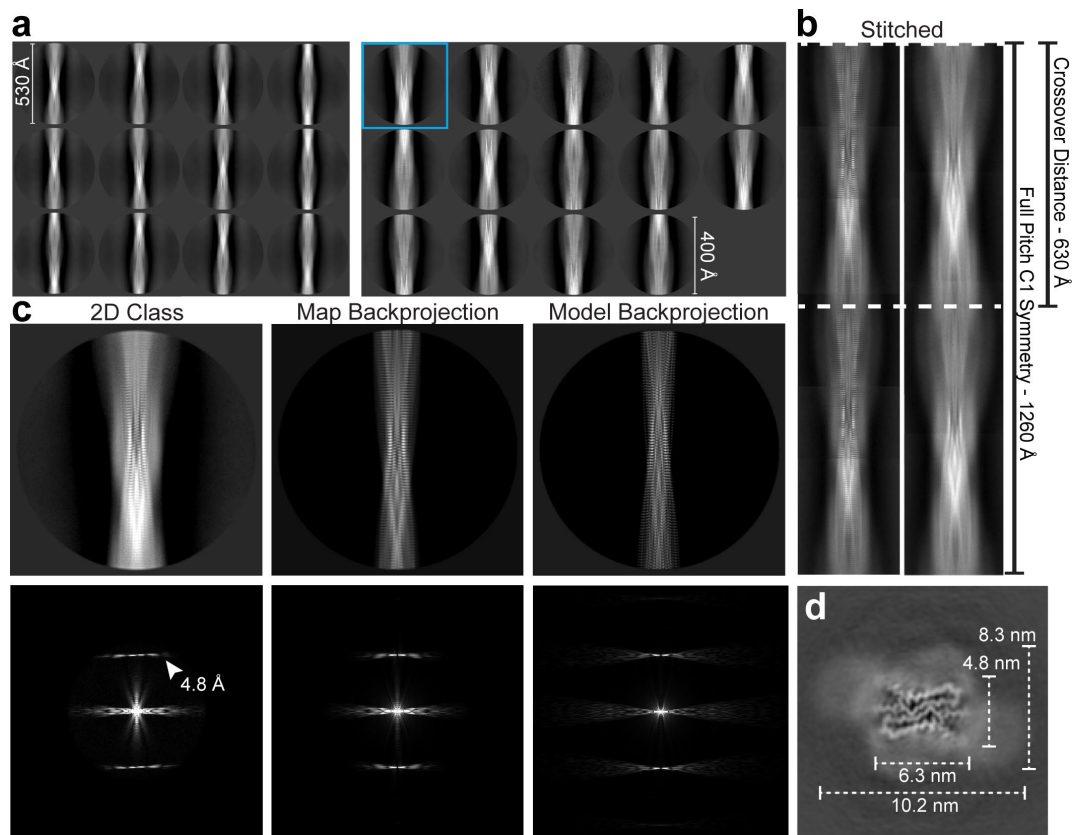
Reprints and permissions information is available at www.nature.com/reprints.



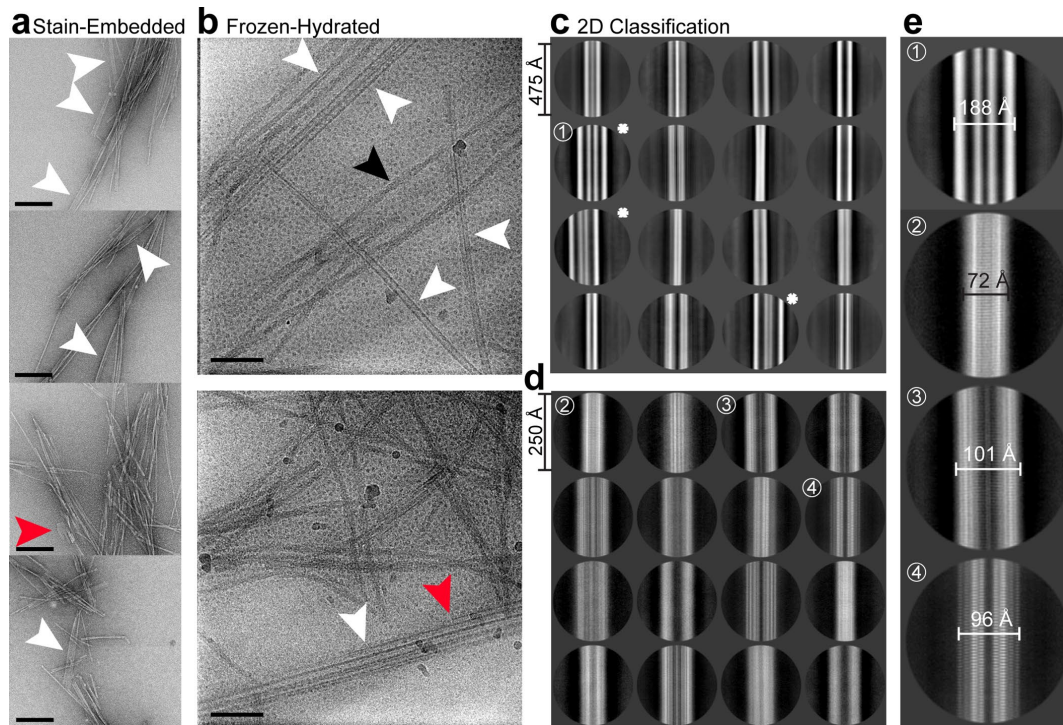
Extended Data Fig. 1 | Production, isolation and characterization of rPrP⁹⁴⁻¹⁷⁸ fibrils. **a**, Purification of untagged human PrP⁹⁴⁻¹⁷⁸. Samples collected during size exclusion chromatography of PrP⁹⁴⁻¹⁷⁸ were run on a 4–12 % SDS-PAGE gel corresponding to labeled peaks **b**. Gel lanes left to right: 1. Pre-stained protein ladder (Thermo), 2. Post-induction fraction, 3. Solubilized material from inclusion bodies, 4. Void fraction (A) from (a), 5. Peak (B) from (a), 6. Peak (C) from (a) containing 9.7 kDa monomeric rPrP⁹⁴⁻¹⁷⁸, 7–9. Peaks D–F from (a) containing excess guanidine or other UV active small eluates. **c**, Mass spectrum showing most abundant peaks correspond to ions with an extracted molecular weight that matches rPrP⁹⁴⁻¹⁷⁸. **d**, Representative micrographs of a heterogeneous mix of untreated fibrils (left) in Growth Buffer; promising filaments (black arrows) and disordered, clumped, or amorphous material (blue arrows). Adjacent images are of filaments treated with 1:10 molar ratio proteinase K:rPrP⁹⁴⁻¹⁷⁸ monomer and bath sonicated for 10 minutes. Proteinase K-treated and sonicated filaments exchanged into water in a frozen-hydrated state (third column, top) or 2 % SDS (third column, bottom). Representative image of Proteinase K, sonication, and SDS treated filaments used for high-resolution imaging and reconstruction (right). Scale bars, 200 nm. Scale bar for high-resolution image (right) 50 nm.



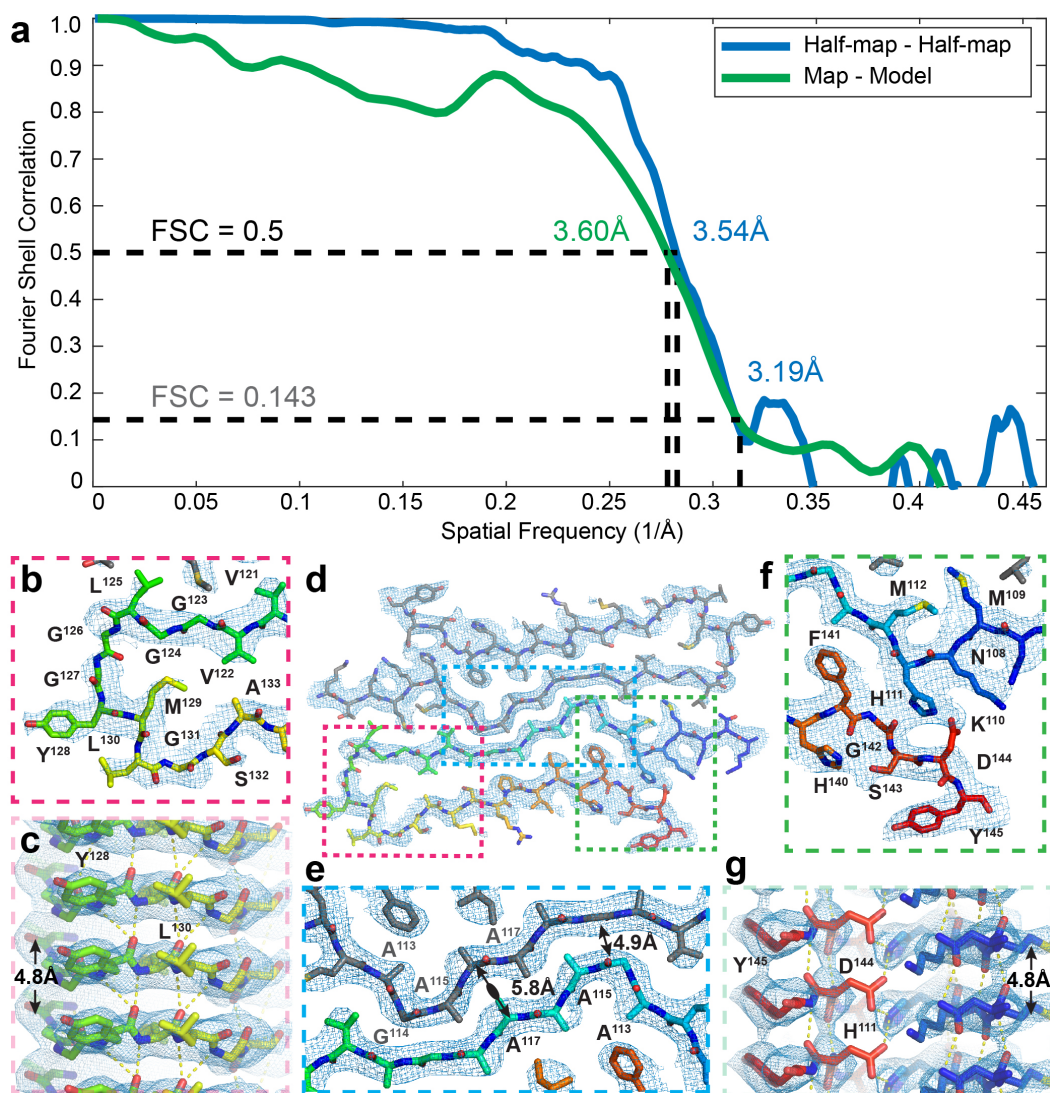
Extended Data Fig. 2 | Partial protease digestion of rPrP⁹⁴⁻¹⁷⁸ fibrils. **a**, Plots of incubation time versus nephelometry units as a measure of insoluble character in fibril suspensions treated with proteinase K compared to proteinase K only and buffer controls. **b**, Representative electron micrographs of each sample in panel (a) before proteinase K digestion at the start of the incubation period. This image of fibrils is the same as that shown in Extended Data Fig. 1d. **c**, Representative micrographs of each sample in Panel (a) after the 24-hour incubation period. All scale bars are 200 nm.



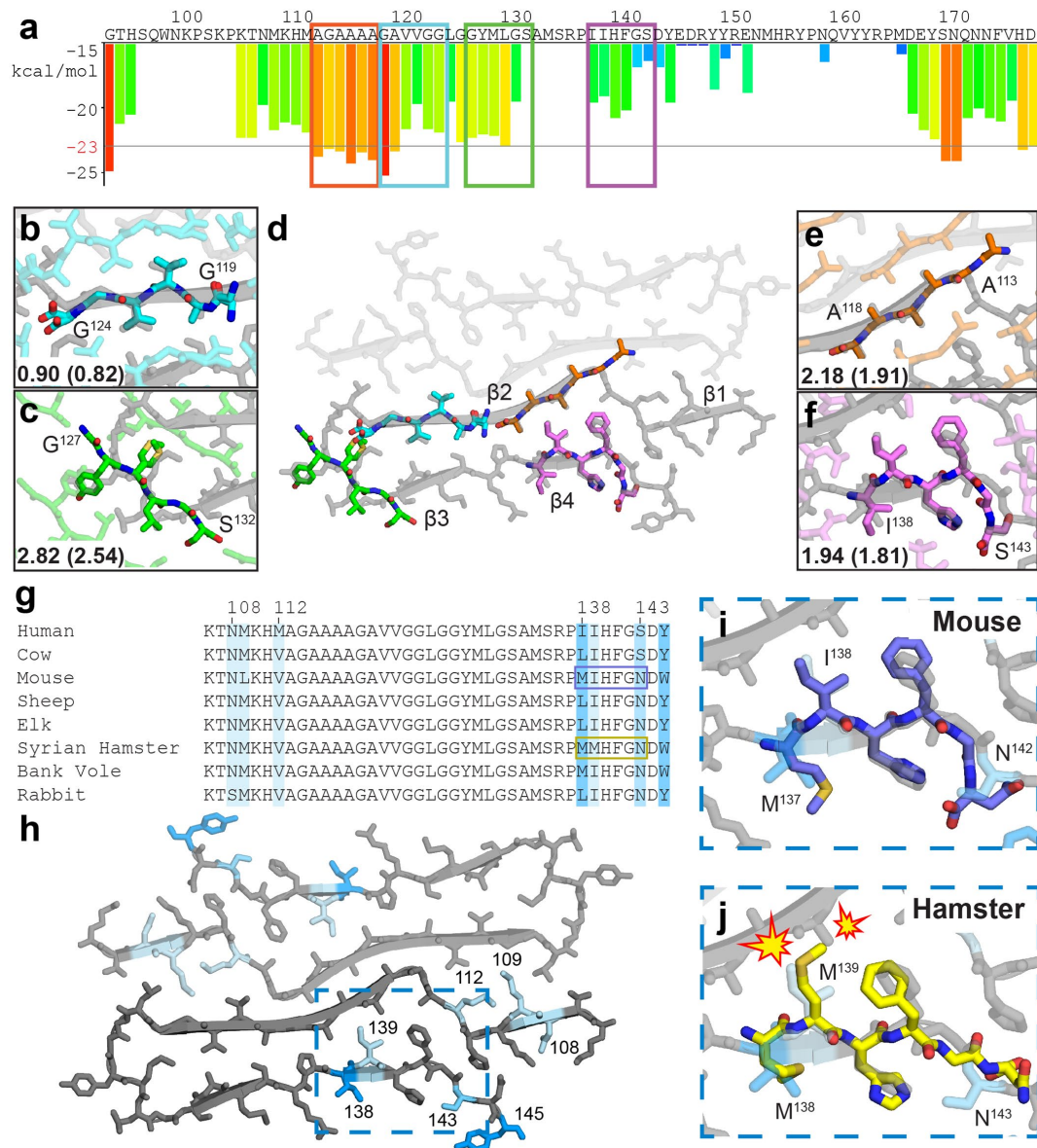
Extended Data Fig. 3 | Classification of major species of protease-resistant rPrP⁹⁴⁻¹⁷⁸ fibrils. **a**, Final 2D class averages from the major species comprising 71 % of defined segments using a 535 Å box with a 530 Å mask (left) or 428 Å (400 pixel) box with a 400 Å mask (right). **b**, A composite image (left) formed by stitching of 2D class averages with the small box shown in (a), agrees with a composite image (right) formed by class averages obtained using a box size encompassing a full crossover distance in (a). The crossover distance and full pitch are both marked. **c**, Comparison between a 2D class (enlarged view of boxed class in (a)), the map backprojection, and model backprojection with accompanying Fourier transforms below. **d**, Slice through the 3D density with dimensions of the ordered region and surrounding diffuse density noted.



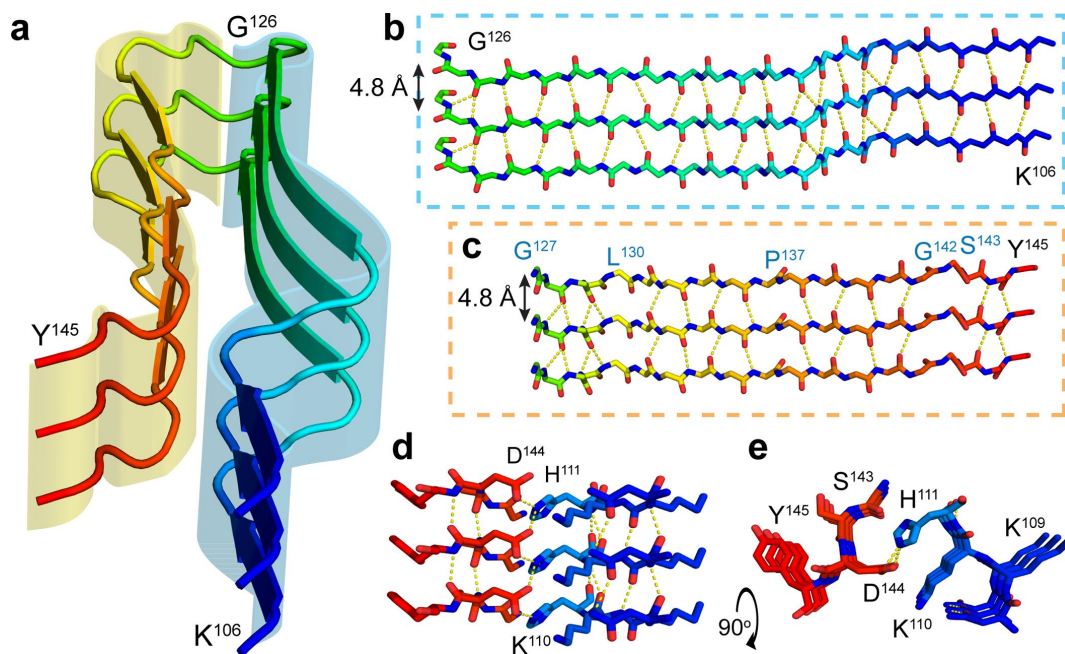
Extended Data Fig. 4 | 2D classification of minor morphological populations formed by rPrP⁹⁴⁻¹⁷⁸. **a**, Stain-embedded, Proteinase K and sonication treated fibrils show several minor populations of ribbon-like polymorphs (white arrows), including wide filaments with regions that abruptly stop (red arrows). These morphologies remain a minor species after vitrification in 2% SDS alongside the major twisted species (single example black arrow) **(b)** A 2D classification round selecting for minor species results in 2D classes that make up 29% (75,860 segments) of total particles sorted into defined classes **(c)** with 2.1% of the total segments across major and minor populations (~6000 particles) being sorted into classes that resemble thick ribbons with columns of alternating electron dense and poor material (stars). Additional 2D classification of segments that did not sort into thick ribbon classes **(d)** revealed several classes containing high-resolution information, in some classes even revealing 4.8 Å strand separation. The 16 best looking classes are shown and make up 62% (47,479 segments) of the particles shown in **(c)**. 3D reconstruction of these filaments was unsuccessful. A magnified view of select classes from **(c)** and **(d)** show the range of fibril widths observed **(e)** Scalebar in **(a)** 200 nm, in **(b)** 50 nm.



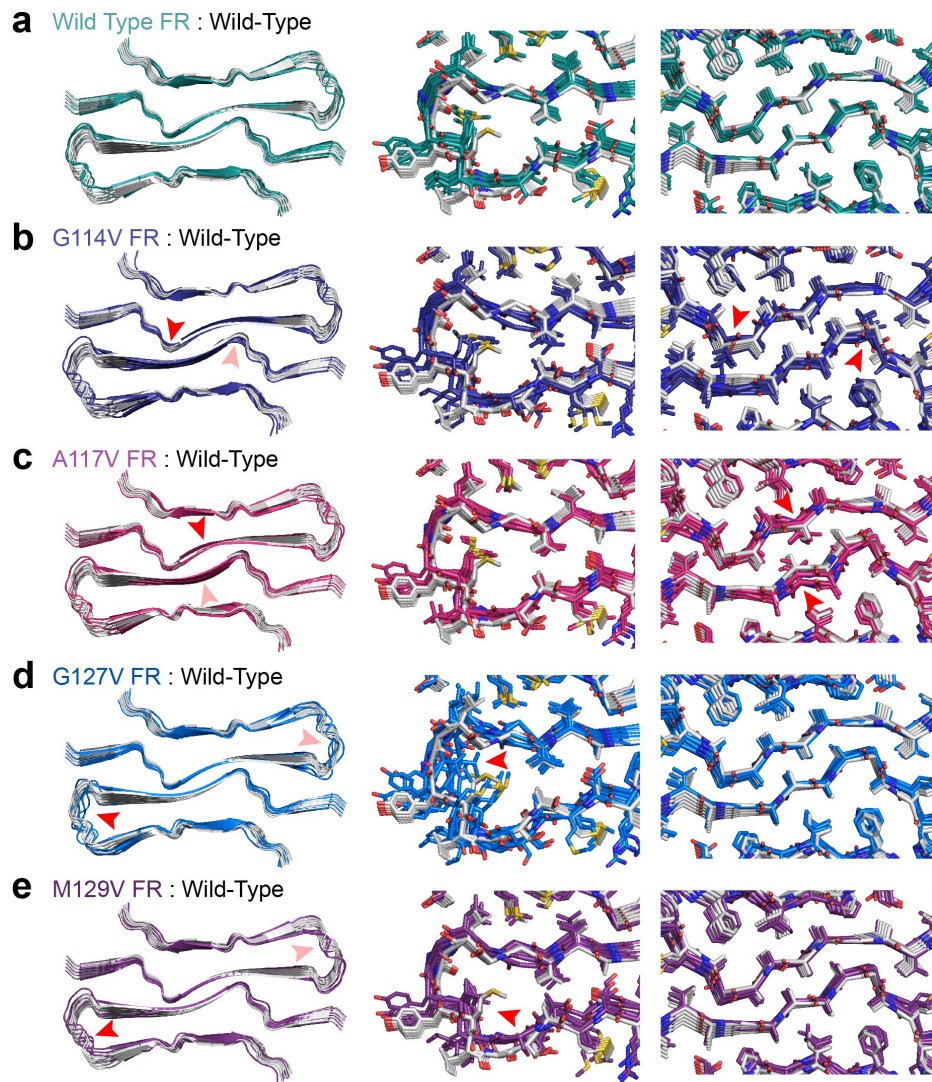
Extended Data Fig. 5 | Agreement of rPrP¹⁰⁶⁻¹⁴⁵ model with core density in rPrP^{Res}. **a**, Fourier Shell Correlation (FSC) between half-maps (blue) and the map and model (green) with resolutions at FSC = 0.5 (black dotted line) and 0.143 (gray dotted line) noted. **b**, Agreement of turn region with density looking down the fibril axis and **(c)** a side view of the same region shows clear strand separation. **d**, Overall fit of model into density. **e**, Magnified view of protofilament interface showing spacing between backbones at the center of the interface is at the most tightly spaced region between G¹¹⁴ of one chain and G¹¹⁹ of its mate. **f**, Magnified view of linchpin region with inferred salt bridge between residues H¹¹¹ and D¹⁴⁴ that seal off the interior of each chain and a side view **(g)** of the same region showing 4.8 Å separation between strands.



Extended Data Fig. 6 | Structural agreement of hexapeptide prion zipper structures to rPrP^{Res}. **a**, ZipperDB²⁶ profile of the segment encoding for rPrP^{Res}. Bars extending below the line represent hexapeptides predicted to form steric zippers. Peptides with crystal structures aligned to the fibril model are boxed. Two of these were previously published^{27,28}. Alignment of crystal structures ¹¹⁹GAVVGG¹²⁴ (**b**) ¹¹³AGAAA¹¹⁸ (**c**) ¹²⁷GYMLGS¹³² (**e**) and ¹³⁸IHFSGS¹⁴³ (**f**) with rPrP^{Res}. (**d**) Symmetry mates in plane, all atom, and backbone (parenteses) RMSD values against rPrP^{Res} are shown (**g**) Sequence alignment for species of common interest to prion disease. Darker blues correlates with more variability, loosely defined as more residue mismatches among the species compared, and mapping of these residues onto rPrP^{Res} **h**, Hexapeptide structure for muPrP¹³⁷⁻¹⁴² (mouse numbering) is compatible with rPrP^{Res} (**i**) while shPrP¹³⁸⁻¹⁴³ is not (**j**).



Extended Data Fig. 7 | Molecular contacts at the core of rPrP^{Res}. **a**, Cartoon representation of one protofilament with three stacked chains highlighting outer (yellow) and inner (blue) sheets. Stick representation of stacked inner (**b**) and outer (**c**) sheets with inferred hydrogen bonding networks based on refinement giving weight to standard beta sheet geometry, favored Ramachandran angles, and map to model fit. Residues where geometry deviated during refinement to break backbone hydrogen bonds are noted in blue text. **d**, Side view highlighting a favorable salt bridge between H¹¹¹ and D¹⁴⁴ and the same region rotated to show the interaction looking down the fiber axis (**e**).



Extended Data Fig. 8 | Alignment of fast-relaxed mutant sequences to rPrP^{Res}. **a**, Left: Alignment of eight chains of the model (gray) to the same number of chains of a fast-relaxed model produced in PyRosetta (green) along with magnified views with sidechains in the turn region (center) and the protofilament interface (right). **b**, Eight chains of fast relaxed G114V mutant (left) along with magnified view of the protofilament interface near the mutation site (red arrow). The same set of three views are also shown for A117V (**c**) G127V (**d**) and M129V (**e**) mutants or polymorphisms.

In the format provided by the authors and unedited.

Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core

Calina Glynn¹, Michael R. Sawaya^{1,2}, Peng Ge^{1,3}, Marcus Gallagher-Jones¹,
Connor W. Short¹, Ronquiajah Bowman¹, Marcin Apostol^{1,2,5}, Z. Hong Zhou^{3,4}, David S. Eisenberg^{1,2}
and Jose A. Rodriguez¹✉

¹Department of Chemistry and Biochemistry; UCLA-DOE Institute for Genomics and Proteomics; STROBE, NSF Science and Technology Center, University of California, Los Angeles, Los Angeles, CA, USA. ²Department of Biological Chemistry and Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA, USA. ³California NanoSystems Institute, University of California Los Angeles, Los Angeles, CA, USA. ⁴Department of Microbiology Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, CA, USA. ⁵Present address: ADRx, Thousand Oaks, CA, USA. ✉e-mail: jrodriguez@mbi.ucla.edu

Supplementary Table 1. Properties of PrP¹¹³AGAAAA¹¹⁸ and ¹¹⁹GAVVGG¹²⁴

Steric Zipper Features	¹¹³ AGAAAA ¹¹⁸	¹¹⁹ GAVVGG ¹²⁴
	(PDB 6PQ5)	(PDB 6PQA)
Surface Complementarity of Interface (Sc)	0.708	0.874
Surface Area Buried in Interface (Å ²)	130 ± 2	122 ± 4
Sheet-to-sheet Distance (Å)	5.23 ± 0.06	6.9 ± 0.9
Steric Zipper Class	Class 7 (anti-parallel)	Class 4 (parallel)

Supplementary Table 2. Rosetta energy and deviation from rPrP^{Res} of repacked mutants and peptides. For Rosetta energy score columns, either repacking of side chains only or fast relaxing to allow for backbone remodeling followed by sidechain repacking. For RMSD columns, the fast relaxed and repacked threaded mutant sequences or raw peptide structures were compared against the rPrP^{Res} structure without fast relaxing or repacking. For peptide structures with multiple sidechain conformations, only the highest occupancy conformation is used to calculate RMSD.

Sequence Aligned*	Rosetta Energy Score (REU)		RMSD (Atoms Aligned)		
	Repack Only	Fast Relax and Repack	CA	Main	All
Wild Type	-522.9 ± 12.9	-768.1 ± 5.4	0.787 (400)	0.780 (1600)	1.415 (2770)
G114V	10610.8 ± 38.7	-711.2 ± 0.9	0.894 (400)	0.903 (1600)	1.516 (2770)
A117V	1776.4 ± 12.1	-699.6 ± 18.0	0.848 (400)	0.843 (1600)	1.444 (2770)
G127V	6674.1 ± 11.5	-410.3 ± 22.2	0.866 (400)	0.864 (1600)	1.468 (2770)
M129V	-406.4 ± 14.6	-674.8 ± 8.9	0.824 (400)	0.834 (1600)	1.414 (2740)
¹¹³ AGAAAA ¹¹⁸	-	-	2.051 (6)	1.910 (24)	2.183 (29)
¹¹⁹ GAVVGG ¹²⁴	-	-	0.572 (6)	0.822 (24)	0.899 (31)
¹²⁷ GYMLGS ¹³²	-	-	2.129 (6)	2.541 (24)	2.820 (42)
¹³⁸ IHFHGS ¹⁴³	-	-	1.189 (6)	1.813 (24)	1.938 (47)

*The full 40 residue sequence with the specified mutation was used for threading and fast relaxing. Experimentally determined peptide structures were aligned directly to the model with no fast relaxing.

Supplementary Table 3. Solvation energy of rPrP^{Res} compared to a subset of other amyloid fibril structures.

Fibril	Method	# of Residues in Calculation	ΔG° (kcal/mol /chain)	ΔG° (kcal/mol /residue)	PDB ID
Amyloid-β 1-42	ssNMR	26	-19.8	-0.76	2BEG
rPrP ^{Res}	cryo-EM	40	-29.6	-0.74	6UUR
AD Tau PHF	cryo-EM	73	-31.4	-0.43	5O3L
Alpha synuclein (twister)	cryo-EM	41	-12.4	-0.30	6CU8
HET-s	ssNMR	61	-18.8	-0.30	2RNM
FUS	ssNMR	60	-12.2	-0.20	5W3N

Chapter 4

Cryo-EM Studies of Bank Vole and Elk Prion Fibrils

Cryo-EM Studies of Bank Vole and Elk Prion Fibrils

Calina Glynn, Jose A. Rodriguez

Abstract

Prion diseases are absolutely fatal neurodegenerative diseases that arise from genetic, sporadic, and acquired origins. There are over a dozen prion diseases in humans and animals, each associated with a unique set of clinical, pathological, biophysical, and transmission related characteristics. Within each disease, there are also many subtypes, called strains, that present with slight variations to each characteristic. These strains can be transmitted from one animal to another of the same species through ingestion of infectious aggregates of the prion protein, or PrP^{Sc}. Horizontal transmission is more readily achieved for some prion diseases, such as Chronic Wasting Disease in cervids, than others for reasons that remain a mystery. Prion disease transmission between members of different species is hampered by what is known as the species barrier, whereby the natively folded prion protein, PrP^C, cannot be readily templated by PrP^{Sc} of a differing protein sequence. The bank vole prion protein presents a unique case whereby bank vole PrP^C can be readily converted to infectious PrP^{Sc} capable of inducing disease both spontaneously and through templating by any prion strain of any mammalian origin. The structural basis for species barriers and prion transmission, including what uniquely allows bank vole prions to be converted by all isolates, has yet to be realized. Here I present the first views of fibrillar structures spontaneously adopted by elk and bank vole prion sequences that allow for visualization of the chains making up the cores of each fibril. Despite being formed in a similar chemical environment and differing at only five amino acids, these structures display some distinguishing features in structure and flexibility.

Introduction

The Bank Vole: A New Model Organism for Rapid Study of All Strains of Prion Disease

In the early days of prion disease research, before interspecies disease transmission was established, study of the original prion disease, scrapie, was confined to sheep and goats. Studies on transmission of scrapie were exceptionally long, with incubation periods for sheep spanning multiple years and goats being around two years.¹ Transmission to small rodents sped up prion disease research, with the time from inoculation with the scrapie agent to presentation of disease taking 8 months in golden hamsters, 7 months in rats, and 4 months in mice¹ after the original inoculum had undergone a few passages into the same animal species. In 1972, it was hypothesized that perhaps the incubation period for scrapie was influenced by the size, lifespan, and metabolic activity of the animal being passaged into¹, thus other active rodents were infected in search of a more rapid means to study scrapie prions. Additionally, rodents represented a potential reservoir for prion diseases, as rodents are ubiquitous, and transmission to and from rodents would be relevant to understanding interspecies prion disease transmission in the wild. Thus the interest in studying transmission of prion disease to and from rodents extended beyond their use as an animal model. The 1972 study found that field voles had the shortest incubation period at 3.5 months, comparable or perhaps slightly shorter than for mice, and that short incubation period was consistent regardless of whether the vole was inoculated with scrapie prions first passaged through mice, rats, or other voles¹. This was an unusual observation, given that most other animals that had been studied thus far displayed a species barrier, whereby the time for a new host species to present with symptoms would be longer if the scrapie strain was passaged through a different animal species than through the same species.

The species barrier further slowed and complicated study of prion diseases. Study of transmission in commonly used animal models gave variable timelines depending on the prion strain. Additionally, some animals, including commonly used mice, were not susceptible to infection at all, sometimes unexpectedly and for unknown reason². After 1974, genetic modification of animals became possible and the gene for the prion protein (PrP) from one animal species could be knocked out and replaced by the PrP gene of another species. This allowed for the realization that the species barrier could be overcome if only the PrP amino acid sequence of the source inoculum and new host matched.^{3,4} This allowed for prion diseases like scrapie to be studied in transgenic mice expressing sheep PrP without first having to overcome the sheep to mouse species barrier. However, overcoming the species barrier in this way required a new transgenic animal model to be created to study each strain of prion disease, of which there are many due to the various familial mutations associated with prion diseases in humans and many animal species susceptible to prion disease. For this reason, an animal model, or PrP sequence, that is universally susceptible to rapid infection by all prion isolates is highly desirable to speed up prion research and make uncommon genetic variants more accessible to study.

In 2005, another species of vole, the bank vole, was found to be susceptible to infection with multiple scrapie strains with comparable and short incubation periods without any genetic modification to the PrP gene⁵. In 2006, the same group found that nontransgenic bank voles were also able to acquire the human prion disease Creutzfeldt-Jakob disease (CJD) in a similar time frame to humanized mice⁶. The resulting aggregates were found to have protease-resistant PrP fragments that matched the size and relative abundance of the inoculum strain^{6,7} and cerebral deposition patterns⁸. Because of the many different sizes and variable protease-resistance patterns of prions found in different strains, these types of patterns are often used to help identify prion strains. This lent credence to the hypothesis that the inoculum strain was being faithfully reproduced in bank voles. Likewise, the *in vitro* assay, real-time quaking-induced conversion (RT-

QuIC), showed that recombinant bank vole PrP was able to propagate all prion strains in a way that did not change the size or protease-resistance of the resulting aggregates⁹. This work implied that not only does animal-derived bank vole PrP propagate all tested prion strains with fidelity, but the recombinant protein was also able to do so.

Bank Vole PrP Provides an Avenue for Study of Spontaneously Occurring Prion Disease

While there are many studies of both overexpressed and normal expression levels resulting in prion disease, they all utilize transgenic animals expressing PrP with familial mutations seen in rare genetic human prion diseases¹⁰⁻¹⁵. Spontaneous disease does not occur at high enough frequency in wild type animals, even with overexpression of PrP, to study spontaneous disease. In 2012, this changed when bank vole PrP was found to be particularly prone to misfolding into a transmissible infectious agent spontaneously without the need for a seed strain¹⁶. Transgenic mice overexpressing wild type bank vole PrP developed and succumbed to prion disease that was transmissible to transgenic and nontransgenic mice, which also developed and succumbed to prion disease¹⁶. This demonstrated that not only was the bank vole prion particularly prone to misfolding into an aggregate conformation, but that it was prone to spontaneously misfolding into a bona fide infectious conformation.

Structural Value of Bank Vole Prions

Infectious prions with a 100% attack rate can also be generated from recombinant bank vole PrP by protein misfolding shaking amplification (PMSA)^{17,18}. This illustrates that even recombinant bank vole prions are capable of retaining a hallmark feature of prions that distinguishes them from other amyloids: infectivity. Taken together, all these factors illustrate that bank vole PrP may provide unique opportunities for amplification and study of prion strain regardless of origin species

and rarity, study of spontaneous prion formation, and study of recombinant prions capable of infectivity. If bank vole prions truly retain the structure of the seed strain, they present a massive breakthrough in prion research whereby all strains can be amplified with fidelity and studied from biological and structural perspectives with the aim of understanding the seed strain. The range of structures bank vole PrP is capable of taking on spontaneously is also of great interest for understanding what may represent the ultimate prion: a spontaneously forming infectious structure without regards for substrate PrP origin, be it recombinant or animal derived.

An Emerging Prion Threat: Chronic Wasting Disease

A prion disease affecting cervids, Chronic Wasting Disease (CWD), was first noted in 1967 and first described in print in 1980¹⁹ however, it is still a relatively new prion disease compared to related diseases scrapie (described at least as early as 1853)²⁰, CJD (1921)²¹, and kuru (1957)²². Bovine Spongiform Encephalopathy (BSE) was first noticed in 1985 and first described in 1987²³, but received a much larger amount of attention due to its high prevalence in cattle in the meat industry and the subsequent transmission of BSE to humans in the form of variant CJD (vCJD). CWD received considerably less attention as efforts shifted out of urgency to BSE and prion disease transmission to humans. Elk and deer are consumed on a much smaller scale than beef, and CWD has not been shown to be transmissible to humans. With ~6000 CWD cases reported in the US from 1996-2010²⁴ compared to the 190,000 BSE cases reported in Europe during a similar time frame in 2011²⁵, it is no surprise that CWD has been a much less studied, and much less understood, prion disease.

Nevertheless, CWD poses an environmental and human health risk in the early stages of development. CWD is the only recognized prion disease of wild animals^{26,27} meaning there is massive potential for unchecked spread and undetected transmission across species lines. It is

not known whether CWD originated in captive or free-range animals, but the first cases were noted in Colorado^{19,26,28} followed by Wyoming²⁷, where the farms and research facilities in the two states regularly exchanged animals. By 2002, new surveillance for CWD found herds harboring infected animals in seven US states - Colorado, Wyoming, South Dakota, Nebraska, Oklahoma, Kansas, and New Mexico - and one Canadian province²⁷. But the surveillance came too late and CWD was brought to South Korea in 1997 through exchange of infected animals from Canada to South Korea^{29,30}. As of January 2021, 25 US states, 3 Canadian provinces, 3 European countries, and South Korea have all reported captive or free-ranging animals with CWD^{31,32}. Among these were naturally occurring cases in animals where CWD had not been seen before - moose³³ and reindeer³⁴. In captive herds where the disease has taken hold, infection rates of 1 in 10 animals are not uncommon with localized infection rates reaching 1 in 4³². To combat the spread, entire herds - not just animals with known infection - throughout the world have been depopulated. In this way, CWD has depleted deer, elk, moose, and reindeer populations through both direct (CWD itself) and indirect (human controlled herd depopulation) means.

Why Study CWD: The Potential for an Epidemic, Interspecies Transmission, and Ecological Impact

The rapid spread of this disease poses a direct threat to population dynamics for cervids and may cause sudden unforeseen ecosystem imbalances³⁵. Infectious prions can survive and remain transmissible in the soil in absence of an animal host³⁶, leaving plants and earth exposed to infected elk threatening to cohabitating species susceptible to transmission. With transmission barriers shifting as strain adaptation occurs in each new species host, transmission of CWD to humans may be possible through indirect routes just a few species transfers away we have yet to consider or explore.

Transmission barriers are not reflective of phylogenetic relationships between species or even overall sequence similarity, making interspecies transmission difficult to infer based on current knowledge³⁷. Disease transmission to predators of animals with a high incidence of prion disease is also not well understood, creating a potential new challenge of tracing prion diseases through food chains. These factors along with the prolonged incubation period, up to several years for animals and decades for humans, have made intervention via quarantine and herd depopulation an unsuitable means of complete disease containment and prevention. The unpredictable nature and impending threat of a prion disease epidemic in the United States makes the need for development of treatments urgent, but at present there are no cures, therapies, or preventative treatments on the market or in clinical trials for either humans or animals. This is largely due to a lack of molecular information about key features required for transmission and stability along with how to disrupt those features.

The Possibility of CWD Strains

In many animals, multiple prion strains exist with a variety of associated symptoms, progressions, and biophysical properties. The ratio between the three different glycosylation states of PrP is a common marker used to distinguish some prion strains. Strains are also distinguishable based on location of aggregate deposition, clinical progression, survival time, and other biophysical properties like protease resistance and molecular weights of the protease-resistant PrP fragments. Not unlike other mammalian prion diseases, CWD samples show a range of ratios between glycosylation states³⁸, as well as two distinguishable survival times and aggregate deposition patterns when passaged into ferrets³⁹ or syrian golden hamsters⁴⁰. Since 2020, it has also been proposed that cervids in Norway and North America have distinguishable strains of CWD⁴¹ and, similar to familial human prion diseases, there may be genetic factors that influence CWD prion strain⁴². Most of the work looking into the possibility of strains in CWD has been

conducted in the last 15 years, with key findings noting naturally occurring strains just coming to light in the past 1-2 years. This illustrates that the possibility of CWD strains is just beginning to be explored, and there may be strains we have yet to discover.

A Cervid Prion Structure: Importance for Establishing Structural Features Governing Prion Disease Transmission

The CWD transmission rate between cervids is much higher than for other animal prion diseases for unknown reasons. Knowledge of the misfolded conformation(s) of cervid prions may offer hints about what structural features allow for ease of transmission that could be extended to inferences about other prions. Prion structures from different strains would help pinpoint atomic level features that control transmissibility, protease-resistance and other biophysical properties, as well as link structures to clinical features like survival time. This type of information will require systematic studies of prion structures, not only those found in diseased animals, but also those formed *in vitro* that either do not form in animals or lack a property of interest, like infectivity. Lastly, structures from CWD prions would allow for systemic comparison to prion structures from other mammals to aid in understanding and prediction of cross species transmission relationships based on the structural compatibility of different prion sequences and strain structures.

Results

In this chapter, I present current progress in understanding the structures taken on by spontaneously forming bank vole and elk prions. In order to provide a direct comparison between preferred structures, the same stretch of residues used in chapter 3⁴³ was incorporated into each protein construct - residues 94-178 based on human numbering. This fragment contains all

residues proposed to be involved in the original proposed beta solenoid core⁴⁴, all residues found in at the core of fibrils found in patients with the familial prion disease Gerstmann-Sträussler-Scheinker disease GSS^{45,46}, and all residues found in the ordered core of an infectious Y145Stop variant that presents species-specific biophysical characteristics when studied by solid state Nuclear Magnetic Resonance (ssNMR)⁴⁷.

It is also worth noting that recombinant bank vole prions have been shown to be capable of producing the same 6-8kDa protease-resistant fragments seen in GSS patients⁷. GSS has yet to be demonstrated to be transmissible to other rodents, thus a model organism - and prion protein - where this rare disease can be studied and its associated prion aggregates can be amplified is desirable.

A Single Bank Vole Prion Fibril Morphology

Fibrils formed by recombinant bank vole PrP⁹⁴⁻¹⁷⁸ (rBvPrP⁹⁴⁻¹⁷⁸) under the conditions used were exceptionally homogeneous. A single morphology was detected after examination of hundreds of negative stain transmission electron micrographs and cryo-EM micrographs. This polymorph was easily identified by its short pitch - just 27 nm from one crossover to the next - and formed nearly instantaneously - within 10 minutes - of dilution of the monomeric protein into the growth buffer and initiation of acoustic resonance mixing (Figure 4.1a-c). Continued acoustic resonance mixing up to the longest incubation period of 30 hours did not alter the observed morphology. The only noteworthy change in fibrils left shaking between 12 and 30 hours was an increase in fibrils coalescing together to form tangled clumps, but all individual fibrils within these nets appeared to maintain the same structure. It is also worth noting that when fibrils were removed from the acoustic mixer, no further evolution of the sample took place. Fibrils kept their level of monodispersity and individual morphology after being kept under quiescent conditions on the

benchtop at room temperature for at least five months. These same results were achieved with multiple batches of fibrils formed from several protein expressions and purifications, illustrating that formation of this single morphology was reproducible.

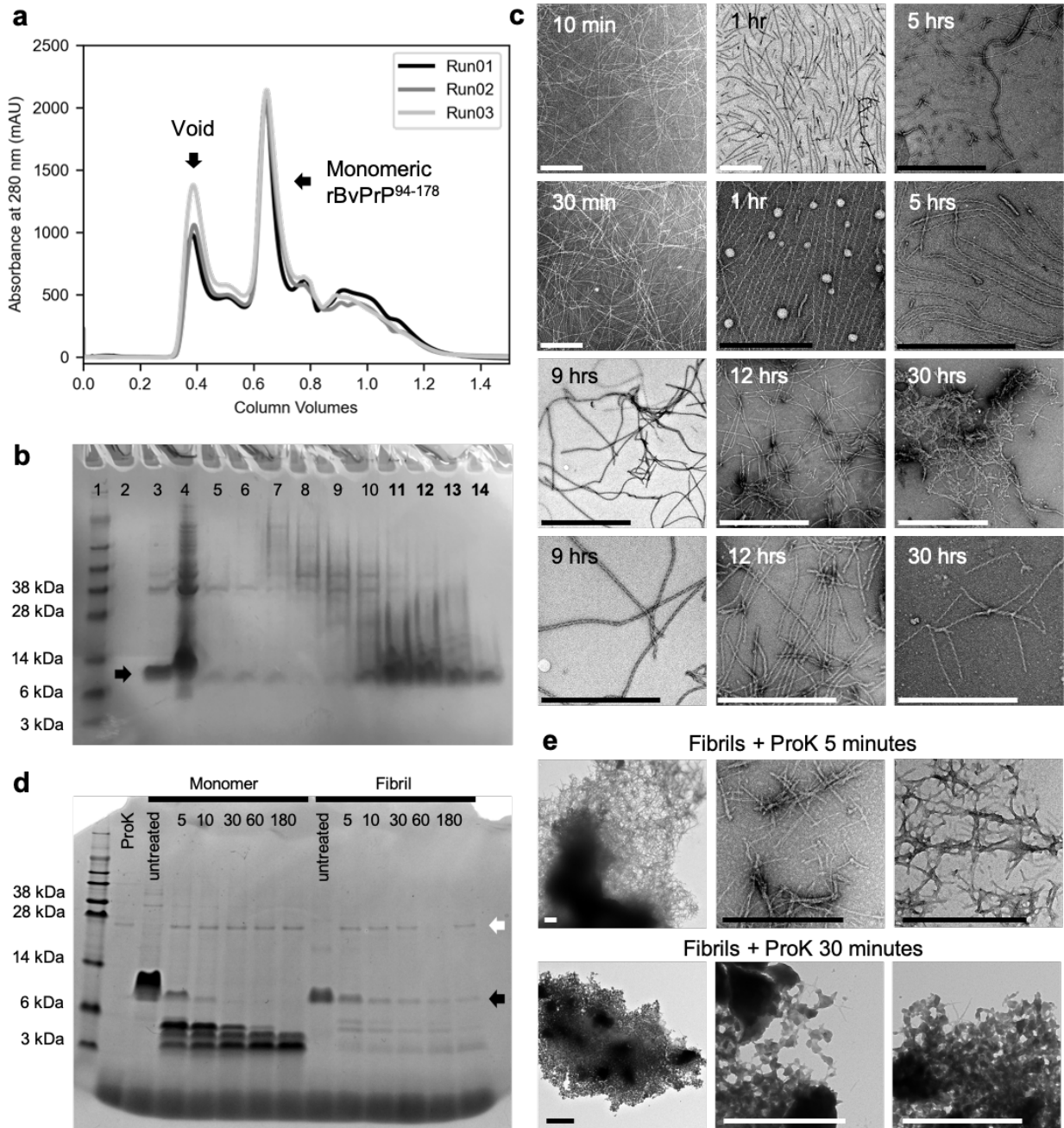


Figure 4.1 Formation and Characterization of rBvPrP⁹⁴⁻¹⁷⁸ Fibrils. (a) rBvPrP⁹⁴⁻¹⁷⁸ was purified by size exclusion chromatography (SEC) with most protein eluting as a monomer based on retention time and SDS-PAGE (b). Lanes in (b) are as follows: 1. SeeBlue Plus2 Protein Ladder 2. Supernatant after lysis 3. Pellet after solubilization in 8M guanidine HCl 4. Supernatant injected over an SEC column after solubilization in 8M guanidine HCl 5-10. Fractions eluting in the void volume and small peak before the large peak (lanes 11-14, bold) indicating ~9.7kDa monomeric rBvPrP⁹⁴⁻¹⁷⁸ (black arrow). (c) Fibrils formed within 10 minutes of dilution of rBvPrP⁹⁴⁻¹⁷⁸ into fibril growth buffer and agitation via acoustic resonance mixing. Fibrils with a single rapidly twisting morphology were the only morphology found in the sample regardless of whether they continued undergoing acoustic resonance mixing for at least 30 hours. (d) SDS-PAGE indicates these fibrils were somewhat more proteinase K resistant than monomeric, unfolded rBvPrP⁹⁴⁻¹⁷⁸, but could overall be described as protease-sensitive. Proteinase K is indicated with a white arrow (e) Proteinase K treated fibrils were imaged by negative stain electron microscopy after five or thirty minutes of incubation with proteinase K. Both samples experienced alterations that were immediately visible, including increased clumping, a loss of clearly identifiable fibril morphology, and accumulation of nonfibrillar aggregates. All scalebars 500nm.

Proteinase K Sensitivity of rBvPrP⁹⁴⁻¹⁷⁸

Unlike my previously shown rHuPrP⁹⁴⁻¹⁷⁸ fibrils⁴³, rBvPrP⁹⁴⁻¹⁷⁸ fibrils are protease-sensitive by comparison, but achieving a quantitative assessment of ProK sensitivity here was fraught with challenges. The monodispersity of fibrils in solution made nephelometry an unsuitable method for attempts to quantify fibril disaggregation. This method, which I previously used to illustrate dissolution of aggregates of human prion fibrils⁴³, is best suited for detection of large aggregates of fibrils and not the well separated individual fibrils present in the bank vole prion sample being studied here. Prions are also known to not always bind ThT, and upon adding ThT to these bank vole prion fibrils as well as control amyloid fibrils known to bind ThT, I was able to confirm that this bank prion fibril polymorph does not bind ThT, making this method also unsuitable for measurement of either fibril formation or dissociation after exposure to ProK.

In my final attempt at quantifying ProK sensitivity of this bank vole prion polymorph, I compared bands observed by SDS-PAGE from monomeric and fibrillar forms of the bank vole prion construct after incubation with ProK for different lengths of time (Figure 4.1d). This also presented with challenges that restricted quantitative measurements, the largest being incomplete entry of fibrillar samples into the gel. Nonetheless, from the SDS-PAGE gel and electron micrographs (Figure 4.1d and e) it is clear that the fibril polymorph studied here was less sensitive to ProK than its unfolded, monomeric counterpart, but much weaker against ProK than my previously studied fibrils of a human prion formed from the same stretch of residues. The greater ProK sensitivity of this bank vole prion fibril polymorph compared to my previously studied human prion fibril polymorph can be understood without atomic detail just based on the relative size and compactness of the two fibrils cores. The protease-resistant human prion (HuPrP^{Res}) core is made up of two symmetry related hairpins, constituting four layers of beta sheets in total. The bank vole

prion fibril core, on the other hand, is only composed of one beta strand per protofilament (Figure 4.2), corresponding to only two, less protected, beta sheets making up the fibril core.

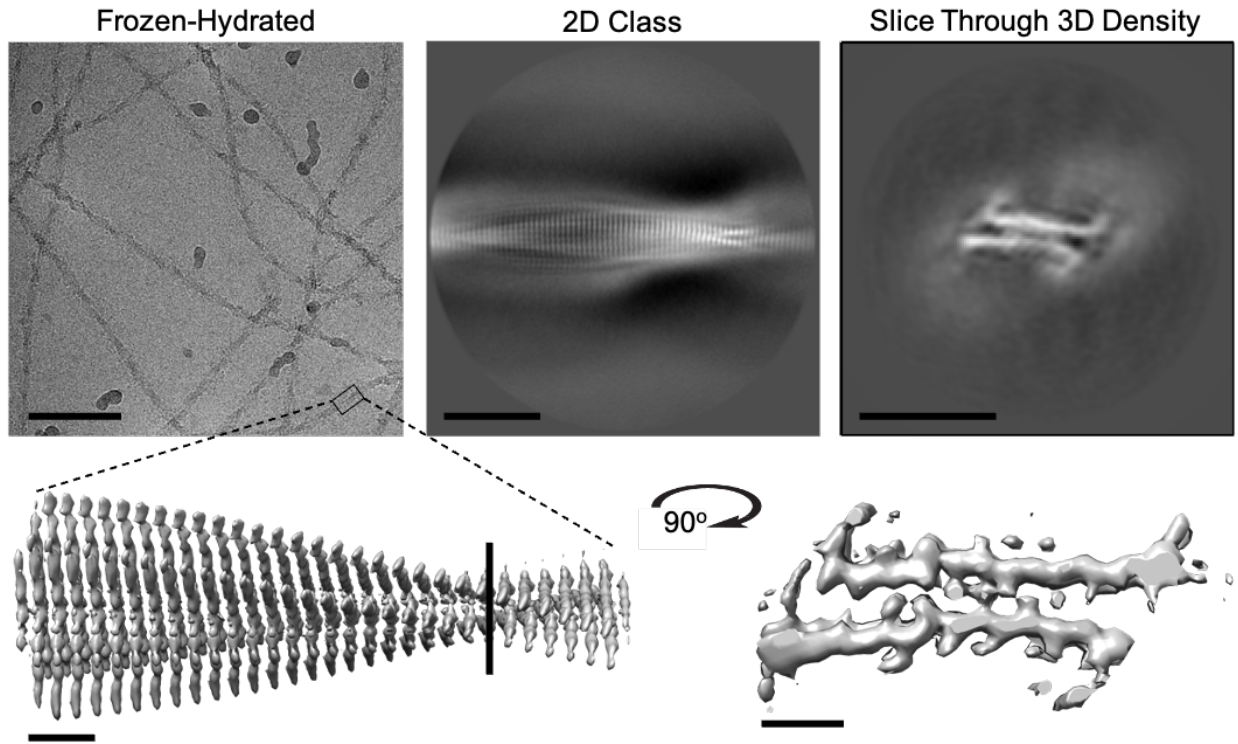


Figure 4.2. Cryo-EM Reconstruction of rBvPrP⁹⁴⁻¹⁷⁸ Fibrils. rBvPrP⁹⁴⁻¹⁷⁸ fibrils were imaged at high-resolution in a frozen-hydrated state (top left, scalebar 100nm). A representative 2D class for the single morphology found is shown in the top center panel (scalebar 100Å). Gold standard refinement in 3D revealed an approximately 20Å by 60Å cross section composed of two easily identifiable protofilaments (top right, scalebar 100Å). Strands on each rung are clearly separated in the density map (bottom left, scalebar 20Å) and a cross section after masking and sharpening shows ~15 residues per chain making up the ordered core with some side chains visible (bottom right, scalebar 20Å).

rElkPrP⁹⁴⁻¹⁷⁸ Prion Filaments: A Continuum of Structures

For rElkPrP⁹⁴⁻¹⁷⁸, nearly all fibril growth conditions tested resulted in a polymorphic mixture of fibrils forming within 24 hours. More separated fibrils were more common under very acidic conditions while more clumped fibrils were more common under neutral conditions (Figure 4.3), thus acidic conditions were pursued for optimizing fibril morphology for cryo-EM. For most growth conditions, untwisted species were favored. Upon reduction from pH 3 sodium citrate to pH 2 citric acid, some untwisted species remained but twisting fibrils that were otherwise similar in appearance and dimensions became favored. It is therefore expected that the fibril polymorphs found in the pH 2 condition represent the same untwisted species seen at pH 3, but with flexible twisting behavior. Determining structures from this mixture of what appear to be related polymorphs may aid in identification of features giving rise to flexibility in amyloid structures, a topic that has yet to be thoroughly explored but becomes relevant when considering mechanism of monomer recruitment, misfolding, and crossing strain or species barriers.

A mixture of ~7 polymorphs with different pitch and thickness could be identified in negatively stained fibrils in electron micrographs. Likewise, fibrils isolated from the brains of transgenic mice expressing the elk prion protein and inoculated with CWD prions have been noted to be highly structurally polymorphic in terms of dimensions and twisting behavior⁴⁸, some of which resemble the dimensions found in my fibril sample. This mixture of polymorphs may be an important feature involved in the high transmission rate of CWD between cervids, thus structure determination of a single fibril polymorph may be insufficient to explain the collective biophysical and structural properties of CWD prion aggregates. Based on these two considerations - understanding flexibility in amyloid structure and the presence of multiple polymorphs in CWD - cryo-EM data was collected from this polymorphic elk fibril sample.

In order to minimize misclassification, each of three types of polymorphs based on crossover distance were manually picked separately. These consisted of “short” “medium” and “long” pitch polymorphs. While there was variation in crossover distance for each group, cutoffs for each group were chosen after visual assessment of the most commonly occurring crossover distances centering around a few values. Within some groups, there were also polymorphs that could be more readily distinguished during two-dimensional (2D) classification based on fibril width. These morphologies were not manually picked separately from other fibrils of similar pitch.

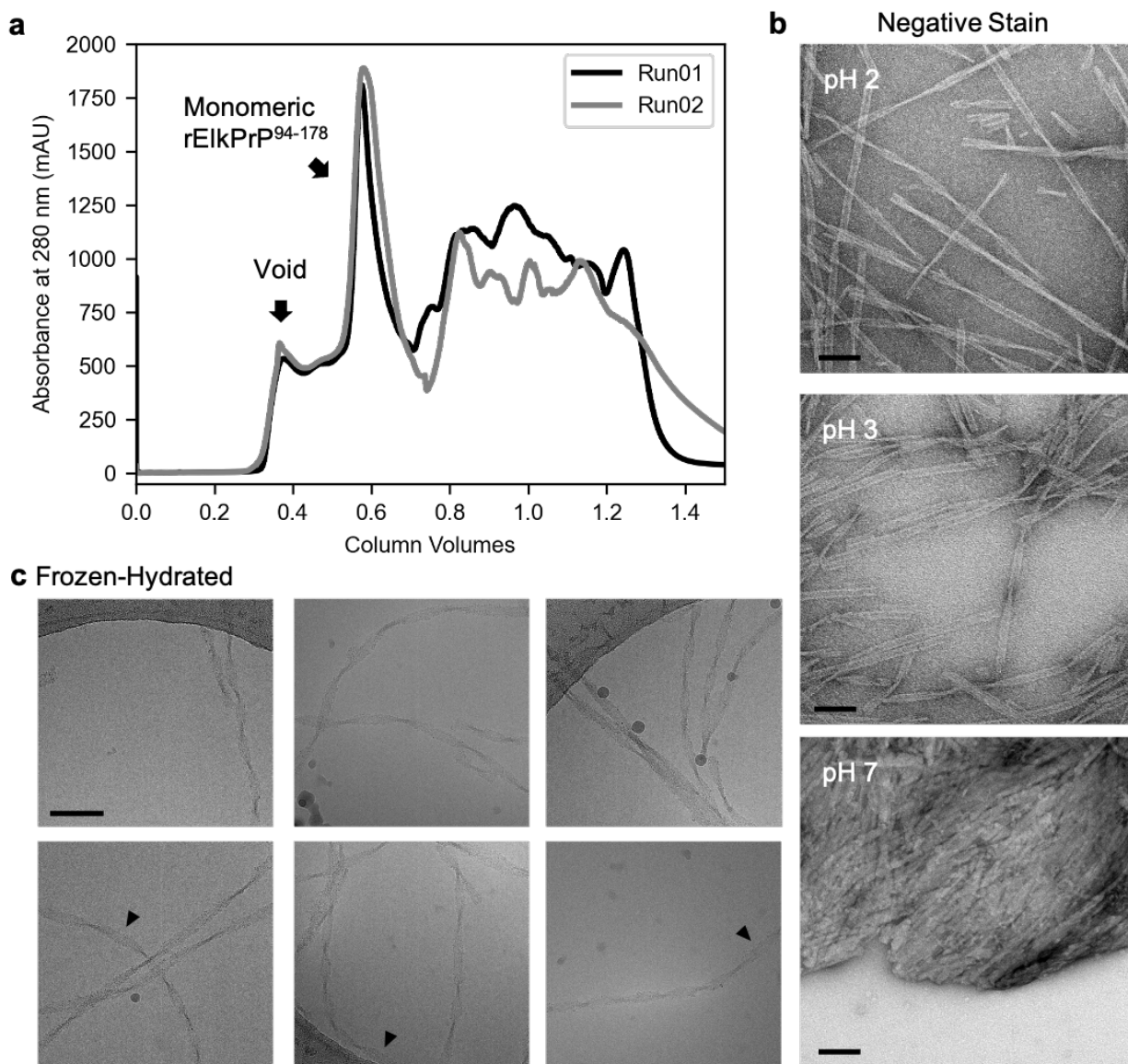


Figure 4.3. Formation of rElkPrP⁹⁴⁻¹⁷⁸ Fibrils. (a) Monomeric rElkPrP⁹⁴⁻¹⁷⁸ was purified by size exclusion chromatography. (b) Fibrils were formed under various conditions, but conditions using neutral pH buffers tended to result in tightly clumped fibrils (bottom) while acidic pH buffers yielded more monodisperse fibrils (top and middle). Fibrils formed at pH 2 appear similar in maximum width and appearance as fibrils formed at pH 3, but have taken on twisted arrangements. (c) Fibrils formed at pH 2 preserved in a frozen-hydrated state imaged at high resolution by cryo-EM.

Many fibrils seem to adopt a regular twist for some portion of the fibril before changing pitch (black arrows). All scalebars 100nm.

Fibrils falling into the short pitch group exhibited three distinguishable morphologies based on how bent sheets at the core appeared in cross sections (Figure 4.4). Because fibrils seemed to change from one morphology to another within an individual fibril (Figure 4.3c), these classes may represent the most commonly observed points along a continuum of flexible structures sharing a common core. This is supported by the similar appearance of fibrils formed at pH 2 and pH 3, with the main apparent difference being whether the fibrils were twisting or not. This hypothesis would require higher resolution maps where sequence assignment is possible to confirm.

Fibrils falling into the medium pitch group also exhibited some variability in morphology, largely in the pitch. Thin medium pitch fibrils closely resembled the short pitch fibrils found in class 1, albeit with a longer stretch of the two strand core being visible and a much longer crossover distance of ~110nm (Figure 4.5). Thicker fibrils of medium pitch were very rare, making up just 12% of total medium pitch fibrils identified during 2D classification (Figure 4.6a). Preliminary 3D classification using a small number of mixed quality particles yielded fibril cross sections that appeared to be composed of the same two strand building blocks arranged in pairs (Figure 4.6b and c).

Discussion

A wide range of sequence dependent structures are taken on by mammalian prions, both in tissue derived and *in vitro* generated prion samples. Dozens of prion diseases and strains exist across humans and animals, some of which have only recently been discovered^{38-40,42}. While it has yet to be seen whether structures found in humans and animals infected with prion disease can be reproduced over a short time frame *in vitro* given the right circumstances, it may be a long wait to find out given the large number of strains for which structures have yet to be uncovered. It is therefore critical to forge ahead with structural studies of *in vitro* generated material with the aim

of gaining information about conformations accessible to prions and what features bestow - or take away - a structure's stable and infectious properties. With the additional restrictions and challenges associated with determining structures from tissue-derived prion strains, a method known to generate the same structure in a more readily available *in vitro* environment will be well worth the investment to improve both basic and translational science.

There is currently no robust way to predict amyloid folds the way globular structures can be predicted, including why some folds recur in all patients sharing a diagnosis while a distinct fold is adopted by the same protein in all patients sharing a different diagnosis. There is also no way of knowing why some structures are formed in humans and animals naturally while others may not occur naturally, but still have properties of interest like stability or infectivity.

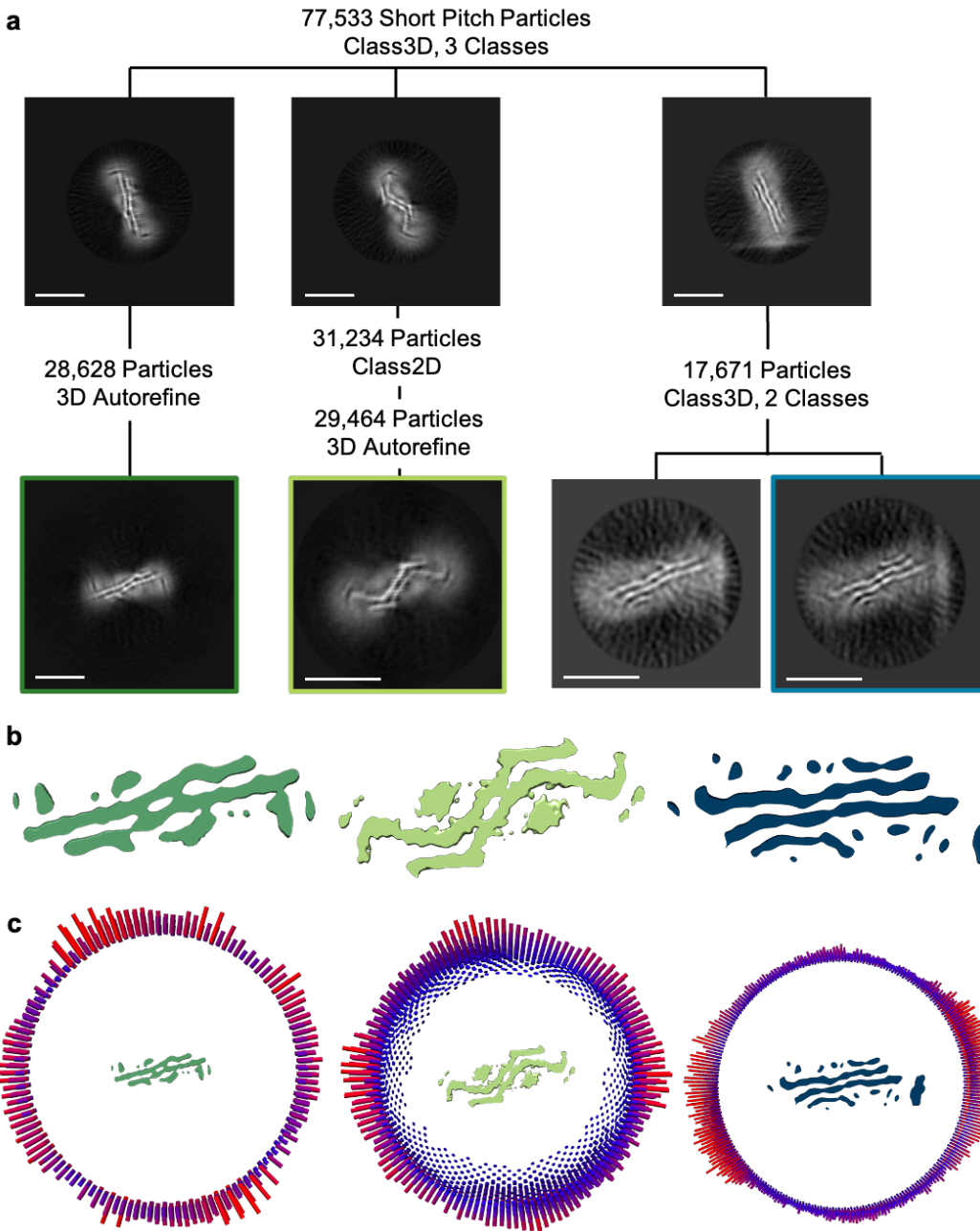
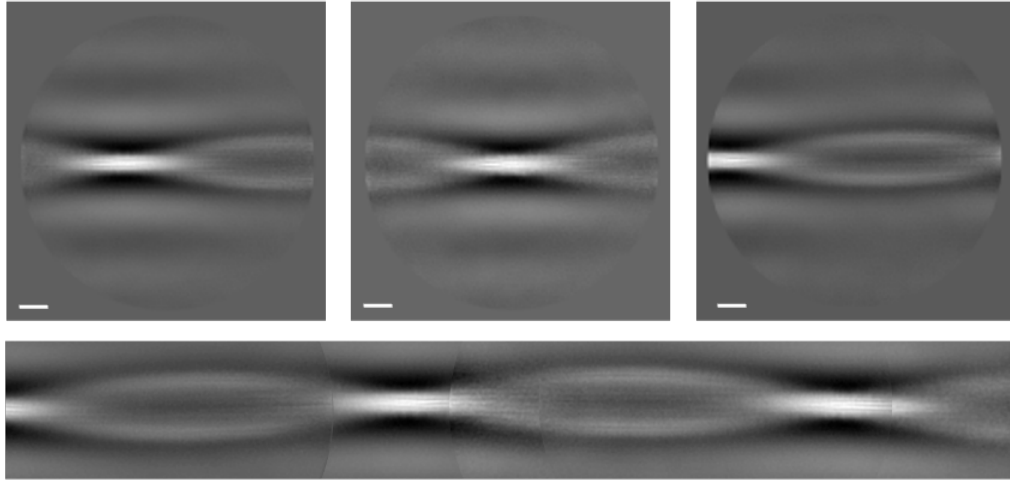


Figure 4.4. 3D Classification and Refinement of Short Pitch rElkPr^{P94-178} Polymorphs. (a) All particles belonging to fibrils with a short crossover distance (near 66nm) were separated into three 3D classes. Each class then underwent subsequent 2D or 3D classification to optimize particles included, helical parameters, and box size for the class being studied. Scalebars = 100A. (b) Gold standard refinements produced maps for short class 1 and 2 where clear chains can be traced in a cross section of the density. Likewise, 3D classification of short class 3 yielded maps

where a chain could be traced in the cross section. Maps are scaled to one another. (c) Angular distributions of particles included in short class 1-3. Because of the low resolution ($\sim 7\text{-}8\text{\AA}$) for classes 1 and 3, tilt angles were kept restricted.

a

Thin Polymorph (65,477 Particles)



$\sim 1,186\text{\AA}$ crossover (-0.728° twist)

b

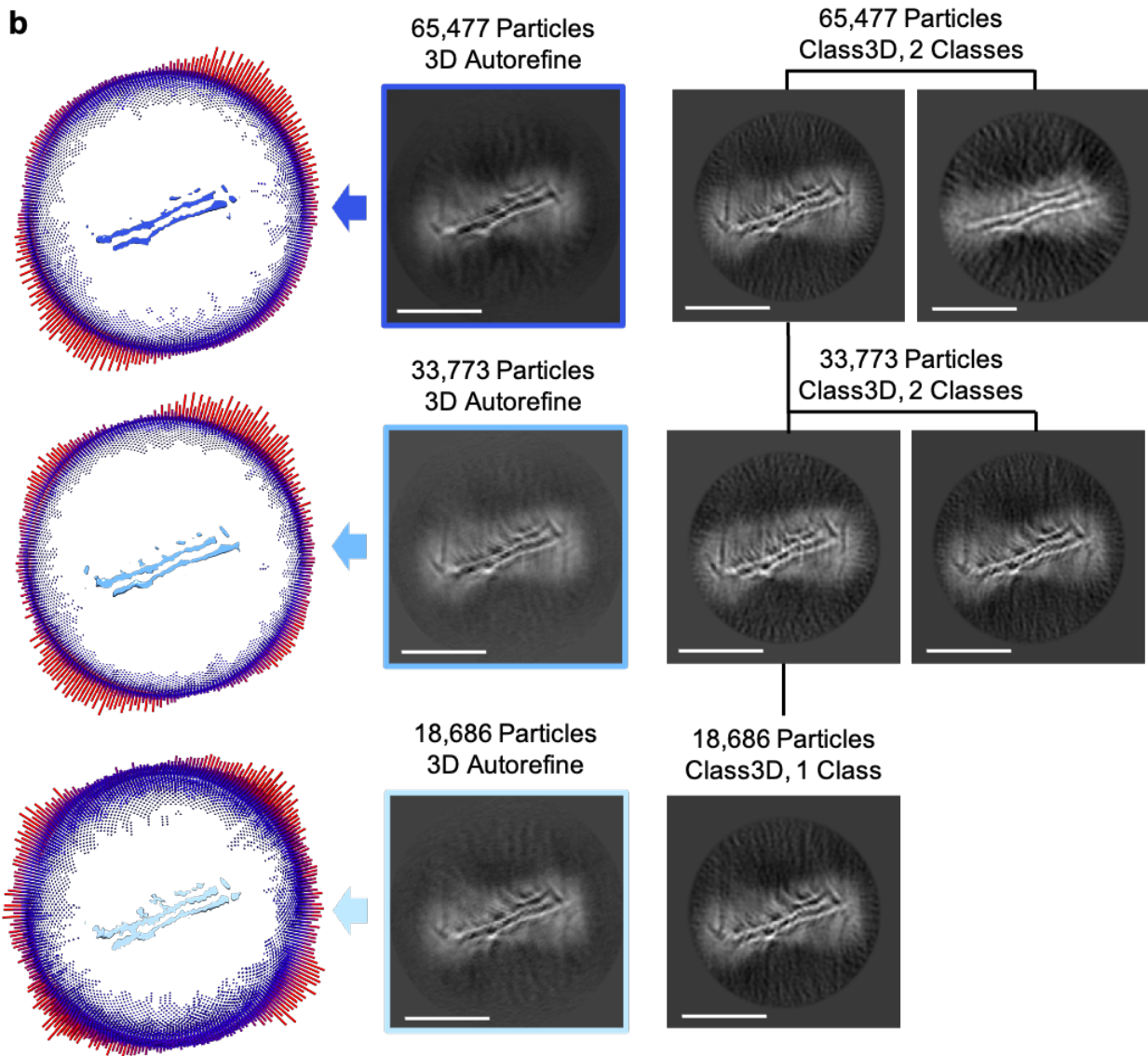


Figure 4.5. 3D Classification and Refinement of a Thin Medium Pitch rElkPrP⁹⁴⁻¹⁷⁸ Polymorph. (a) 2D classification of medium pitch rElkPrP⁹⁴⁻¹⁷⁸ fibrils revealed two morphologies of similar pitch, one thin shown here and one thick (Figure 4.6). Stitching of 2D classes led to an initial crossover distance estimate of 1,186Å and a corresponding twist of -0.728 degrees. (b) Rounds of 3D classification and 3D autorefinement were used to illustrate that a single thin morphology was present and remove all but the highest resolution particles. Particle orientation distributions are also shown (left). All scalebars = 100Å.

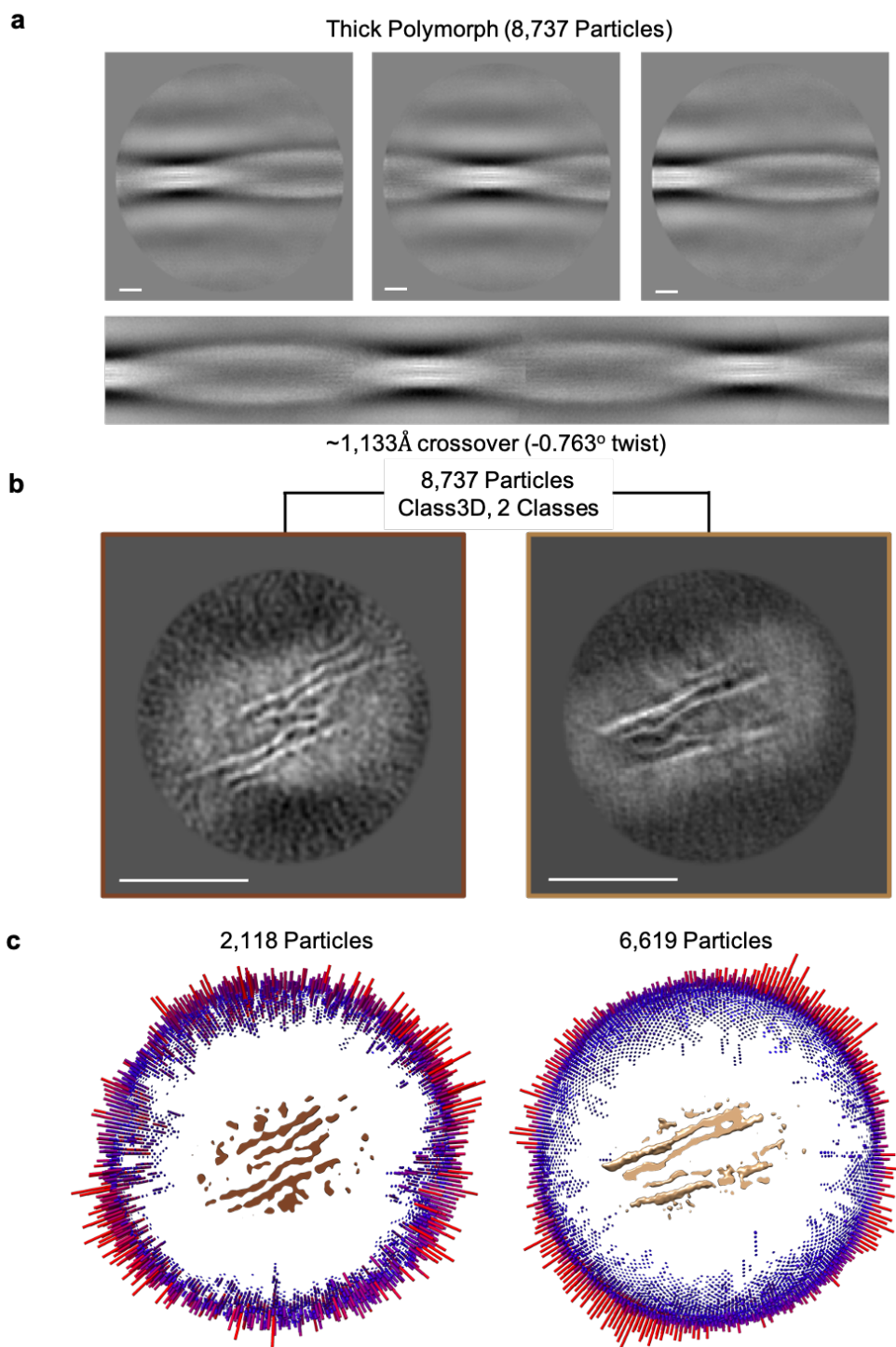


Figure 4.6. 3D Classification of Thick Medium Pitch rElkPrP⁹⁴⁻¹⁷⁸ Polymorphs. (a) 2D classification of medium pitch rElkPrP⁹⁴⁻¹⁷⁸ fibrils revealed two morphologies of similar pitch, one thick shown here and one thin (Figure 4.5). Stitching of 2D classes led to an initial crossover distance estimate of 1,133Å and a corresponding twist of -0.763 degrees. (b) One round of 3D

classification was used to separate particles into two classes with particle distributions shown in
(c). All scalebars = 100Å.

Assigning physiological properties to a specific fibril morphology is also complicated by heterogeneity in disease strains where a minor, uncharacterized polymorph may be responsible for a noteworthy phenotype such as toxicity or infectivity. Isolation of a single polymorph - which may only be possible *in vitro* for some diseases - offers a rare opportunity to assign biological properties to a single structure. By characterizing properties of each polymorph in isolation, the opportunity to explain prion behavior extends beyond the strain being studied and into any strain including the characterized polymorph. With the establishment of a direct line between biophysical properties and phenotype for more prion polymorphs, we will be able to model how strains with a disease-specific recipe of ingredient polymorphs will manifest.

Here I have demonstrated that recombinant fibrillar prions formed under similar conditions can result in different structures depending on the sequence of origin. Just a few amino acid substitutions are sufficient to push the equilibrium from rapid formation of a single polymorph to formation of a continuum of flexible structures.

rBvPrP⁹⁴⁻¹⁷⁸ rapidly and spontaneously forms fibrils that can be uniform in morphology. This implies that not only is BvPrP prone to misfolding into amyloid conformations, but these misfolded conformations are not random, and some may be much more favorable than others. This bank vole prion polymorph is also relatively protease-sensitive compared to a previously described polymorph of HuPrP formed from the same stretch of residues⁴³. This difference may be at least partly explained by the comparably small core in rBvPrP⁹⁴⁻¹⁷⁸ made up of only two sheets compared to 4 in rHuPrP^{Res}. It is also noteworthy that not all prions share the same biophysical properties. Naturally occurring prion diseases display a wide range of aggregate locations, protease-sensitivities, clumping behavior, fibrillar or sub-fibrillar morphology, and infectivity. Study of polymorphs known to harbor or omit each of these properties will be crucial to tease out how each property is encoded through structure.

rElkPrP⁹⁴⁻¹⁷⁸, on the other hand, displayed a wide range of morphologies with variable pitch and thickness that could be readily distinguished in electron micrographs. Cryo-EM revealed these morphologies may share a common, flexible core that can bend or twist from one morphology to the next even within the same filament (Figure 4.3 and 4.7). A continuum of amyloid structures like this with more than just slight variations in pitch has only been described with high resolution detail once before for tau⁴⁹. What role flexibility plays in the properties of amyloids remains to be understood, but flexible amyloid structures like this are likely more common than is reflected in the recent literature aimed only at high resolution structure determination. Flexibility in amyloid filaments creates challenges in data processing that restricts achievable resolution, thus samples displaying flexibility are often abandoned or not pursued.

This elk prion fibril sample also raises philosophical questions about what constitutes a distinct polymorph. Because these flexible structures can be found along the same fibril, some might answer that no, they are not polymorphs and only an incompatible arrangement of residues or a new quaternary structure would be necessary for this designation. This introduces limitations on what conclusions can be drawn about the polymorphism present in a sample visualized by common methods that are only reliable out to modest resolution, such as negative stain electron microscopy.

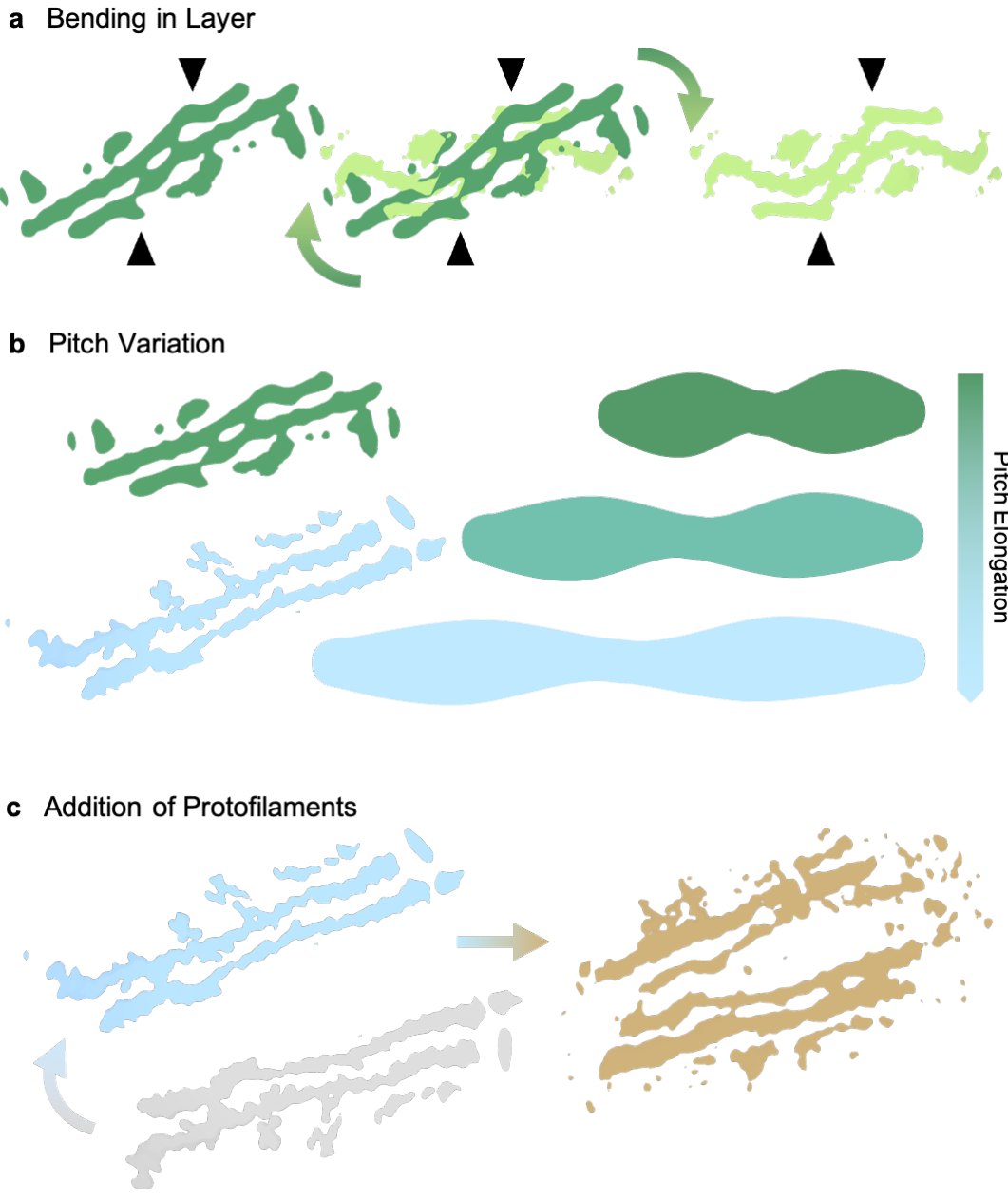


Figure 4.7. Schematic of Types of Polymorphism Seen in rElkPrP⁹⁴⁻¹⁷⁸ Fibrils. (a) Illustration of how polymorphs belonging to short class 1 and 2 may convert to one another via bending within a fibril layer, (b) how polymorphs in short and medium pitch classes could interconvert via pitch elongation achieved by shrinking the degrees rotated in each layer, and (c) how less abundant thick fibrils found in the sample appear to be the result of protofilament addition along the sides of the fibril.

Materials and Methods

Recombinant Expression and Purification of rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸

As with previous work on rHuPrP⁹⁴⁻¹⁷⁸ ⁴³, a gene encoding either the equivalent bank vole PrP residues 94-178 (IDT) with the M109 genotype, or elk PrP residues 94-178 (human / bank vole numbering) was cloned into a pet28a+ *Escherichia coli* expression vector derivative containing no purification or solubility tags via Gibson assembly between NcoI and XhoI restriction sites. Correct gene insertion and sequence were confirmed by DNA sequencing (Genewiz). Recombinant bank vole PrP 94-178 (rBvPrP⁹⁴⁻¹⁷⁸) or recombinant elk PrP 94-178 (rElkPrP⁹⁴⁻¹⁷⁸) was expressed in BL21 GOLD cells via growth at 37 degrees C while shaking at 180-230 rpm until the optical density measured at 600nm (OD₆₀₀) reached 0.6-0.8. Cells were induced with 1mM IPTG and allowed to overexpress target proteins for 3.5-6 hours under the same cell growth conditions. Cells were then harvested via centrifugation at 8,000g for 10 minutes and stored at -80 degrees C until protein purification.

Purification was carried out as previously described^{43,50} and reiterated in chapter 2 with only distinctions highlighted here. After lysis, insoluble material was pelleted via centrifugation at 30,000g for 1 hr (rBvPrP⁹⁴⁻¹⁷⁸) or at 10,000g for 30 minutes (rElkPrP⁹⁴⁻¹⁷⁸) at 4 degrees C. After washing the pellet with Buffer A (chapter 2) and solubilizing in freshly made Buffer B (chapter 2), remaining insoluble material was pelleted via centrifugation for 20 minutes at 20,000g (rBvPrP⁹⁴⁻¹⁷⁸) or 35,000g (rElkPrP⁹⁴⁻¹⁷⁸) based on how readily soluble and insoluble material separated. Size exclusion chromatography was carried out as previously described in the literature (Wan 2015, Glynn 2020) and in chapter 2. After desalting, protein containing fractions were pooled and concentrated to 3.2 - 8 mg/ml before being flash frozen using liquid nitrogen and stored at -80 degrees C until use.

Growth of rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸ Fibrils

Fibrils of rBvPrP⁹⁴⁻¹⁷⁸ were grown at 0.5mg/ml in 0.5M urea, 50mM NaCitrate pH 3, and 200mM NaCl. Fibrils used for cryo-EM were grown via acoustic resonance mixing^{43,51,52} at 38Hz for 9 hours. It was later found that acoustic resonance mixing times as short as 10 minutes were able to produce abundance fibrils of the same morphology that did not evolve over time. After approximately 12 hours of shaking, individual fibrils maintained the same morphology, but began to irreversibly aggregate into clumps. Fibrils removed from the mixer after less than 12 hours of shaking did not clump or change morphology over the course of at least five months while they were stored quiescent at room temperature. Fibrils of rElkPrP⁹⁴⁻¹⁷⁸ were grown at 1mg/ml in 2M urea, 50mM Citric Acid pH 2, and 200mM NaCl. Fibrils later used for cryo-EM were also allowed to grow via acoustic resonance mixing at 38Hz for three days.

Negative Stain Transmission Electron Microscopy of rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸ Fibrils

Fibril morphology was screened by negative stain transmission electron microscopy (TEM). After appropriate dilution (10-20x for rBvPrP⁹⁴⁻¹⁷⁸, 50x for rElkPrP⁹⁴⁻¹⁷⁸) of concentrated fibrils into water, 3μL was applied to a 300 mesh formvar carbon (F/C) grid (Ted Pella). After being allowed to incubate for two minutes, 3μL of 2% uranyl acetate (UA) solution was applied to the grid and also allowed to incubate for two minutes before excess liquid was wicked away with filter paper. The grid was then allowed to air dry for a few minutes before storage or imaging using a Tecnai 12 electron microscope (FEI).

Partial Protease Digestion of rBvPrP⁹⁴⁻¹⁷⁸

Several methods were initially pursued for assessment of Proteinase K (ProK) resistance of rBvPrP⁹⁴⁻¹⁷⁸ fibrils. Nephelometry, which I previously used to illustrate dissolution of aggregates of human prion fibrils⁴³. Even concentrated rBvPrP⁹⁴⁻¹⁷⁸ fibrils were not able to yield nephelometry readings readily distinguishable from readings of fibril growth buffer with proteinase K alone, thus this method was not pursued for quantification of aggregates in this case.

For the ThT assay, 25 μ M rBvPrP⁹⁴⁻¹⁷⁸ and 174 μ M insulin were first confirmed to contain fibrils by negative stain electron microscopy. To assess the ability of rBvPrP⁹⁴⁻¹⁷⁸ to bind ThT, 25 μ M ThT was added to either unproteolyzed rBvPrP⁹⁴⁻¹⁷⁸ fibrils, rBvPrP⁹⁴⁻¹⁷⁸ fibril growth buffer alone, or insulin fibrils in a total reaction volume of 15 μ L in a 384 well black flat bottom plate (Corning). Emission at 482 nm was recorded using a Varioskan plate reader with a 440nm excitation wavelength for a 100ms or 1s integration time. Fluorescence for rBvPrP⁹⁴⁻¹⁷⁸ fibrils (2.4 AU) barely registered above the buffer only control (0.7 AU), while fluorescence for insulin gave a typical reading for ThT binding samples for this instrument⁵³ (133.9 AU). Thus, this rBvPrP⁹⁴⁻¹⁷⁸ polymorph did not readily bind ThT and this method was also not pursued for quantification of aggregates of this sample.

In the final attempt to quantify proteinase K resistance of rBvPrP⁹⁴⁻⁷⁸, SDS-PAGE was used. Enzyme was dissolved in rBvPrP⁹⁴⁻¹⁷⁸ growth buffer and added to either fibrillar or unfibrillized, unfolded rBvPrP⁹⁴⁻¹⁷⁸ in a 1:100 molar ratio of ProK : rBvPrP⁹⁴⁻¹⁷⁸ monomer. After addition of enzyme, samples were incubated for 5, 10, 30, 60, or 180 minutes at room temperature under quiescent conditions. After the incubation time was reached, SDS-PAGE loading buffer was added to each sample and heated to 98 degrees for 10 minutes to stop the reaction. For each time point, a sample of the digestion was also immediately diluted 5x into water and 3 μ L was applied to a glow-discharged (PELCO easiGlow) 300 mesh carbon F/C electron microscopy grid (Ted Pella). The sample was allowed to incubate for two minutes before 3 μ L of 2% UA was added.

After two minutes of incubation with UA stain, excess liquid was wicked away with filter paper and the grid was allowed to air dry for a few minutes.

Preparation of Frozen-Hydrated rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸ Fibril Grids for Single Particle

Cryo-EM

For both rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸, unproteolyzed fibrils were used for cryo-EM data collection. After 9 hours of fibril formation and 16 days under quiescent conditions on the benchtop, rBvPrP⁹⁴⁻¹⁷⁸ fibrils were diluted 25 fold into 50mM pH 3 NaCitrate buffer. R1.2/1.3 300 mesh Copper grids (Quantifoil, Electron Microscopy Sciences) were glow-discharged for 90 seconds on each side using a PELCO easiGlow glow discharger. 1.8microl of sample was applied to each side of the grid before being plunge-frozen into liquid ethane using an FEI Vitrobot Mark IV set to 100% humidity and 12 degrees C. A blot time of 11 and a blot force of 1 were used.

For the first round of cryo-EM data collection, rElkPrP⁹⁴⁻¹⁷⁸ fibrils were diluted 200x into water before being sonicated in an ultrasonic bath (Fisher Scientific) for 2 seconds to break up fibers slightly. The grids, glow-discharging, and sample application for rElk^{PrP⁹⁴⁻¹⁷⁸} was identical to rBvPrP⁹⁴⁻¹⁷⁸, with the only changes to the plunge-freezing process being the vitrobot settings of temperature (4 degrees), blot time (4 seconds), and drain time (1 second). For the second round of data collection, sample concentration was increased to allow more particles from the many polymorphs to be imaged. These samples were diluted 100x into water. For grid preparation, all parameters were unchanged except blot time, which was adjusted to 5 seconds.

Cryo-EM Data Collection of rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸ Fibrils

For rBvPrP⁹⁴⁻¹⁷⁸ fibrils, images were collected on a Titan Krios (FEI) operating at an accelerating voltage of 300keV equipped with a Gatan K2 direct electron detector attached to a Gatan Image Filter (GIF) with a slit width of 20eV. Automated image collection was carried out using SerialEM⁵⁴ at 130,000x in super resolution counting mode. For the first 350 images, 40 200ms second frames were collected for a total exposure time of 7 seconds at a dose rate of $7.1e^-/A^2/s$ (or $1.42e^-/A^2/frame$). After the first 350 images, the illuminated area was shrunk from 700nm to 660nm to allow capture of two images per hole. For these images, 30 200ms frames were collected for a total exposure time of 6s at a dose rate of $8.6e^-/A^2/s$ (or $1.72e^-/A^2/frame$). In total, 8991 images were collected over three days with a target defocus range of -1.6 - -2.2 microns.

For the second dataset collected on rBvPrP⁹⁴⁻¹⁷⁸ fibrils, images were collected on a Titan Krios (FEI) operating at an accelerating voltage of 300keV equipped with a Gatan K3 direct electron detector attached to a GIF with a slit width of 20eV. Automated image collection was carried out using Legikon⁵⁵ at 105,000x in super resolution counting mode. 36 83ms frames were collected for a total exposure time of 3 seconds at a dose rate of $13.2e^-/A^2/s$ (or $1.10e^-/A^2/frame$). In total, 2630 images were collected over the course of one day with a target defocus of -2 microns.

For rElkPrP⁹⁴⁻¹⁷⁸, two datasets were also collected. For the first dataset, the same instrument and data collection parameters were used as in the first rBvPrP⁹⁴⁻¹⁷⁸ dataset with the following distinction. 30 200ms frames were automatically collected using legikon for a total exposure time of 6s at a dose rate of $6e^-/A^2/s$ for a total dose of $36e^-/A^2$ (or $1.20e^-/A^2/frame$). 5732 images were collected over the course of four days.

For the second rElkPrP⁹⁴⁻¹⁷⁸ dataset, images were collected in a manner similar to the second rBvPrP⁹⁴⁻¹⁷⁸ dataset with the following distinctions. Automated image acquisition was carried out at 81,000x in super resolution counting mode. 36 83ms frames were collected for a total exposure

time of 3s with a dose rate of $50e^-/A^2$ (or $1.39e^-/A^2/frame$). 7453 images were collected over the course of three days.

Cryo-EM Data Pre-Processing of rBvPrP⁹⁴⁻¹⁷⁸ and rElkPrP⁹⁴⁻¹⁷⁸ Fibrils

For all datasets, motion correction was carried out in MotionCorr2⁵⁶ implemented in RELION 3.0⁵⁷ using 5 x 5 or 4 x 4 (rElkPrP⁹⁴⁻¹⁷⁸, dataset one only) patches per micrograph. All datasets were collected in super resolution counting mode, and were therefore images were all binned 2x after motion correction with the exception of the first rElkPrP⁹⁴⁻¹⁷⁸ dataset, where images were left unbinned at this stage. For the first rBvPrP⁹⁴⁻¹⁷⁸ dataset, there was an abundance of images and a large number of fibrils of a single polymorph captured in each image. Thus, before contrast transfer function (CTF) correction using Gctf⁵⁸ images with either poor quality or no ice or images that were too far from focus were manually removed. 80% of all images in the dataset, or 7219 out of 8991, had potential to contain usable images of fibrils and passed this initial filtering step. For all other datasets, images were not manually removed before CTF correction.

For rBvPrP⁹⁴⁻¹⁷⁸ dataset one, after ctf correction images with an estimated resolution better than 4Å were carried forward. 6346 micrographs met this resolution cutoff, representing 88% of images selected during the first filtering round and 71% of the total dataset collected. For all other datasets, no filtration was carried out based on estimated micrograph resolution and all micrographs were used for manual filament selection in RELION 3.0. For both rElkPrP⁹⁴⁻¹⁷⁸ datasets, fibrils that had “short” (~66nm) and “medium” (~100nm) crossover distances were picked separately. Fibrils with very long crossover distances (>140nm) were not selected due to their extreme variability in pitch and flexibility that would further hinder efforts to determine the structure of a fibril with such a long pitch. After manual selection of particles from rBvPrP⁹⁴⁻¹⁷⁸ dataset one, 6219 micrographs were found to contain picked particles. This represented 98% of

images that went into the manual picking step and 69% of all micrographs collected in the dataset. The second rBvPrP⁹⁴⁻¹⁷⁸ dataset has yet to undergo preprocessing.

Two-Dimensional Classification of rBvPrP⁹⁴⁻¹⁷⁸ Fibrils

At this time, only the first dataset from rBvPrP⁹⁴⁻¹⁷⁸ fibrils have undergone two-dimensional (2D) classification and will be described here. An initial box size of 256 pixels with an interbox distance of 19.2Å was used to extract 2,002,105 particles from 6219 micrographs. These particles were binned 2x to a working box size of 128 pixels with 2.14Å / pixel to speed up early classification steps. Particles were subjected to a 200Å high pass filter in RELION before being divided into 4 groups of ~500,000 particles each to undergo 2D classification in 200 classes separately. After a few rounds of removing poor quality 2D classes, the remaining 750,008 particles were pooled, reextracted without binning, and subjected to another round of 2D classification into 200 classes together. The short fibril crossover distance allowed it to be measured directly via 2D classification using either a 384 or 512 pixel box, with both box sizes yielding a similar crossover estimate near 270Å. This measurement varied within ~20Å depending on the class being measured. 476,317 particles belonging to the best looking and highest resolution classes were carried forward to three-dimensional (3D) classification.

Two-Dimensional Classification of rElkPrP⁹⁴⁻¹⁷⁸ Fibrils: Short Pitch

At this time, only the first dataset from rElkPrP⁹⁴⁻¹⁷⁸ fibrils have undergone two-dimensional (2D) classification and will be described here. 109,098 particles in the short pitch group were initially extracted with a 1600 pixel box rescaled to 800 pixels at 1.07Å/pixel with a 28.8Å interbox distance. Particles were classified into 25 classes in order to obtain an estimate of the crossover distance that could be captured within a single image to minimize errors introduced by stitching

together classes of smaller box size. All particles were then reextracted with 800 pixel boxes binned to 400 pixels at 1.07Å/pixel. These particles underwent several rounds of 2D classification until 77,533 particles from high quality classes displaying 4.8Å strand spacing along the helical axis remained.

After an initial round of 3D classification described in the next section three slightly different structures could be identified and are subsequently described as short class 1, 2, and 3. 31,234 particles belonging to short class 2 were subjected to an additional round of 2D classification with the aim of removing the poorest quality particles. Ten classes were used and seven classes composed of 29,644 particles with visible 4.8Å features were selected and subjected to additional rounds of 3D classification to be described in the next section. 17,671 particles belonging to short class 3 were also subjected to an additional round of 2D classification using a 2048 pixel box binned to 512 pixels at 2.14Å/pixel. The goal of 2D classification for this group of particles was to obtain a more accurate pitch estimate of short class 3 specifically from a single large box without errors introduced by stitching classes with minimal defining features.

Two-Dimensional Classification of rElkPrP⁹⁴⁻¹⁷⁸ Fibrils: Medium Pitch

98,808 particles with crossover distances of ~100nm were extracted using a box size of 2048 pixels binned to 512 pixels at 2.14Å/pixel. 2D classification was carried out with 50 classes, and two readily identifiable morphologies with similar crossover distances emerged - a thick morphology and a thin morphology. Because of the long pitch, a few classes needed to be stitched together, even with this large box, in order to obtain an estimate of the pitch for both thick and thin fibrils. Thin fibrils with crossover distances of ~1,186Å made up 88% of medium pitch fibrils selected with 65,477 particles falling into these classes and moved into 3D classification steps.

The remaining 12% of selected fibrils belonged to classes with fibrils of thicker morphology with a similar estimated crossover distance of $\sim 1,133\text{\AA}$. This translates to 8,737 particles belonging to the “thick” subset being carried to 3D classification.

Three-Dimensional Helical Reconstruction of rBvPrP⁹⁴⁻¹⁷⁸ Fibrils

For thin filaments with a short pitch, smaller than typical box sizes are often used to best align particles taking up a relatively small number of pixels⁵⁹⁻⁶³. The short $\sim 270\text{\AA}$ crossover distance and $\sim 130\text{\AA}$ thickness of rBvPrP⁹⁴⁻¹⁷⁸ fibrils at their thickest point led to 3D classification and gold standard autorefinement attempts at various stages of the classification and autorefinement process using particles that were reextracted using various box sizes - 160, 200, 256, and 384 pixels. All box sizes used yielded cross sections similar in appearance, but the very small ordered core of the fibril - approximately 20\AA by 60\AA - could be best resolved using a 160 pixel box. Thus, this small box size was used for subsequent 3D classification and autorefinement.

Several rounds of 3D classification aligning the central 30% of the box were carried out using three classes and C1 symmetry. A gaussian ellipsoid was used as an initial reference model. The crossover distance measured from 2D classes was used to calculate an initial estimate of the twist value that was optimized to -3.11 degrees for a left handed helix. All three classes in each round of classification represented the same ultrastructure, but only 79,114 particles belonging to the highest resolution class were carried forward for CTF refinement and subsequent 3D autorefinement. At no point during the cryo-EM data processing pipeline - manual picking, 2D classification, and 3D classification - was another polymorph observed in the dataset, thus the polymorph described here was conclusively the only fibrillar polymorph present in the sample.

Similar to the procedure used for autorefinement of rHuPrP^{Res} 43, a bimodal tilt angle distribution was discouraged by addition of the flag `-helical_keep_tilt_prior_fixed` and setting the sigma tilt value to 1 for early iterations where a low healpix order (4, or 1.875 degrees) was used. Once a solution started to be converged upon after a few iterations, finer sampling was achieved by increasing the healpix order and the tilt values were allowed to expand to two and tilt priors were no longer fixed to 90 degrees. This achieved a smooth distribution of tilt angles about the expected 90 degrees.

After the twist value had been optimized, similar autorefinements were carried out using twist and rise values corresponding to two, three, and four rung beta solenoids but no discrepancies in layers indicative of a beta solenoid structure were detected and thus were not pursued.

After gold standard autorefinement had been carried out using two independently refined half maps, helical symmetry in real space was imposed using the central 20% of the box to ensure uniform resolution throughout the fibril. Masking and postprocessing steps were carried out in RELION 3.0 using the flag `--inimask_threshold 0.01`. Sharpening was carried out in PHENIX⁶⁴ using `phenix.auto_sharpen` to 3.4Å and a final B factor of -163.65.

Three-Dimensional Classification of rElkPrP⁹⁴⁻¹⁷⁸ Fibrils: Short Pitch

After 2D classification of short short pitched particles, classes belonging to different polymorphs were not readily distinguishable. Thus, all 77,533 particles were included in 3D classification runs with 3 classes using an initial twist value of -1.3 degrees calculated from the estimated crossover distance of 665Å measured in 2D classes with a 800 pixel (856Å) box and by stitching together 2D classes of smaller box sizes. This 3D classification yielded three distinguishable cross sections, which will subsequently be described as short class 1, 2, and 3. These three

morphologies could be more easily classified after a second round of 3D classification using the same particles, but now also using the three classes from the previous run as references and allowing the twist value to refine between -1.15 and -1.35 degrees to allow particles to separate more easily into these classes based on their slightly different twist values.

Short class 1 was composed of 28,628 particles that were subjected to 3D autorefinement with C1 symmetry. This refinement used a gaussian ellipsoid as a reference model, a 400 pixel box with 1.07Å/pixel, and a -1.25 degree twist. Due to the moderate resolution of the reconstruction (7.25Å), postprocessing and sharpening were not pursued.

Short class 2 was composed of 31,234 particles that were subjected to an additional round of 2D classification described in the section above in order to remove poor quality particles. The remaining 29,644 particles were then subjected to additional rounds of 3D classification to refine the twist and rise values and to visualize a staggering of strands indicative of a pseudo-2₁ screw axis. The twist value was further refined to 179.42 degrees and the rise refined to 2.415Å. The quality of the reconstruction obtained using several box sizes (256, 300, 400, and 512 pixels) was also assessed here, and a 256 pixel box was pursued for 3D autorefinement. Gold standard 3D autorefinement was carried out using a reference of the same polymorph from a previous 3D classification run including the same particles and helical parameters filtered to 15Å. The flag – helical_keep_tilt_prior_fixed was used to discourage particles adopting a bimodal tilt angle distribution for early refinement iterations using a low healpix order (4, 1.875 degrees). This flag was removed after a few iterations where particles began to converge to their correct orientations and the healpix order had been increased to refine particle orientations. Throughout this process, sigma tilt did not to be as restricted for early iterations as for some other dataset and was set to three.

The estimated resolution for short class 2 should have been high enough to see some large sidechain features. This estimate was 4.4Å directly out of autorefinement and 3.7Å after imposing symmetry in real space from the central 10% of the box and postprocessing using the flag --inimask_threshold 0.01. However, directional striping artifacts were visible in cross sections of the map and in maps sharpened past 4.2Å that could not be explained by severe orientation bias (Figure 4.4). These stripes are likely the result of data of insufficient quality for a high-resolution structure where noise bias has built up in the reference arising from very little mass in the box for alignment (Sjores Scheres, personal correspondence). Thus, this estimated resolution is not reflective of what protein features can be seen in the map and conclusions were not drawn beyond the shape of the main chain. No attempts were made to assign a sequence into this unreliable density.

Short class 3 was the least abundant of the short classes, and was composed of 17,671 particles. An estimated crossover distance of 754Å was obtained from a 2D classification run using a box size corresponding to 1,096Å, this translates to a -1.146 degree twist. 3D classification was carried out using a 256 pixel box and a gaussian ellipsoid as a reference model with particles allowed to split into two classes. A small initial number of particles belonging to this class did not allow for many particles to be removed without loss of signal, and 3D autorefinement was not pursued for this class. The resolution obtained during 3D classification (~7-8Å) did not allow for visualization of strand separation, but did allow two large beta strands and two small beta strands to be visible in cross sections making up the core of the fibril.

Three-Dimensional Classification of rElkPrP⁹⁴⁻¹⁷⁸ Fibrils: Medium Pitch

For particles belonging to the medium pitch thin fibril group, two rounds of 3D classification with two classes and a 256 pixel box with 1.07Å/pixel were used to try and separate out the best

particles. 3D autorefinement runs were carried out using all 65,477 particles, 33,773 particles falling into the best class after the first round of 3D classification, and 18,686 particles falling into the best class after a second round of 3D classification. All three autorefinements achieved similar results with twist values ranging from -0.663 - -0.689 degrees and estimated resolutions reaching 4.4Å after autorefinement and 3.7-4.1Å after postprocessing. Similar to short class 2, directional striping was visible in cross sections from all runs and sharpening could not be carried out without clear artifacts appearing at sharpening values past 4.2Å. Thus, these estimated resolutions are also not reflective of the true resolution of protein features in the map and conclusions beyond tracing a chain could not be drawn.

For particles belonging to the medium pitch thick fibril group, 8,737 particles were subjected to 3D classification into two classes using a 256 pixel box with 1.07Å/pixel. Two ultrastructures that appeared to be composed of the same two beta strand building blocks arranged slightly differently emerged. While chains could be traced for both of these thick filament morphologies, The very small pool of mixed quality particles did not allow for further improvements in homogeneity or resolution during subsequent rounds of 3D classification, thus 3D autorefinement was not pursued for either thick filament morphology.

Model Building Attempts for BvPrP⁹⁴⁻¹⁷⁸ Using PyRosetta

Initial attempts at sequence assignment centered on identifying the largest sidechains and manually assessing the fit of surrounding residues to the cryo-EM density, but sequence assignment was not conclusive based on visual assessment alone. The small ordered core was composed of 15 residues making up each chain out of a total of 84 residues in the protein construct. Landmark features in the density that ruled out all other sequence assignments were not apparent, and many solutions appeared equally likely. The directionality of each chain and

whether they were truly related by symmetry was also not taken for granted, further multiplying the number of possible sequence combinations.

Thus, exhaustive computational searches of sequences fit against the cryo-EM density were conducted in PyRosetta⁶⁵. Protein backbones composed of ten identical layers were built using COOT⁶⁶ into each of two chains traveling both parallel and antiparallel to one another for a total of four directionality combinations between the two sheets. All ten poly-alanine layers were iteratively refined against the cryo-EM density in PHENIX⁶⁴ with manual adjustments in COOT to ensure the best starting point for threading sidechains onto each backbone. Changes made to one layer were propagated to all ten layers throughout refinement to ensure model compatibility with adjacent protein layers. Clashscores and ramachandran angles for all four models were minimized to similar values of 8.97-9.93 and 92.31% favored respectively to minimize bias in subsequent model scores based on starting backbone geometry.

All 15 residue sequence combinations in the 84 residue segment (71 combinations) were independently thread onto each of the two strands (5041 combinations) in each of the four models (20,164 combinations) with each combination of chain directionalities. After threading, all ten layers were allowed to repack against the cryo-EM density. The six central layers were then allowed to undergo a fast-relax protocol heavily weighing fit against the cryo-EM density while the two top and bottom layers remained fixed. The repulsive term was also down weighted since amyloid structures are known to pack more tightly than the natively folded proteins the Rosetta energy function was optimized for. The middle six layers were then scored based on Rosetta energy and Fourier shell correlation (FSC) for a resolutions range indicative of the highest resolution features in the map that could be used to best assess discrepancies in sidechain fit (4.0-3.4Å).

The deviations in Rosetta energy and the FSC at 4.0-3.4Å between the best and worst scoring models were too small to confidently assign a correct sequence, thus no sequence or chain directionality assignments were made and further computational efforts would be needed to achieve this end.

References

1. Chandler, R. L. & Turefy, B. A. Inoculation of Voles, Chinese Hamsters, Gerbils and Guinea-pigs with Scrapie Brain Material. *Research in Veterinary Science* **13**, 219–226 (1972).
2. Kimberlin, R. H., Cole, S. & Walker, C. A. Y. 1987. Temporary and Permanent Modifications to a Single Strain of Mouse Scrapie on Transmission to Rats and Hamsters. *Journal of General Virology* **68**, 1875–1881.
3. Scott, M. *et al.* Transgenic mice expressing hamster prion protein produce species-specific scrapie infectivity and amyloid plaques. *Cell* **59**, 847–857 (1989).
4. Prusiner, S. B. *et al.* Transgenic studies implicate interactions between homologous PrP isoforms in scrapie prion replication. *Cell* **63**, 673–686 (1990).
5. Cartoni, C. *et al.* Identification of the pathological prion protein allotypes in scrapie-infected heterozygous bank voles (*Clethrionomys glareolus*) by high-performance liquid chromatography–mass spectrometry. *Journal of Chromatography A* **1081**, 122–126 (2005).
6. Nonno, R. *et al.* Efficient Transmission and Characterization of Creutzfeldt–Jakob Disease Strains in Bank Voles. *PLOS Pathogens* **2**, e12 (2006).
7. Pirisinu, L. *et al.* Gerstmann-Sträussler-Scheinker disease subtypes efficiently transmit in bank voles as genuine prion diseases. *Sci Rep* **6**, 20443 (2016).
8. Watts, J. C. *et al.* Evidence That Bank Vole PrP Is a Universal Acceptor for Prions. *PLOS Pathogens* **10**, e1003990 (2014).
9. Orrú, C. D. *et al.* Bank Vole Prion Protein As an Apparently Universal Substrate for RT-QuIC-Based Detection and Discrimination of Prion Strains. *PLoS Pathog* **11**, (2015).
10. Hsiao, K. K. *et al.* Spontaneous neurodegeneration in transgenic mice with mutant prion protein. *Science* **250**, 1587–1590 (1990).
11. Telling, G. C. *et al.* Interactions between wild-type and mutant prion proteins modulate neurodegeneration in transgenic mice. *Genes & Development* **10**, 1736–1750 (1996).

12. Chiesa, R., Piccardo, P., Ghetti, B. & Harris, D. A. Neurological Illness in Transgenic Mice Expressing a Prion Protein with an Insertional Mutation. *Neuron* **21**, 1339–1351 (1998).
13. Dossena, S. *et al.* Mutant Prion Protein Expression Causes Motor and Memory Deficits and Abnormal Sleep Patterns in a Transgenic Mouse Model. *Neuron* **60**, 598–609 (2008).
14. Jackson, W. S. *et al.* Spontaneous Generation of Prion Infectivity in Fatal Familial Insomnia Knockin Mice. *Neuron* **63**, 438–450 (2009).
15. Friedman-Levi, Y. *et al.* Fatal Prion Disease in a Mouse Model of Genetic E200K Creutzfeldt-Jakob Disease. *PLOS Pathogens* **7**, e1002350 (2011).
16. Watts, J. C. *et al.* Spontaneous generation of rapidly transmissible prions in transgenic mice expressing wild-type bank vole prion protein. *Proc Natl Acad Sci U S A* **109**, 3498–3503 (2012).
17. Eraña, H. *et al.* Development of a new largely scalable in vitro prion propagation method for the production of infectious recombinant prions for high resolution structural studies. *PLOS Pathogens* **15**, e1008117 (2019).
18. Martín-Pastor, M. *et al.* Structural features of an infectious recombinant PrP^{Sc} prion using solid state NMR. <http://biorxiv.org/lookup/doi/10.1101/2020.04.08.032839> (2020)
doi:10.1101/2020.04.08.032839.
19. WILLIAMS, E. S. & YOUNG, S. CHRONIC WASTING DISEASE OF CAPTIVE MULE DEER: A SPONGIFORM ENCEPHALOPATHY¹. *Journal of Wildlife Diseases* **16**, 89–98 (1980).
20. Stockman, S. Scrapie: An Obscure Disease of Sheep. *Journal of Comparative Pathology and Therapeutics* **26**, 317–327 (1913).
21. Jakob, A. Über eigenartige erkrankungen des zentralnervensystems mit bemerkenswertem anatomischen befunde: (Spastische pseudosklerose — encephalomyelopathie mit disseminirten degenerationsherden.). *Z. f. d. g. Neur. u. Psych.* **64**, 147–228 (1921).

22. Zigas, V. & Gajdusek, D. C. Kuru: Clinical Study of a New Syndrome resembling Paralysis Agitans in Natives of the Eastern Highlands of Australian New Guinea. *Medical Journal of Australia* **2**, 745–54 (1957).
23. Wells, G. a. H. *et al.* A novel progressive spongiform encephalopathy in cattle. *Veterinary Record* **121**, 419–420 (1987).
24. Saunders, S. E., Bartelt-Hunt, S. L. & Bartz, J. C. Occurrence, Transmission, and Zoonotic Potential of Chronic Wasting Disease. *Emerg Infect Dis* **18**, 369–376 (2012).
25. Brown, P. W. & Detwiler, L. A. Bovine Spongiform Encephalopathy. in *Food Microbiology* 651–671 (John Wiley & Sons, Ltd, 2012). doi:10.1128/9781555818463.ch25.
26. Spraker, T. R. *et al.* SPONGIFORM ENCEPHALOPATHY IN FREE-RANGING MULE DEER (*ODOCOILEUS HEMIONUS*), WHITE-TAILED DEER (*ODOCOILEUS VIRGINIANUS*) AND ROCKY MOUNTAIN ELK (*CERVUS ELAPHUS NELSONI*) IN NORTHCENTRAL COLORADO. *Journal of Wildlife Diseases* **33**, 1–6 (1997).
27. Williams, E. S., Miller, M. W., Kreeger, T. J., Kahn, R. H. & Thorne, E. T. Chronic Wasting Disease of Deer and Elk: A Review with Recommendations for Management. *The Journal of Wildlife Management* **66**, 551–563 (2002).
28. WILLIAMS, E. S. & YOUNG, S. SPONGIFORM ENCEPHALOPATHY OF ROCKY MOUNTAIN ELK¹. *Journal of Wildlife Diseases* **18**, 465–471 (1982).
29. Sohn, H.-J. *et al.* A Case of Chronic Wasting Disease in an Elk Imported to Korea from Canada. *Journal of Veterinary Medical Science* **64**, 855–858 (2002).
30. Kim, T.-Y. *et al.* Additional Cases of Chronic Wasting Disease in Imported Deer in Korea. *Journal of Veterinary Medical Science* **67**, 753–759 (2005).
31. Rivera, N. A., Brandt, A. L., Novakofski, J. E. & Mateus-Pinilla, N. E. Chronic Wasting Disease In Cervids: Prevalence, Impact And Management Strategies. *Vet Med (Auckl)* **10**, 123–139 (2019).

32. Occurrence | Chronic Wasting Disease (CWD) | Prion Disease | CDC.
<https://www.cdc.gov/prions/cwd/occurrence.html> (2021).
33. Baeten, L. A., Powers, B. E., Jewell, J. E., Spraker, T. R. & Miller, M. W. A Natural Case of Chronic Wasting Disease in a Free-ranging Moose (*Alces alces shirasi*). *Journal of Wildlife Diseases* **43**, 309–314 (2007).
34. Benestad, S. L., Mitchell, G., Simmons, M., Ytrehus, B. & Vikøren, T. First case of chronic wasting disease in Europe in a Norwegian free-ranging reindeer. *Veterinary Research* **47**, 88 (2016).
35. Miller, M. W. *et al.* Epizootiology of chronic wasting disease in free-ranging cervids in colorado and wyoming. *Journal of Wildlife Diseases* **36**, 676–690 (2000).
36. Bocharova, O. V., Breydo, L., Salnikov, V. V., Gill, A. C. & Baskakov, I. V. Synthetic prions generated in vitro are similar to a newly identified subpopulation of PrPSc from sporadic Creutzfeldt-Jakob Disease. *Protein Sci* **14**, 1222–1232 (2005).
37. Wopfner, F. *et al.* Analysis of 27 mammalian and 9 avian PrPs reveals high conservation of flexible regions of the prion protein 11Edited by A. R. Fersht. *Journal of Molecular Biology* **289**, 1163–1178 (1999).
38. Race, R. E. *et al.* Comparison of Abnormal Prion Protein Glycoform Patterns from Transmissible Spongiform Encephalopathy Agent-Infected Deer, Elk, Sheep, and Cattle. *Journal of Virology* **76**, 12365–12368 (2002).
39. Perrott, M. R., Sigurdson, C. J., Mason, G. L. & Hoover, E. A. Evidence for distinct chronic wasting disease (CWD) strains in experimental CWD in ferrets. *Journal of General Virology* **93**, 212–221 (2012).
40. Raymond, G. J. *et al.* Transmission and Adaptation of Chronic Wasting Disease to Hamsters and Transgenic Mice: Evidence for Strains. *Journal of Virology* **81**, 4305–4314 (2007).

41. Nonno, R. *et al.* Studies in bank voles reveal strain differences between chronic wasting disease prions from Norway and North America. *PNAS* **117**, 31417–31426 (2020).
42. Moore, J. *et al.* Novel Strain of the Chronic Wasting Disease Agent Isolated From Experimentally Inoculated Elk With LL132 Prion Protein. *Sci Rep* **10**, 3148 (2020).
43. Glynn, C. *et al.* Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. *Nature Structural & Molecular Biology* 1–7 (2020) doi:10.1038/s41594-020-0403-y.
44. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. Evidence for assembly of prions with left-handed β -helices into trimers. *PNAS* **101**, 8342–8347 (2004).
45. Ghetti, B. *et al.* Vascular variant of prion protein cerebral amyloidosis with tau-positive neurofibrillary tangles: the phenotype of the stop codon 145 mutation in PRNP. *PNAS* **93**, 744–748 (1996).
46. Piccardo, P. *et al.* Prion Proteins with Different Conformations Accumulate in Gerstmann-Sträussler-Scheinker Disease Caused by A117V and F198S Mutations. *Am J Pathol* **158**, 2201–2207 (2001).
47. Theint, T. *et al.* Species-dependent structural polymorphism of Y145Stop prion protein amyloid revealed by solid-state NMR spectroscopy. *Nature Communications* **8**, 1–10 (2017).
48. Westaway, D. New approaches to CWD prevention: A conversation on possibilities. (2019).
49. Zhang, W. *et al.* Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *eLife* **8**, e43584 (2019).
50. Wan, W. *et al.* Structural Studies of Truncated Forms of the Prion Protein PrP. *Biophysical Journal* **108**, 1548–1554 (2015).
51. Gao, Y., Tran, P., Petkovic-Duran, K., Swallow, T. & Zhu, Y. Acoustic micromixing increases antibody-antigen binding in immunoassays. *Biomed Microdevices* **17**, 79 (2015).

52. Nagapudi, K., Umanzor, E. Y. & Masui, C. High-throughput screening and scale-up of cocrystals using resonant acoustic mixing. *International Journal of Pharmaceutics* **521**, 337–345 (2017).
53. Warmack, R. A. *et al.* Structure of amyloid- β (20-34) with Alzheimer's-associated isomerization at Asp23 reveals a distinct protofilament interface. *Nat Commun* **10**, 3357 (2019).
54. Mastronarde, D. N. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microscopy and Microanalysis* **9**, 1182–1183 (2003).
55. Carragher, B. *et al.* Legion: An Automated System for Acquisition of Images from Vitreous Ice Specimens. *Journal of Structural Biology* **132**, 33–45 (2000).
56. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
57. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
58. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1–12 (2016).
59. Schmidt, M. *et al.* Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat Commun* **10**, 5008 (2019).
60. Röder, C. *et al.* Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid- β fibrils. *Nat Struct Mol Biol* **27**, 660–667 (2020).
61. Gallardo, R. *et al.* Fibril structures of diabetes-related amylin variants reveal a basis for surface-templated assembly. *Nat Struct Mol Biol* **27**, 1048–1056 (2020).
62. Hervas, R. *et al.* Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in *Drosophila*. *Science* **367**, 1230–1234 (2020).

63. Bansal, A. *et al.* AA amyloid fibrils from diseased tissue are structurally different from in vitro formed SAA fibrils. *Nat Commun* **12**, 1013 (2021).
64. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* **66**, 213–221 (2010).
65. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
66. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst D* **66**, 486–501 (2010).

Conclusions, Discussion, and Looking Forward

Technological Advancements Allow for Visualization of Prion Cores

The uncovering of amyloid structures with atomic detail has been the result of decades of technological advancements. The first two chapters described how the boundaries of crystallography were pushed by micro-crystal electron diffraction (microED). Peptide structures arising from microfocus x-ray beamlines at synchrotrons and from electron microscopes allowed for visualization of the types of interactions present in amyloids that bestow them with their trademark stability and longevity. The steric zippers first described by Michael Sawaya and David Eisenberg¹ consisted of interdigitating polar residues whose hydrogen bonds could be satisfied through long stacks of polar ladders²⁻⁴ along the entire length of the amyloid fibril. Besides interdigitating polar residues, hydrophobic residues were also found to form tightly packed steric zippers where their sidechains could be buried away from solvents. In chapters 1 and 2, an intricate three-dimensional hydrogen bonding network, termed a polar clasp, was described. The clasps are composed of two interlocking polar ladders and illustrate how hydrogen bonds formed between neighboring residues on the same chain could be another form of stabilizing interaction in amyloids that blocks access to the peptide backbone by solvents, denaturants, or proteases.

In chapter 2, I illustrated how the same amyloid structure may be accessible to prion $\beta 2\alpha 2$ loop peptides, but in different chemical environments and with different stabilities. The reluctance to form the same structure under the same conditions may offer the beginnings of an explanation for the species barriers observed for mammalian prions of different origin species. I also demonstrated how single residue substitutions can impact more than just their own interactions and have an influence that propagates across the peptide chain to favor new interactions and create new interfaces between protofilament chains. Here, the influence of this residue - 174 - could not have been inferred from a single structure alone. In the crystal structure of bank vole PrP¹⁶⁸⁻¹⁷⁶ presented in chapter 1⁵, substituting N174 for an S or T would not be strictly prohibited.

This substitution would not have induced any clashes that would force a structural rearrangement, but nonetheless a new structure was taken on for peptides containing S or T174 in place of N. This highlights the importance of more than just the endpoint structure itself in deciding the fold an amyloid will take on.

The Cryo-EM Resolution Revolution Reaches Prions

The theme of incorporating new technology and methodology in pursuit of amyloid structure continued in chapters 3 and 4 through utilization of single particle cryo-electron microscopy (cryo-EM). While the cryo-EM resolution revolution began a few years earlier, the revolution made its mark on the amyloid field starting in 2017. More elaborate amyloid structures than ever before started to be resolved with atomic detail, both for tissue-derived and *in vitro* reconstituted amyloid fibrils. In chapter 3, I presented the first of these structures from a recombinant human prion fibril with chaotrope-, protease-, and SDS-resistant properties. In chapter 4, I demonstrated that bank vole and elk prions are also capable of spontaneously forming fibrils in the absence of a seed strain and form structures distinct from those taken on by the same stretch of residues of the human prion protein. In the subsequent sections, all cryo-EM structures of prions resolved to date will be discussed with implications for strains, biophysical properties, and species barriers discussed.

Comparison to Other Prion Structures: Recombinant, Infectious Y145Stop from Human, Animal, and Familial Disease Associated Variants

While the most common prion diseases are associated with aggregation of full length PrP into 27-30kDa protease-resistant aggregates, there are prion diseases associated with PrP truncations and accumulation of much smaller 6-8kDa protein fragments spanning residues ~90-145 with

ragged N and C termini⁶. The most noteworthy of these diseases is the human familial prion disease Gerstmann–Straüssler–Scheinker disease (GSS), but animal equivalents with aggregates composed of the same comparably small stretch of residues exist⁷. The naturally occurring and disease associated Y145Stop truncation has been characterized extensively by solid state nuclear magnetic resonance (ssNMR)^{8–10} and has been demonstrated to be capable of inducing disease in mice after intracerebral injection¹¹. Based on this ssNMR data, human, mouse, and syrian hamster prion fibrils share a common ordered region spanning residues 112–140. Residues 112 and 139 were also shown to have a disproportionate influence on the structure of the resulting aggregates. A genetic match at these residues allowed for faithful reconstitution of known species barriers between human, mouse, and syrian hamster prions⁸.

The prion constructs used in these ssNMR and mouse experiments were similar to those used in chapters 3 and 4 (residues 23–144 there and 94–178 here) and produced a core composed of nearly the same residues (112–140 there and 106–145 here). Both structures are modelled to include two symmetry related hairpin-like structures coming together to form a tight interface, and observations about the effect of mutations on structure and species barriers can be understood through the structure presented in chapter 3¹².

Most recently, a cryo-EM structure of fibrils formed from human PrP^{23–144} was determined¹³. Fibrils were prepared in a way similar, but not identical, to those of human prion fibrils studied in ssNMR experiments^{8,9} and in mouse experiments where the resulting fibrils were capable of inducing disease¹¹. The ordered core of these fibrils was composed of residues 108–141 – a near identical stretch to residues most clearly resolved in ssNMR experiments and those seen in the cryo-EM structure presented in chapter 3. Despite inclusion of the same residues in the core, the sharing of two short beta sheets composed of residues 133–135 and 138–140, and many of the same residues being inaccessible to solvents or in positions that would have relevant consequences for

species barriers, cryo-EM revealed two entirely different architectures for the fibrils used in each study. These two structures taken together highlight the potential for many different prion structures that share the same features and properties to exist.

Despite a similar fragment of mouse sequence being capable of inducing disease in mice¹¹, the ability of either of these recombinant human fibril polymorphs to either induce disease or demonstrate infectivity has yet to be tested. The relevance and consequences for disease and infectivity of different atomic structures with many indistinguishable structural features and biophysical properties has yet to be realized, but may be useful for identifying and designing prions in the future.

Many mutations known to be associated with the familial prion disease GSS cannot be accommodated in the HuPrP^{Res} structure presented in chapter 3¹². In agreement with this, ssNMR experiments have recently demonstrated that GSS associated mutation A117V and the naturally occurring sequence polymorphism M129V influence the size and sheets making up the ordered core of prion fibrils¹⁰. The ordered core shrank from residues 112-140 in wild type human PrP to 120-139 for A117V and 121-139 for M129V variants. This illustrates that these mutations are both predicted to, and do in fact, alter the conformation and flexibility of the resulting fibril cores.

Also based on ssNMR, substitution of human I138 for mouse M138 - a solvent facing residue in HuPrP^{Res} - did not alter structure or species barriers while substitution of human I139 to syrian hamster M139 - an inward facing residue in HuPrP^{Res} that could not accommodate such a substitution - did alter structure and species-specific infectivity relationships to more closely match patterns observed for syrian hamster prions⁸.

Comparison to Other Prion Structures: Full Length Recombinant Human PrP of Wild Type and Familial Mutant E196K Sequences

Fibrils formed by the full length human prion protein under nonreducing conditions have shed light on a new polymorph with the C terminal domain at its core¹⁴. Likewise, fibrils formed by the full length human prion protein containing the familial disease associated E196K mutation also favor a core composed of the C terminal domain¹⁵. However, the structure taken on is readily distinguishable from the structure favored by the aforementioned wild type full length human prion protein. The structure presented in chapter 3 is composed of residues 106-145, while these two structures are composed of residues 170-229 and 175-217 respectively. Thus each section - the N terminal and C terminal region - is able to form a large compact structure with a sizable buried surface area on it's own without the other half of the protein. This leads to the hypothesis that prions may be composed of two potential semi-independent amyloid cores, one composed of the N terminal region and the other the C terminal region of the protein. The residues not seen in any of the structures described thusfar - residues 146-169 - includes a portion of the $\beta 2\alpha 2$ loop that may influence how the two core modules come together and form species-specific structures. At this time, this hypothesis is purely speculation, but may have some support from a prion structure formed by the 263K scrapie strain isolated from hamsters¹⁶ to be discussed in a subsequent section.

Familial Mutations: General Trends and Implications for Prion Structure

Both works using A117V and M129V¹⁰ as well as E196K variants¹⁵ bring to light a pattern for familial prion disease associated variants - a smaller ordered core structure. The ordered region of the N terminal core shifts from ~30 residues down to ~20 residues when the A117V or M129V mutation is introduced and the C terminal core shifts from ~60 residues down to ~40 residues.

More structures will be needed from both recombinant and tissue sources to find out whether this is mere coincidence or a meaningful trend for prion structures in general.

Comparison to Other Prion Structures: Bank Vole and Elk Prions

A single residue has been shown to alter the fibril structure of not only human prions, but also other amyloid forming proteins^{17,23}. It should serve as no surprise, therefore, that animal prions - which have several residue differences from human prions - could take on different structures from human prions. In chapter 4 I demonstrated that bank vole and elk prion fibrils generated under similar conditions produce distinct structures from one another that are also distinct from human prions formed by the same stretch of residues. Both bank vole and elk prion core structures are composed of two straight layers of beta strands compared to the four layers making up the core of HuPrP^{Res}. The smaller, less protected core of recombinant bank vole PrP⁹⁴⁻¹⁷⁸ (rBvPrP⁹⁴⁻¹⁷⁸) fibrils studied here are also protease-sensitive while HuPrP^{Res} fibrils are not, which can be explained by the structures even without resolutions compatible with residue assignment. The flexibility seen in recombinant elk PrP⁹⁴⁻¹⁷⁸ (rElkPrP⁹⁴⁻¹⁷⁸) fibrils is likely a common feature of amyloids that was more commonly observed in the 1990s and 2000s^{24,26} before high resolution cryo-EM compatible with residue assignment became commonplace for amyloids. The contribution of flexibility may be being unintentionally erased, and the disease- and species barrier-relevance of core flexibility remains to be explored.

Comparison to Other Prion Structures: Infectious 263K Scrapie Strain Isolated from Hamsters

Recently, fibrils found in the 263K scrapie strain were isolated from hamsters and had their structure determined by cryo-EM¹⁶. These fibrils were then intracerebrally injected back into mice after manipulations required for cryo-EM and were found to remain infectious. The core of these

fibrils includes residue 95-227 - the entire canonical protease-resistant core of PrP^{Sc}. The structure has two domains - the N terminal and C terminal - linked by the β 2 α 2 loop.

Still, the question of what in the structure encodes for infectivity remains open. Speculation introduces many possible explanations, one being the large 133 residue core. However, the sheer size of the ordered core alone is insufficient to explain infectivity, as a recently determined structure of TDP-43 with 139 residues making up the core²⁷ is not considered infectious. Likewise, much smaller prion cores generated from recombinant Y145Stop truncations with residues 112-140 making up the ordered core have been shown to be infectious when injected intracerebrally¹¹. These constructs were generated without the two glycosylation sites at residues 181 and 197, thus these infectious fibrils were unglycosylated and rule out incorporation of glycans as a sole cause of infectivity.

The structure of 263K scrapie prions displays asymmetry at the poles of the fibril¹⁶. Each chain spans multiple layers allowing for interactions between *i*, *i*-1, and *i*-2 layers, but these types of multilayer interactions have also been shown for other amyloid structures²⁷⁻³⁰ and are also insufficient to explain the infectious nature of prions on their own. The 263K scrapie prions also included a GPI anchor which is hypothesized to interact with cell membranes to induce membrane spiraling¹⁶, but this behavior cannot explain how GPI anchorless prion strains are also infectious^{11,31,32}. One such strain, the anchorless Rocky Mountain Lab (aRML) mouse prion strain lacks a GPI anchor and N-linked glycans. The structure formed by aRML prions isolated from mouse brain also harbors a large cross section that appears to be composed of a single protofilament¹⁶, but the ability to trace a chain with certainty remains just out of reach, preventing a detailed comparison between this strain and others at the residue level.

The surface of 263K prions is unusually polar with a large number of unbalanced positive charges on the N terminal half of the fibril and unbalanced negative charges on the C terminal half of the fibril. The protein is folded in such a way that one face of the fibril carries a net positive charge while the other face carries a net negative charge. While unbalanced charges are not uncommon for brain-derived amyloids³³⁻³⁶, observations of each face of a fibril carrying a net negative or net positive charge have been largely confined to systemic amyloids^{37,38} and have all been less extreme than the polarity seen in 263K prions and this feature may set this structure apart from other amyloids.

Lastly, 263K prions are asymmetric and composed of two core modules made from the N and C terminal halves of the protein. Each half is capable of making fibrils on its own with hairpin structures similar to those seen in 263K prions at the core^{12,14,16}. A modular core like this where each half of the protein can form its own core has been proposed for FUS^{39,40}, but thus far both cores have not been shown to exist concurrently as in 263K prions. Likewise, fibrils of the low complexity domain of TDP-43 also appear to have a two module core²⁷ where the N terminal half has been shown to form fibrils on its own³⁰. The modular cores of TDP-43 and FUS in combination with the infectious nature of prion truncations lacking the C terminal module illustrate that a multi module core is also an incomplete explanation for the infectious properties seen in prions. Without more structures from both known infectious and tested and shown to be noninfectious prion aggregates, a definitive explanation for what makes a structure infectious cannot be put forward. It is also possible that no single factor described above can engender the binary declaration of infectivity on its own, and infectivity is better described as a spectrum with many contributing factors. Describing infectivity as a spectrum rather than a binary characteristic is supported by the range of transmissibility observed for different prion diseases.

The Two Core Hypothesis: Implications for Strains

It is well established that different prion strains take on different biophysical properties that manifest as different clinical, pathological, and transmission related behaviors⁴¹. It is to be expected, therefore, that each human or animal prion disease strain may be associated with a unique structural fingerprint. This fingerprint may be modified by regions such as the $\beta 2\alpha 2$ loop, a segment that varies between species, is highly amyloidogenic, and found at the interface of the two core modules comprising 263K scrapie prions¹⁶. The bank vole $\beta 2\alpha 2$ loop structure presented in chapter 1 showed a core prone to forming stable fibrils regardless of growth conditions⁵. This $\beta 2\alpha 2$ loop sequence is shared between bank voles and hamsters, and it is easy to understand why such a segment would serve as a glue holding together N and C terminal prion cores in a fibril. In chapter 2, I showed that the amyloid formed by the $\beta 2\alpha 2$ loop is not equally stable for all species, and thus fibrils with different loop sequences are unlikely to take on the same arrangement as 263K prions under the same physiological conditions, even when mutation would not introduce steric clashes.

Like the tau structures with different isoform compositions isolated from patients with different diseases^{42,43}, different forms of the PrP are also likely incorporated into fibrils in a selective manner. Un-, mono-, and di-glycosylated PrP species are known to be present in different ratios in different disease strains⁴⁴. Prion strains with and without the GPI anchor as well as those with a Y145Stop truncation are also known to have different behavior even when no other modifications to the protein are present. The number of strains - and possible structures - is compounded by the number of species susceptible to prion disease as well as familial variants with alterations to PrP sequence.

In summary, a tremendous amount of progress has been made in determining prion structures during the course of this dissertation. It remains to be seen what in a structure encodes for

infectivity, but the answer may come soon with technological advancements allowing for atomic structures of prion strains to come to light. Challenges in how different structures induce different diseases, mechanism of spontaneous misfolding, and mechanisms of transmission are all outstanding questions that stand to benefit from the herculean efforts currently being made in prion structural biology.

References

1. Sawaya, M. R. *et al.* Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453 (2007).
2. Perutz, M. F., Staden, R., Moens, L. & De Baere, I. Polar zippers. *Current Biology* **3**, 249–253 (1993).
3. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proceedings of the National Academy of Sciences* **91**, 5355–5358 (1994).
4. Stott, K., Blackburn, J. M., Butler, P. J. & Perutz, M. Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proceedings of the National Academy of Sciences* **92**, 6509–6513 (1995).
5. Gallagher-Jones, M. *et al.* Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nature Structural & Molecular Biology* **25**, 131 (2018).
6. Piccardo, P. *et al.* Prion Proteins with Different Conformations Accumulate in Gerstmann-Sträussler-Scheinker Disease Caused by A117V and F198S Mutations. *Am J Pathol* **158**, 2201–2207 (2001).
7. Benestad, S. L., Arsac, J.-N., Goldmann, W. & Nöremark, M. Atypical/Nor98 scrapie: properties of the agent, genetics, and epidemiology. *Vet. Res.* **39**, 19 (2008).
8. Theint, T. *et al.* Species-dependent structural polymorphism of Y145Stop prion protein amyloid revealed by solid-state NMR spectroscopy. *Nature Communications* **8**, 1–10 (2017).
9. Theint, T. *et al.* Structural Studies of Amyloid Fibrils by Paramagnetic Solid-State Nuclear Magnetic Resonance Spectroscopy. *J. Am. Chem. Soc.* **140**, 13161–13166 (2018).
10. Dao, H. H. *et al.* ^{13}C and ^{15}N chemical shift assignments of A117V and M129V human Y145Stop prion protein amyloid fibrils. *Biomol NMR Assign* **15**, 45–51 (2021).

11. Choi, J.-K. *et al.* Amyloid fibrils from the N-terminal prion protein fragment are infectious. *PNAS* **113**, 13851–13856 (2016).
12. Glynn, C. *et al.* Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. *Nature Structural & Molecular Biology* 1–7 (2020) doi:10.1038/s41594-020-0403-y.
13. Li, Q., Jaroniec, C. P. & Surewicz, W. K. *Cryo-EM structure of disease-related prion fibrils provides insights into seeding barriers.* 2021.08.10.455830
<https://www.biorxiv.org/content/10.1101/2021.08.10.455830v1> (2021)
doi:10.1101/2021.08.10.455830.
14. Wang, L.-Q. *et al.* Cryo-EM structure of an amyloid fibril formed by full-length human prion protein. *Nat Struct Mol Biol* **27**, 598–602 (2020).
15. Wang, L.-Q. *et al.* Familial prion disease-related mutation E196K displays a novel amyloid fibril structure revealed by cryo-EM. *bioRxiv* 2021.02.18.431846 (2021)
doi:10.1101/2021.02.18.431846.
16. Kraus, A. *et al.* High-resolution structure and strain comparison of infectious mammalian prions. *Molecular Cell* (2021) doi:10.1016/j.molcel.2021.08.011.
17. Boyer, D. R. *et al.* Structures of fibrils formed by α -synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nat Struct Mol Biol* **26**, 1044–1052 (2019).
18. Sun, Y. *et al.* Cryo-EM structure of full-length α -synuclein amyloid fibril with Parkinson's disease familial A53T mutation. *Cell Res* **30**, 360–362 (2020).
19. Ni, X., McGlinchey, R. P., Jiang, J. & Lee, J. C. Structural Insights into α -Synuclein Fibril Polymorphism: Effects of Parkinson's Disease-Related C-Terminal Truncations. *J Mol Biol* **431**, 3913–3919 (2019).
20. Boyer, D. R. *et al.* The α -synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *PNAS* **117**, 3592–3602 (2020).

21. Zhao, K. *et al.* Parkinson's disease associated mutation E46K of α -synuclein triggers the formation of a distinct fibril structure. *Nat Commun* **11**, 2643 (2020).
22. Zhao, K. *et al.* Parkinson's disease-related phosphorylation at Tyr39 rearranges α -synuclein amyloid fibril structure revealed by cryo-EM. *PNAS* **117**, 20305–20315 (2020).
23. Gallardo, R. *et al.* Fibril structures of diabetes-related amylin variants reveal a basis for surface-templated assembly. *Nat Struct Mol Biol* **27**, 1048–1056 (2020).
24. Schmidt, M. *et al.* Comparison of Alzheimer A β (1–40) and A β (1–42) amyloid fibrils reveals similar protofilament structures. *PNAS* **106**, 19813–19818 (2009).
25. Zhang, R. *et al.* Interprotofilament interactions between Alzheimer's A β 1–42 peptides in amyloid fibrils revealed by cryoEM. *PNAS* **106**, 4653–4658 (2009).
26. Fändrich, M., Meinhardt, J. & Grigorieff, N. Structural polymorphism of Alzheimer A β and other amyloid fibrils. *Prion* **3**, 89–93 (2009).
27. Li, Q., Babinchak, W. M. & Surewicz, W. K. Cryo-EM structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat Commun* **12**, 1620 (2021).
28. Gremer, L. *et al.* Fibril structure of amyloid- β (1–42) by cryo-electron microscopy. *Science* **358**, 116–119 (2017).
29. Röder, C. *et al.* Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy. *Nat Commun* **10**, 3754 (2019).
30. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat Struct Mol Biol* **26**, 619–627 (2019).
31. Vázquez-Fernández, E. *et al.* The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy. *PLOS Pathogens* **12**, e1005835 (2016).
32. Chesebro, B. *et al.* Anchorless Prion Protein Results in Infectious Amyloid Disease Without Clinical Scrapie. *Science* **308**, 1435–1439 (2005).
33. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).

34. Falcon, B. *et al.* Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
35. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
36. Zhang, W. *et al.* Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
37. Rademaker, L. *et al.* Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nat Commun* **10**, 1103 (2019).
38. Schmidt, M. *et al.* Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nat Commun* **10**, 5008 (2019).
39. 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).
40. Lee, M., Ghosh, U., Thurber, K. R., Kato, M. & Tycko, R. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. *Nat Commun* **11**, 5735 (2020).
41. Bruce, M. E. TSE strain variation: An investigation into prion disease diversity. *British Medical Bulletin* **66**, 99–108 (2003).
42. Scheres, S. H., Zhang, W., Falcon, B. & Goedert, M. Cryo-EM structures of tau filaments. *Current Opinion in Structural Biology* **64**, 17–25 (2020).
43. Shi, Y. *et al.* Structure-based Classification of Tauopathies. *bioRxiv* 2021.05.28.446130 (2021) doi:10.1101/2021.05.28.446130.
44. Lawson, V. A., Collins, S. J., Masters, C. L. & Hill, A. F. Prion protein glycosylation. *Journal of Neurochemistry* **93**, 793–801 (2005).

APPENDIX

Table 1: All protein constructs

Plasmid Short Name	Construct	Parent Plasmid	Plasmid Resistance	Prion Protein Length (Residues)	Sequence	Expressed	Purified	Fibrils Formed	Cryo Screening Pursued
HuPrP13	His-MBP-TEV-HuPrP94-178	#	Amp	85	Y	Y	Y		
HuPrP14	NoTag-HuPrP94-178	p4995*	Kan	85	Y	Y	Y	Y	Y
HuPrP15	NoTag-HuPrP94-178_M129V	p4995	Kan	85	Y	Y	Y	Y	
HuPrP16	NoTag-HuPrP94-178_	p4995	Kan	85					
HuPrP17	NoTag-HuPrP94-178_A117V	p4995	Kan	85					
HuPrP18	NoTag-HuPrP94-178_G127V	p4995	Kan	85	Y	Y	Y	Y	Y
HuPrP19	NoTag-HuPrP82-178_G114V_M129V	p4995	Kan	97	Y	Y	Y	Y	
HuPrP20	NoTag_HuPrP82-178_A117V_M129V	pET24a+	Kan	97	Y	Y	Y		
HuPrP21	NoTag_HuPrP82-178_P102L	pET24a+	Kan	97	Y	Y			
HuPrP22	NoTag_HuPrP82-178_P105L	pET24a+	Kan	97	Y	Y			
BvPrP13	His-MBP-TEV-BvPrP94-178	#	Amp	85	Y				
BvPrP14	NoTag_BvPrP94-178	p4995	Kan	85	Y	Y	Y	Y	Y
BvPrP15	NoTag_BvPrP94-178_N174T	p4995	Kan	85	Y	Y	Y	Y	
BvPrP16	NoTag_BvPrP94-178_N174S	p4995	Kan	85	Y	Y	Y	Y	Y
BvPrP17	NoTag_BvPrP94-178_Q168E_N170S	p4995	Kan	85	Y	Y	Y	Y	Y
BvPrP18	NoTag_BvPrP94-178_N170S	p4995	Kan	85	Y	Y	Y	Y	
BvPrP19	NoTag_BvPrP94-178_N170Q_N174Q	p4995	Kan	85	Y	Y	Y		
BvPrP20	NoTag_BvPrP94-178_N170S_N174S	p4995	Kan	85	Y	Y	Y		
BvPrP21	MGSDKI-His-TEV-BvPrP23-230	pET24a+	Kan	208	Y	Y	Y	Y	
ElkPrP13	His-MBP-TEV-ElkPrP94-178		Amp	85					
ElkPrP02	NoTag_ElkPrP94-178	p4995	Kan	85	Y	Y	Y	Y	Y
RbPrP01	NoTag_RbPrP94-178	p4995	Kan	85	Y	Y	Y	Y	Y
CowPrP01	NoTag_CowPrP94-178	p4995	Kan	85	Y	Y	Y	Y	

#vectors inherited from Heather McFarlene of the Eisenberg lab with gblock (IDT) for HuPrP⁹⁴⁻¹⁷⁸ inserted between known cut sites after a His-MBP-TEV tag SacI and XhoI

*p4995 is a custom pET24a+ derivative lacking purification or solubility tags generated at UCLA

Appendix Chapter 1

MicroED Structures from Micrometer Thick Protein Crystals

The work described in this chapter has been reproduced from:

Michael W. Martynowycz, Calina Glynn, Jennifer Miao, M. Jason de la Cruz, Johan Hattne, Dan Shi, Duilio Cascio, Jose Rodriguez, Tamir Gonen. "MicroED Structures from Micrometer Thick Protein Crystals." *bioRxiv*, 2017.

Copyright 2017

Michael W. Martynowycz, Calina Glynn, Jennifer Miao, M. Jason de la Cruz, Johan Hattne, Dan Shi, Duilio Cascio, Jose Rodriguez, Tamir Gonen.

MicroED Structures from Micrometer Thick Protein Crystals

Michael W. Martynowycz¹, Calina Glynn², Jennifer Miao², M. Jason de la Cruz¹, Johan Hattne¹, Dan Shi¹, Duilio Cascio³, Jose Rodriguez², and Tamir Gonen^{1,4*}

¹ Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA

² Department of Chemistry and Biochemistry; UCLA-DOE Institute; University of California, Los Angeles; Los Angeles, California 90095-1570, USA

³ Department of Biological Chemistry, UCLA-DOE Institute; University of California, Los Angeles; Los Angeles, California 90095-1570, USA

⁴ Departments of Biological Chemistry and Physiology, David Geffen Medical School, University of California, Los Angeles, California 90095-1570, USA.

* Correspondence should be addressed to T.G. (tgonen@ucla.edu)

Summary

Atomic resolution protein structures can be determined by MicroED from crystals that surpass the theoretical maximum thickness limit by an order of magnitude.

Abstract

Theoretical calculations suggest that crystals exceeding 100 nm thickness are excluded by dynamical scattering from successful structure determination using microcrystal electron diffraction (MicroED). These calculations are at odds with experimental results where MicroED structures have been determined from significantly thicker crystals. Here we systematically evaluate the influence of thickness on the accuracy of MicroED intensities and the ability to determine structures from protein crystals one micrometer thick. To do so, we compare *ab initio* structures of a human prion protein segment determined from thin crystals to those determined from crystals up to one micrometer thick. We also compare molecular replacement solutions from crystals of varying thickness for a larger globular protein, proteinase K. Our results indicate that structures can be reliably determined from crystals at least an order of magnitude thicker than previously suggested by simulation, opening the possibility for an even broader range of MicroED experiments.

Main Text

Introduction

Electrons interact with matter more strongly than X-rays and offer a larger fraction of useful, elastic scattering events to inelastic scattering events (1). These properties are leveraged by the cryoEM method, MicroED, for structure determination at atomic resolution from nanoscale protein crystals (2). With this method, diffraction is measured from protein nanocrystals in a frozen-hydrated state (3) using a low dose electron beam (typically $\sim 0.01 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$) (4–6). Crystals are continuously and unidirectionally rotated in the beam while diffraction images are collected as a movie on a fast detector (7). A number of previously unknown as well as known structures have been determined by this method to atomic resolution (8).

The strong interaction between electrons and materials also allows for multiple scattering events to take place before electrons exit the specimen (9). According to dynamical scattering theory (10), multiple scattering events produce inaccuracies in the recorded reflections, potentially preventing the solution of structures. Simulations suggest that with crystals thicker than 50-100 nm (11, 12) dynamical scattering can be severe, resulting in nearly random intensities, where the relationship between the intensity and structure factor no longer holds true (10). However, these simulations assume diffraction is recorded from a perfect and stationary crystal - real macromolecular crystals are not perfect. In fact, precession electron diffraction can avoid many of the artifacts associated with dynamical scattering from near perfect crystals of inorganic material by pivoting of the electron beam around the crystal (13).

Similarly, by employing continuous rotation MicroED (7), useful diffraction is routinely collected from protein crystals hundreds of nanometers thick (14). Even 1.5 μm -thick crystals of lysozyme were shown to produce diffraction that when integrated produced reasonable statistics (2). Moreover, recent structures determined by *ab initio* methods to 1Å resolution further indicate that the diffraction intensities obtained by continuous rotation MicroED are accurate and maintain the relationship between amplitude and phase (4, 15).

Here we systematically investigate the relationship between crystal thickness, dynamical scattering and the quality of structure solutions obtained by MicroED. We determine *ab initio* structures of a segment from the β_2 - α_2 loop of human prion protein in the amyloid state as well as the structure of proteinase K by molecular replacement from crystals up to a micrometer thick.

Results

A comparison of structures determined from thin and thick crystals of a segment of the β_2 - α_2 loop of human prion protein. As a model system for evaluating the effects of crystal thickness on diffraction intensities, we use a segment from the β_2 - α_2 loop of human prion protein (hPrP) that contains a glycine residue at its amino terminus (sequence GSNQNNF), hereafter referred to as hPrP- $\beta_2\alpha_2$, for MicroED structure analysis from crystals that vary in thickness. The crystals of this segment appear as micrometer-long needles and vary in thickness from several nanometers to over a micrometer. Seven data sets originating from thin crystals were collected and combined

to yield a reference data set that was 80.3% complete in P1 with constants {**a**, **b**, **c**} (Å) = {4.86, 14.11, 18.41}, and angles { α , β , γ } ($^{\circ}$) = {90.00, 93.71, 101.21}. A second data set was constructed from 7 thick crystals between 500 nanometers and one micrometer thick, that were combined to yield a 75.6% complete data set with the same space group and unit cell dimensions as above. Thin and thick crystal structures were determined *ab initio* by direct methods using SHELXT (16) (See **SI Methods**) and refined to atomic resolution using BUSTER-TNT (17) showing clear atomic density (**Figure 1**). Refinement statistics for hPrP- $\beta 2\alpha 2$ are presented in **Table 1**.

The structure of hPrP- $\beta 2\alpha 2$ represents a prion protofilament with amyloid features. The protofilament is a class 2 steric zipper, composed of parallel, face-to-back beta sheets (18). One pair of sheets makes the protofilament, as observed with other amyloid structures (15, 18). At this resolution, the density shows hydrogen atoms as well as the presence of zinc and acetate ions that facilitate crystallographic contacts; both were present in the crystallization condition (**Figure 1**). Structures from both thick and thin crystals show clear densities for waters in the $2F_o - F_c$ map and have multiple hydrogens appearing in the $F_o - F_c$ density at the 3σ level (**Figure 1**). The appearance of resolvable hydrogens in electron diffraction was first reported for proteins by Rodriguez et al. (15) and later by Palatinus (19) for small molecules. Our *ab initio* solutions from thin and thick crystals are very similar, with a backbone RMSD of 0.07 Å and an all-atom RMSD of 0.09 Å (**Figure 1**). In summary, we found no significant differences between structures determined from thin and thick crystals of hPrP- $\beta 2\alpha 2$.

A comparison of structures determined from thin and thick crystals of a globular protein. We collected MicroED data from single ~500 nm and ~1 μm -thick crystals of proteinase K (**SI Figure 2**, **SI Figure 3**) and determined structures from each. Reflections were recorded to a resolution of 1.8 \AA in both cases yielding 97% completeness for the 500nm crystal and 79.7% completeness for the 1 μm thick crystal. Data were reduced in XDS (20) and phased by molecular replacement using the atomic coordinates of PDBID **5i9s** as a search model (21). Structures were refined using phenix.refine (22) with a high-resolution cutoff of 2.0 \AA and 3.0 \AA , respectively (**Table 1**; **Figure 2**). As the single, ~1 micrometer thick crystal structure had poor statistics, its data were merged with data gathered from three additional crystals of proteinase K; each of these also ~1 μm thick (**SI Figure 4**, **SI Figure 5**). The structure from the four combined data sets was determined again using the same search model now to a resolution of 2.5 \AA with better overall statistics (**Table 1**, **Figure 2**). A comparison of the resulting structures with the molecular replacement search model indicate good agreement with lower than 0.25 \AA all atom RMSD. In summary, we found no significant differences between structures determined from thin and thick crystals of proteinase K.

Systematic study of the effects of crystal thickness on MicroED data quality.

Diffraction was measured from 19 crystals of hPrP- $\beta 2\alpha 2$ with thicknesses ranging from ~100 nm to ~1100 nm over similarly-sized wedges of reciprocal space corresponding to a real-space angular range of approximately -30° to $+30^\circ$ (**SI Table 1**). Data for all comparisons was indexed and integrated using XDS (20). Estimation of crystal thickness is discussed at length in **SI Materials**. Briefly, average crystal thickness is

estimated by measuring the projected area from images recorded at 0° and 60° tilt, and the aspect ratio fit to an idealized model. These geometrical estimates were corroborated by intensity ratios using camera counts as previously described (14, 23) and are in good agreement (**SI Figure 6**). Crystal images are presented in **SI Document 2**. To assess the quality of diffraction data, we compare R values (R_{meas}), the ratio of intensity to variance (I/σ), and the half-set correlation coefficient ($CC_{1/2}$). All measured data values and statistics are presented in **SI Table 1**. Measured R values for these 19 crystals show a mean value of 13.03% with a standard deviation of 2.4%; average I/σ values average 4.1 with a standard deviation of 0.91. The half-set correlation coefficient is on average 98.2% with a standard deviation of 1.4%. The structure factor amplitudes for these 19 crystals were compared to the hPrP- $\beta 2\alpha 2$ structure solution from thin crystals discussed above. The correlation coefficient between the solved model and the individual crystals is shown in **Figure-2D** as CC_{model} (**Figure 3; SI Table 1**).

Absorption by protein crystals. The data presented above indicates that dynamical scattering does not inhibit structure solution by MicroED even from micrometer thick crystals when data is collected by continuous rotation. However, we note that the achievable resolution was lower from thick crystals compared to thin crystals (**Figure 1,2 and Table 1**). To evaluate whether absorption from thick crystals is limiting the achievable resolution, we recorded electron energy loss spectroscopy (EELS) spectra from 11 crystals of hPrP- $\beta 2\alpha 2$ and 10 crystals of proteinase K with thicknesses ranging from ~100 nm to ~1400 nm at 300kV (**Figure 4**). Control spectra were recorded from

regions of empty carbon support or within grid holes (**SI Document 3**). Each image was aligned by principal component analysis and a line scan through the zero-loss peak was measured along the first principal component. Intensities for the zero-loss peaks show expected exponential decay (9). A significant energy loss was observed for the carbon support alone with transmitted beam intensity decreasing by more than 40% (**SI Document 3**). An exponential fit to the data (**Figure 4**) suggests that the attenuation length, or (1/e) loss of intensity, due to a carbon film would be ~87 nm at 300kV. However, the (1/e) loss from crystals of proteinase K and hPrP- β 2 α 2 are 261 nm and 323nm, respectively, at 300kV. Our estimates of the actual carbon thickness are about 45nm using intensity ratios (**SI Figure 6**), in good agreement with previous findings on these grids (24). Thus even after accounting for the carbon film support, our estimates are very close to the mean free path of water at 300kV, found experimentally to be 336nm (25, 26). This data suggests that even a three-fold decrease in intensity due to absorption is insufficient to prohibit structure solutions by MicroED for micrometer thick macromolecular crystals at 300kV; the achievable resolution drops quickly near or beyond this thickness limit.

Discussion

These data indicate that accurate data can be collected and structures can be reliably determined by MicroED from crystals an order of magnitude thicker than previously expected. Simulations have suggested that structure determination by MicroED would be inhibited by dynamical scattering events from crystals only ~50-100nm thick (11, 27). Here we show that protein structure solutions can be obtained from substantially thicker

crystals than simulated limits suggest for both globular proteins and peptides using both molecular replacement and *ab initio* methods, respectively.

High quality structure solutions can be obtained for both hPrP- $\beta 2\alpha 2$ and proteinase K regardless of data originating from thin crystals or crystals nearly one micrometer thick. Clear atomic density in the $2F_o - F_c$ map of hPrP- $\beta 2\alpha 2$ reveals hydrogen atoms and the presence of zinc and acetate ligands as well as ordered water molecules. Multiple hydrogens are apparent in the $F_o - F_c$ density at the 3σ level (**Figure 1**). Likewise, the density for proteinase K is of high quality with well-defined density in both the σ -weighted difference maps and the SA composite omit maps for data obtained from thin (<200nm), intermediate (~500nm) and thick (>1000nm) crystals. Thickness does not prevent data collection, integration, or structure determination from these crystals.

The initial structure solution for proteinase K determined by MicroED (**5i9s**) was refined to a resolution of 1.75 Å from merged data of four thin crystals (20). Our present models from single crystals of ~500nm and ~1000nm thickness are determined to 2 Å and 3Å resolution, respectively. We credit this difference in resolution to increased absorption as demonstrated by our EELS analyses from thicker crystals. Together these results indicate that the achievable resolution can be limited by thick specimens. However, merging data from multiple thick crystals yielded in improved statistics and therefore improved resolution, indicating that absorption phenomena can be overcome to some extent by increased redundancy in measurements. Thicker protein crystals are ultimately limited in resolution by absorption.

Why the discrepancy between the experimental data presented here and the limits found in previous simulations? The inherent difficulty of performing simulations requires assumptions of experimental conditions that fail to capture the complexity of MicroED experiments. Specifically, simulations assume that crystals are perfect, all the electrons are scattered and data is recorded from stationary crystals from major zone axes (11, 28). These conditions are often encountered in crystals of small molecules such as inorganic compounds in material science and from two-dimensional (2D) protein crystals, but are rarely encountered in macromolecular crystallography using 3D protein crystals and MicroED protocols (29). Crystals of macromolecules are highly mosaic compared with those of simple organic compounds or inorganic crystals. Macromolecular crystals are imperfect and bent at the nanoscale. MicroED data from highly mosaic crystals benefits from their disorder in analogy to the pivoting of an electron beam in precession electron diffraction. Moreover, MicroED data is collected by continuous rotation so integration over the rocking curve further curbs dynamical effects. Our results confirm that dynamic scattering is not a major problem in solving protein crystal structures, and that dynamic scattering effects do not increase linearly with crystal thickness. However, the question of why significant artifacts are not observed from multiple scattering still remains, and requires further study.

The fact that multiple scattering artifacts do not limit structure determination in MicroED is especially surprising for crystals thicker than ~300nm, where the crystalline thickness easily exceeds the mean free path of an electron at either 200 or 300kV (**Figure 4**) (25,

26, 30). For these crystals, few electrons are transmitted through the crystal without any interaction with the specimen. At very high thicknesses the diffraction spots begin to widen due to energy loss. Larger spots make crystals with many atoms per unit cell difficult to measure and may give poor statistics as a large fraction of scattering will be reduced by absorption. Integrating large broadened spots from crystals with large unit cells may result in spot-overlap and other complications that may limit the achievable resolution; we have yet to encounter this problem. The largest protein determined to date by MicroED is catalase at ~240kDa and even with such a large unit cell and peak broadening, spot overlap was not observed and did not hinder structure determination (14). While broadening of diffraction spots can be curbed using an energy filter, this comes at the cost of reducing the transmitted signal. Nonetheless, use of an energy filter could improve signal to noise in MicroED experiments and the influence of thick crystals with large unit cells on diffraction quality remains to be further investigated.

In determining protein structures from crystals up to 1 micrometer in thickness using both *ab initio* phase retrieval and molecular replacement, we demonstrate that MicroED experiments are ultimately limited by absorption effects and crystal quality. At present, we estimate crystals ~500nm or thinner maximize data quality (~2x mean free path for 200kv), but structures from thicker crystals are not necessarily precluded. In fact, our results from even thicker crystals provides confidence that observed diffraction can be free of multiple scattering artifacts in a typical continuous rotation MicroED experiment, even from micrometer thick protein crystals.

Concluding Remarks

We demonstrate that MicroED yields accurate diffraction intensities, and that structures can be determined by molecular replacement or *ab initio* methods from protein crystals much thicker than previously suggested. Our data demonstrates that, as with X-ray diffraction, dynamical scattering is not a prohibitive source of error when MicroED data is collected on protein crystals by continuous rotation. Instead, absorption phenomena limit the achievable resolution in a way that can be curbed by merging multiple data sets. We suggest a stringent upper bound on crystal thickness of about one micrometer on MicroED experiments and a soft upper bound of about 500nm to ensure high-quality diffraction with the best possible resolution; these limits are not imposed by dynamical scattering, but instead by absorption. Our study expands the usefulness of MicroED as a general method for structure determination to atomic resolution from specimens up to a micrometer thick, opening new avenues of research that may broadly impact structural biology.

Materials and Methods

Protein preparation. hPrP- $\beta 2\alpha 2$ with greater than 98% purity was purchased from Genscript, dissolved in water at 10-20mg/ml and screened by the hanging drop method in a high throughput screen. Initial hits were optimized in 24-well hanging drop trays. The best crystals were observed in a condition containing 10% (w/v) PEG-8000; 0.1M MES pH 6.0; Zn(OAc)₂. This condition was used as the basis for a batch crystallization of the peptide at 10 mg/ml at a 1:1 ratio of peptide solution to mother liquor. In this condition, crystals grew as needle clusters that could be broken by force of pipetting and applied to grids for cryopreservation. Proteinase K (*E. Album*) was purchased from

Sigma and used without further purification. Crystals were grown by adding 5 μL of protein solution (50 mg/ml) to 5 μL of precipitant solution (1.5 M ammonium sulfate, 0.1 M Tris pH 8.0) in a sitting drop vapor diffusion tray. Large proteinase crystals were collected from sitting drops and sonicated into smaller crystals as previously described (4).

MicroED data collection. MicroED data was collected for 19 crystals of the peptide hPrP- $\beta 2\alpha 2$ (Gly-Ser-Asn-Gln-Asn-Asn-Phe) of varying thickness over the real space wedge corresponding to -30° to $+30^\circ$ under continuous rotation. All MicroED experiments were performed on an FEI Tecnai F20 microscope at an accelerating voltage of 200 kV, corresponding to a wavelength of 0.0251 \AA . Data was collected using a constant rotation rate of 0.2° s^{-1} on a TVIPS TemCam-F416 CMOS detector in rolling-shutter mode with 5 s exposures. Beam intensity for all hPrP- $\beta 2\alpha 2$ crystals was constant with an average dose rate of $0.003 \text{ e}^- \text{\AA}^{-1} \text{ sec}^{-1}$, or an overall exposure of $\sim 1 \text{ e}^- \text{\AA}^{-2}$. Proteinase K crystals were collected at a dose rate of $0.01 \text{ e}^- \text{\AA}^{-1} \text{ sec}^{-1}$ with an overall total dose per crystal of less than $3 \text{ e}^- \text{\AA}^{-2}$. hPrP- $\beta 2\alpha 2$ samples maintained a camera length of 0.730 m equivalent to a sample–detector distance of 1.313 m in a corresponding lens less system. Proteinase K data were collected at a camera length of 1.2m, corresponding to a detector distance of 2.200m. Diffraction data were collected through a circular selected area aperture of $1 \mu\text{m}^2$ in projection to reduce background noise. All TEM measurements were done at liquid nitrogen temperatures ($\sim 77 \text{ K}$).

Determination of Crystal Thickness. Crystal thickness was assessed in two ways: geometrical measurement and by using counts on the camera. First, Average crystal thickness was initially estimated geometrically by examining the projected area of each crystal from images taken at 0° , 15° , 30° , 45° , and 60° using the pixel length calibration from the known grid hole size of either 1 or 2 μm in ImageJ (NIH). Crystals of hPrP- $\beta 2\alpha 2$ were all rod-shaped and assumed to be ellipsoidal rods. Tilt series assessed the aspect ratio of the major (**a**) and minor (**b**) axis at each angle, with a known length of **2a** being the rod width measured at 0° . The average thickness of a crystal is then $\pi/4 = \sim 0.785\%$ of the maximum thickness, **2b**, perpendicular to the film. Proteinase K crystals were assumed to be cuboids with edge lengths **w**, **l**, **h**. Edge lengths of **w** and **l** were found from the 0° images and the **h** edge was estimated by the change in projected area of the crystal over a fixed length along the **x** and **y** axis. The average thickness of these crystals was assumed to be equal to the estimation of the **h** edge. Measurement accuracy was between 5 and 10 nm, the average variance for each crystal was $\sim 20\%$, and the standard deviation between intensity and geometry estimations was $\sim 100\text{nm}$.

Second, The geometrical approximation of crystal thickness above was validated by using intensity ratios as previously described (2, 14, 26). Each crystal from the hPrP- $\beta 2\alpha 2$ set in **Figure-3/SI Document 2** at 0° were identified at 0° . Each crystal had an area selected from the portion of the crystal collected upon where the median intensity transmitted was calculated. The same area was then used to calculate the median intensity for the nearest clean carbon area and nearest empty grid hole for comparison. The crystal thickness was estimated by first estimating the carbon thickness using the formula given in Feja & Aebi (26) assuming all missing intensity was lost due to energy

loss signals and the incident maximum intensity was that of the beam intensity in a grid hole. The mean free path of the amorphous carbon was calculated using the average energy loss of carbon being 14.1 eV derived from the formula given in (26), with $Z=6$, corresponding to a mean free path of 125nm at 200kV. This gave a consistent value of the carbon thickness being approximately 45nm with a standard deviation of 7nm. The crystal thickness is then calculated by subtracting the contribution of the carbon using the same equation with a mean free path of 242nm, calculated by scaling the value of 203nm mean free path of vitreous ice given by Grimm et al. at 120kV (25, 30). Our two measures were found to have a standard deviation of approximately 100nm. The correlation coefficient between the two measures was found to be 89%. This suggests our geometrical measurements are accurate within 25% of our listed values. The correlation plot and deviations from these intensity measures are given in **SI Figure 6**.

Absorption Experiments. Energy spectra were collected on a JEOL JEM-3200FSC microscope at 300 kV with a 500 nm² aperture in projection and total dose of 0.1 e⁻ Å⁻². Spectra were collected on a TVIPS TemCam-F416 CMOS detector at full resolution in normal (integration) mode and saved as 16-bit signed integer TIFF files. Spectra were scaled in ImageJ to the known 100 eV μm⁻¹ energy filter spacing and 15.6 μm pixel size. Individual images were loaded into Mathematica and aligned with their principal component along the horizontal axis – typically resulting in a 6.7° clockwise rotation. Line scans of width 1 pixel were selected through the zero-loss peak along the principal loss axis, and shifted such that their peak intensities were located at an energy loss of 0 eV. Thickness for crystals used in spectra were determined as described above. Zero-

loss peaks were fit to a general Gaussian model of $I=a*e^{-b(x-c)^2}$ to determine peak maxima and integrated peak intensity. Data points were weighted by the square root of their intensities for the fitting to assure proper solutions. Zero-loss peak intensity was fit to a general exponential decay model with the intercept fixed to the normalized value of the vacuum peak as in **Figure-4**. Exponential fits to the experimental data for the carbon film, proteinase K, and hPrP- $\beta 2\alpha 2$ protein crystals resulted in R^2 residuals of 1.0, 0.85, and 0.88, respectively.

MicroED Data Processing. Diffraction movies were converted to the SMV file format using TVIPS tools (31) and checked for pixel truncation as previously described (21). Indexing and integration were performed in XDS (20, 32). Integrated diffraction intensities were sorted and merged in XSCALE. Merged intensities were converted to amplitudes and the files formatted to SHELX format in XDSCONV. Thick and thin hPrP- $\beta 2\alpha 2$ were solved using SHELXT, placing all of the atoms in the unit cell correctly (33). SHELXT supported the hypothesis of P1 crystallographic symmetry (16). Intensities for proteinase K were input directly from XSCALE into Phaser for molecular replacement (34). Molecular replacement was successful with a LLG > 1000 and a TFZ > 20 using the model **5i9s**. Refinement was carried out using phenix.refine (22). Individual models were adjusted manually in Coot (35) by visual inspection of the atomic model against the F_o-F_c and $2F_o-F_c$ maps. All-atom composite omit maps with simulated annealing were calculated in Phenix (22).

Models and Figures. Figures were generated in either PyMol or VMD (36, 37). Plots and fits to data such as EELS spectra were created in Mathematica using nonlinear model fits to the data. Backbone RMSDs were calculated in VMD. All atom RMSD values were generated by the align command in PyMol.

Acknowledgements

We thank David Eisenberg (HHMI, UCLA), Robert Glaeser (UCB), and Michael Sawaya (UCLA) for helpful discussions and/or critical reading of this manuscript. Jose Rodriguez is supported as a Searle Scholar and a Beckman Young Investigator. The Gonen Laboratory is supported by the Howard Hughes Medical Institute. This work was also supported by the Janelia Research Visitor Program.

References

1. Henderson R (2009) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys* 28(2):171.
2. Shi D, Nannenga BL, Iadanza MG, Gonen T (2013) Three-dimensional electron crystallography of protein microcrystals. *Elife* 2:e01345.
3. Taylor KA, Glaeser RM (1974) Electron Diffraction of Frozen, Hydrated Protein Crystals. *Science (80-)* 186(4168):1036–1037.
4. de la Cruz MJ, et al. (2017) Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat Methods* (February):1–6.
5. Nannenga BL, Gonen T (2016) MicroED opens a new era for biological structure determination. *Curr Opin Struct Biol* 40:128–135.

6. Liu S, et al. (2016) Atomic resolution structure determination by the cryo-EM method MicroED. *Protein Sci* 26:8–15.
7. Nannenga BL, Shi D, Leslie AGW, Gonen T (2014) High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat Methods* 11(9):927–30.
8. Nannenga BL, Gonen T (2016) MicroED opens a new era for biological structure determination. *Curr Opin Struct Biol* 40:128–135.
9. Kirkland EJ (2010) *Advanced computing in electron microscopy* doi:10.1007/978-1-4419-6533-2.
10. Cowley JM, Moodie AF (1959) The scattering of electrons by atoms and crystals. II. The effects of finite source size. *Acta Crystallogr* 12(5):353–359.
11. Subramanian G, Basu S, Liu H, Zuo JM, Spence JCH (2015) Solving protein nanocrystals by cryo-EM diffraction: Multiple scattering artifacts. *Ultramicroscopy* 148:87–93.
12. Glaeser RM, Downing KH (1993) High-resolution electron crystallography of protein molecules. *Ultramicroscopy* 52(3–4):478–486.
13. White TA, Eggeman AS, Midgley PA (2010) Is precession electron diffraction kinematical? Part I: “Phase-scrambling” multislice simulations. *Ultramicroscopy* 110(7):763–770.
14. Nannenga BL, Shi D, Hattne J, Reyes FE, Gonen T (2014) Structure of catalase determined by MicroED. *Elife* 3:e03600.
15. Rodriguez JA, et al. (2015) Structure of the toxic core of α -synuclein from invisible crystals. *Nature advance on*(7570):486–490.

16. Sheldrick GM (2015) SHELXT - Integrated space-group and crystal-structure determination. *Acta Crystallogr Sect A Found Crystallogr* 71(1):3–8.
17. Blanc E, et al. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr Sect D Biol Crystallogr* 60(12 I):2210–2221.
18. Sawaya MR, et al. (2007) Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* 447(7143):453–457.
19. Palatinus L, et al. (2017) Hydrogen positions in single nanocrystals revealed by electron diffraction. *Science* (80-) 355(6321). doi:10.1126/science.aak9652.
20. Kabsch W (2010) Xds. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.
21. Hattne J, Shi D, De La Cruz MJ, Reyes FE, Gonen T (2016) Modeling truncated pixel values of faint reflections in MicroED images. *J Appl Crystallogr* 49:1029–1034.
22. Afonine P V., et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr Sect D Biol Crystallogr* 68(4):352–367.
23. Malis T, Cheng SC, Egerton RF (1988) EELS log-ratio technique for specimen-thickness measurement in the TEM. *J Electron Microsc Tech* 8(2):193–200.
24. Cho H-J, et al. (2013) Measurement of ice thickness on vitreous ice embedded cryo-EM grids: investigation of optimizing condition for visualizing macromolecules. *J Anal Sci Technol* 4(1):7.
25. Grimm R, Typke D, Bärmann M, Baumeister W (1996) Determination of the inelastic mean free path in ice by examination of tilted vesicles and automated

- most probable loss imaging. *Ultramicroscopy* 63(3–4):169–179.
26. Feja B, Aebi U (1999) Determination of the inelastic mean free path of electrons in vitrified ice layers for on-line thickness measurements by zero-loss imaging. *J Microsc* 193(1):15–19.
 27. Grigorieff N, Henderson R (1996) Comparison of calculated and observed dynamical diffraction from purple membrane: Implications. *Ultramicroscopy* 65(1–2):101–107.
 28. Jiang L, Georgieva D, Nederlof I, Liu Z, Abrahams JP (2011) Image Processing and Lattice Determination for Three-Dimensional Nanocrystals. *Microsc Microanal* 17(6):879–885.
 29. Shi D, et al. (2016) The collection of MicroED data for macromolecular crystallography. *Nat Protoc* 11(5):895–904.
 30. Russo CJ, Passmore LA (2016) Progress towards an optimal specimen support for electron cryomicroscopy. *Curr Opin Struct Biol* 37:81–89.
 31. Hattne J, et al. (2015) MicroED data collection and processing. *Acta Crystallogr Sect A Found Adv* 71:353–360.
 32. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):133–144.
 33. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr Sect A Found Crystallogr* 64(1):112–122.
 34. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658–674.
 35. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics.

Acta Crystallogr Sect D Biol Crystallogr 60(12 I):2126–2132.

36. DeLano W (2002) PyMOL. *DeLano Sci San Carlos, CA* (June 1972):1–15.
37. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38.

Table and Figure Legends

Table-1: Data Collection and Refinement Statistics.

Figure-1: Structural comparison of hPrP- β 2 α 2 from thin and thick crystals by direct methods (A) Structural model for hPrP- β 2 α 2 solved from typically <200 nm crystals (B) from a set of crystals with average thicknesses >1 μ m. $2mF_o-DF_c$ and mF_o-DF_c density maps are contoured at 1.5 σ and 3 σ levels, respectively. Positive hydrogen densities for both solutions are visible in A and B in green and are designed by orange arrows. (C) *Ab initio* solutions from both normal (yellow) and >1 μ m (blue) peptide crystals. (D) Final structural model of hPrP- β 2 α 2.

Figure-2: Structural comparison of globular proteins determined from thin, intermediate and thick crystals. Proteinase K sidechain density and corresponding structure solution from (A) thin ~200nm crystals (PDBID 5I9S)(20), (B) a single 600 nm-thick crystal, (C) a single 1000 nm-thick crystal, and (D) four crystals thicker than 1000 nm. All $2F_o-F_c$ density maps are contoured at the 1.5 σ level for residues 226-240 shown as a grey mesh. (E) All structure solutions of proteinase K overlaid for comparison. Overlaid colors correspond to individual colored structures.

Figure-3: Measurements from hPrP- β 2 α 2 crystals of variable thickness. (A) Mean $I/\sigma I$ values, (B) Measured, expected R values and their difference, R_{diff} , (C) internal half set correlation coefficient, and (D) the correlation of the measured structure factors to the solved thin crystal model, or CC_{solved} . Mean values are depicted by solid lines. The dotted black line corresponds to the maximum reported limit from simulations (10).

Figure-4: Observed transmission from crystals of various thicknesses. Peak intensities of the electron energy loss spectroscopy (EELS) spectra obtained from both hPrP- β 2 α 2 (blue) and proteinase K (orange) crystals. Black dots correspond to the recorded vacuum intensity and the carbon film. Exponential decay models are presented as dashed lines coordinated to the sample color. The thickness limit for MicroED experiments suggested by simulations is given as a dashed horizontal line (10).

SI-Table-1: MicroED diffraction data from hPrP- β 2 α 2 crystals of various thicknesses.

SI-Figure-1: (A) Thick crystals used to solve the thick hPrP- β 2 α 2 structure. (B) Structure solution of hPrP- β 2 α 2 solved from 7 crystals with average thickness $>1 \mu\text{m}$. Colors show individually measured angular ranges between -65 and -63 degrees with corresponding crystals having matching border colors. Grey regions correspond to the angular wedge not measured and (orange) showing the region available for measurement. Scale bars correspond to $2 \mu\text{m}$.

SI-Figure-2: (A) Single 500 nm crystal used to solve the proteinase K structure at marked angles and locations. (B) Corresponding structure solution model. Measurable data after inclusion of symmetry (orange) and recorded angular range (blue) from the crystal. Grey angular regions depict ranges not measured. Scale bars correspond to 2 μm .

SI-Figure-3: (A) Single 1 μm crystal used to solve the proteinase K structure at marked angles and locations. (B) Corresponding structure solution model. Measurable data after inclusion of symmetry (orange) and recorded angular range (blue) from the crystal. Grey angular regions depict ranges not measured. Scale bars correspond to 2 μm .

SI-Figure-4: (A) Four $>1 \mu\text{m}$ crystals used to solve the proteinase K structure at marked angles and locations. (B) Corresponding structure solution model. Measurable data after inclusion of symmetry (orange) and recorded angular range (blue) from the crystal. Grey angular regions depict ranges not measured. Scale bars correspond to 2 μm .

SI-Figure-5: $2mF_o-DF_c$ and all-atom SA composite omit potential maps with corresponding structural model for proteinase K from (A) four thin crystals, (B) a single 600 nm, (C) a single 1 μm , and (D) four merged crystals $>1 \mu\text{m}$ thick. $2F_o-F_c$ potential maps are contoured at the 1.5σ level with a 2 \AA curve.

SI-Figure-6: Crystal thickness estimation. (A) Comparison between the geometrically measured thickness (orange) and the thickness estimate from the Beer-Lambert Law

(Grey), and the calculated carbon thickness from the Beer-Lambert calculation (Blue).

(B) Correlation between the two types of measures of crystal thickness.

SI-Document-2: Peptide tilt angle images (diffraction data crystals) - PDF

SI-Document-3: EELS data (crystals, spectra, and fits) - PDF

	hPrP-β2α2	Thick hPrP-β2α2	Proteinase K	Proteinase K (Intermediate Thickness)	Proteinase K (>1μm, single crystal)	Proteinase K (>1μm, multiple crystals)
Wavelength (Å)	0.0251	0.0251	0.0251	0.0251	0.0251	0.0251
Resolution range (Å)	7.608 - 1.001 (1.037 - 1.001)	14.03 - 1.011 (1.047 - 1.011)	19.62 - 1.75 (1.813 - 1.75)	19.39 - 2.0 (2.072 - 2.0)	22.16 - 3.001 (3.108 - 3.001)	22.61 - 2.5 (2.589 - 2.5)
Space group (#)	P 1 (1)	P 1 (1)	P 43 21 2 (96)	P 43 21 2 (96)	P 43 21 2 (96)	P 43 21 2 (96)
Unit cell	4.86 14.11 18.41	14.34 4.89 17.8	67.32 67.32 101.01	67.21 67.21 100.59	68.09 68.09 102.13	67.5 67.5 102.41
a, b, c (Å)	90 93.71 101.21	94.96 91.02 101.82	90 90 90	90 90 90	90 90 90	90 90 90
α, β, γ (°)						
Total reflections	12907 (524)	8654 (307)	104808 (5429)	57404 (4205)	20317 (1505)	78690 (5656)
Unique reflections	2072 (171)	1869 (141)	22698 (2256)	15756 (1543)	4135 (294)	8627 (828)
Multiplicity	6.2 (4.2)	4.6 (3.2)	4.6 (4.4)	3.6 (3.6)	4.9 (5.1)	9.1 (9.2)
Completeness (%)	80.29 (69.08)	75.90 (60.00)	94.02 (94.14)	97.18 (97.60)	79.70 (81.82)	99.41 (99.64)
Mean I/σ(I)	5.05 (2.25)	4.64 (1.65)	3.1 (1.0)	3.0 (0.87)	2.20 (0.92)	4.83 (2.39)
Wilson B-factor	4.38	6.55	11.77	24.23	39.66	17.86
R-merge	0.238 (0.43)	0.233 (0.474)	0.458 (1.989)	0.294 (1.28)	0.551 (1.47)	0.414 (0.960)
R-meas	0.257 (0.485)	0.258 (0.552)	0.516 (2.258)	0.347 (1.51)	0.617 (1.63)	0.44 (0.97)
CC _{1/2}	0.98 (0.86)	0.979 (0.489)	0.929 (0.112)	0.972 (0.289)	0.883 (0.316)	0.939 (0.484)
Reflections used in refinement	2072 (172)	1869 (141)	22670 (2235)	15756 (1544)	4135 (414)	8627 (828)
Reflections used for R-free	207 (17)	180 (12)	1476 (146)	788 (77)	207 (20)	431 (41)
R-work	0.1984 (0.2067)	0.1925 (0.2102)	0.217 (0.348)	0.2218 (0.3112)	0.1987 (0.3059)	0.2085 (0.2858)
R-free	0.2382 (0.222)	0.2523 (0.2920)	0.266 (0.418)	0.2508 (0.3390)	0.2497 (0.4124)	0.2612 (0.4195)
Number of non-hydrogen atoms	66	63	2170	2130	2037	2090
macromolecules	55	55	2027	2027	2027	2027
ligands	5	5	10	10	10	10
solvent	6	3	133	93	0	53
Protein residues	7	7	281	281	281	281
RMS (bonds)	0.014	0.013	0.003	0.004	0.005	0.003
RMS (angles)	1.82	1.69	0.57	0.74	0.92	0.7
Ramachandran favored (%)	100	100	97.11	97.11	97.11	97.11
Ramachandran allowed (%)	0	0	2.53	2.53	2.53	2.53
Ramachandran outliers (%)	0	0	0.36	0.36	0.36	0.36
Rotamer outliers (%)	0	0	0	0	0.94	0
Clashscore	0	0	4.54	4.81	3.29	2.28
Average B-factor	5.85	6.78	11.49	23.19	25.12	12.61
macromolecules	4.76	6.05	11.19	23.28	25.15	12.75
ligands	7.48	9.69	29.54	27.28	19.83	16.75
solvent	14.52	15.18	14.63	20.75	0	6.39

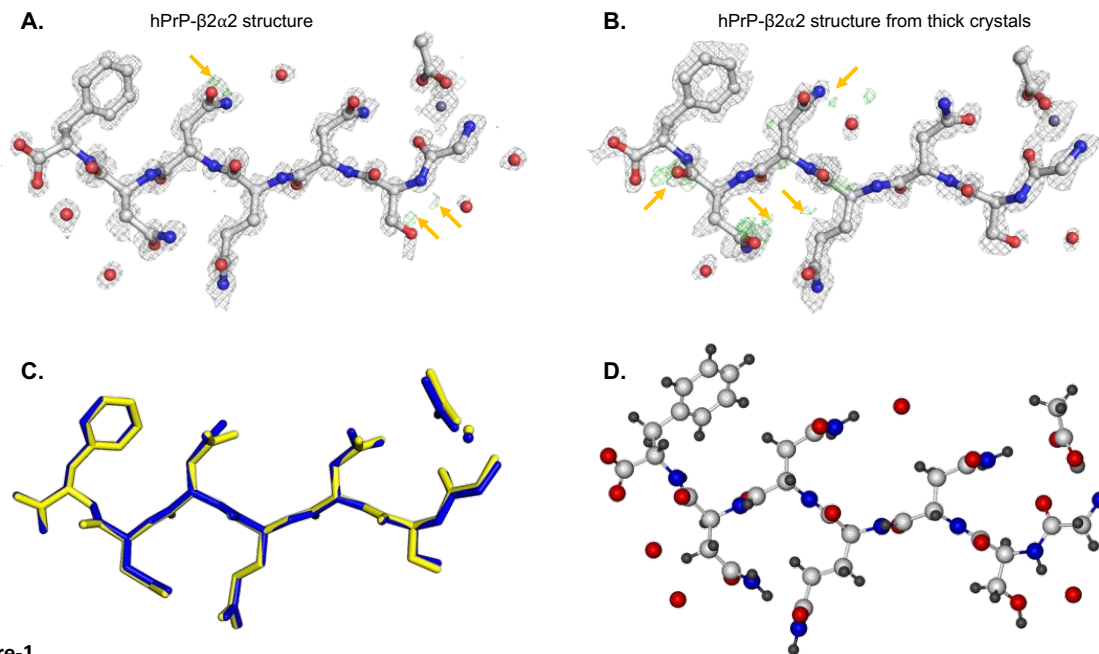


Figure-1

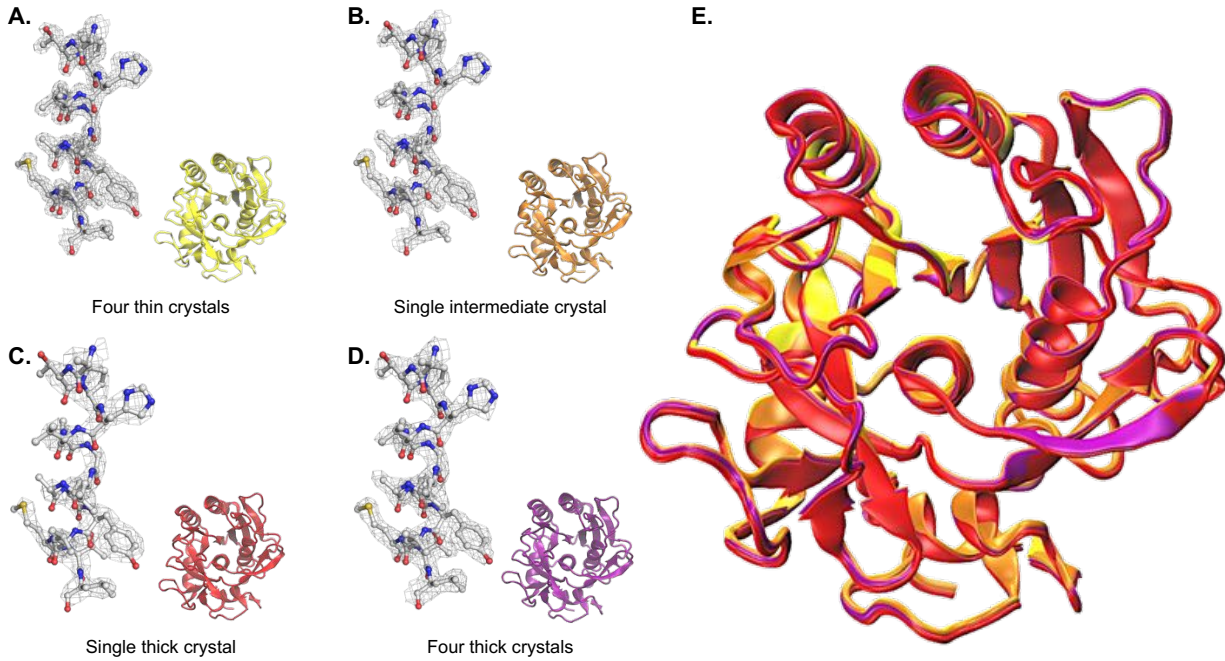
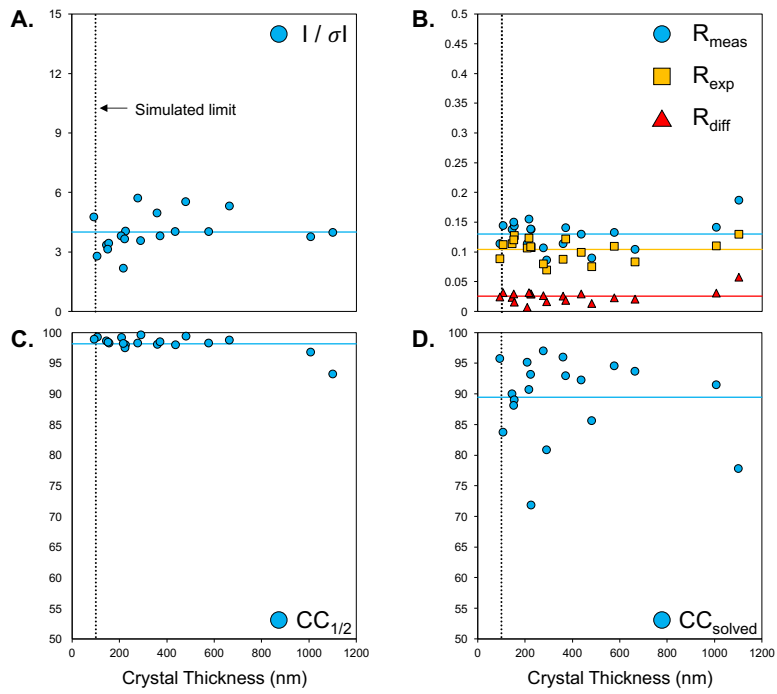


Figure-2



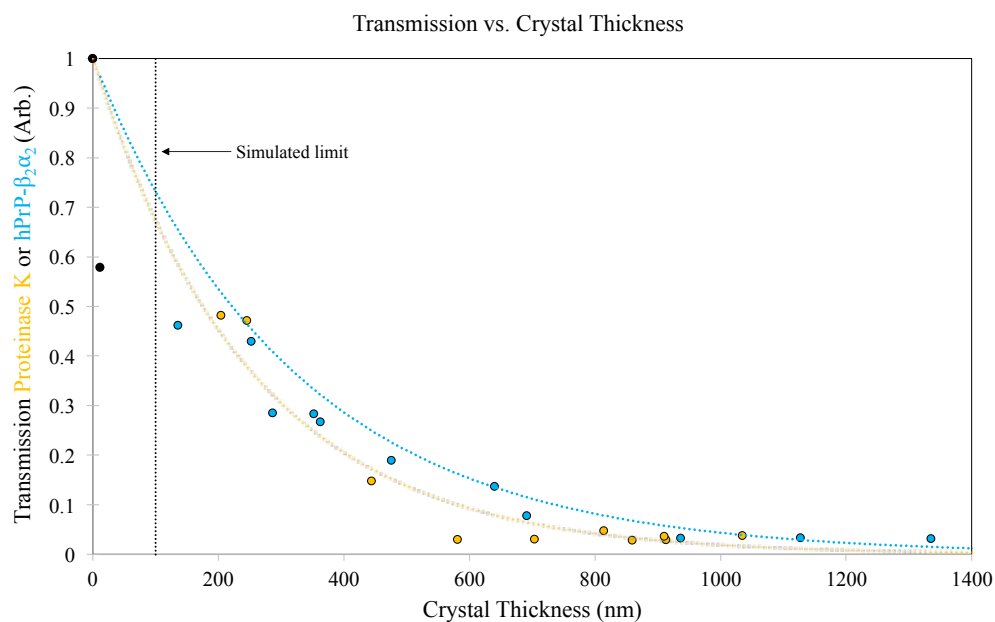


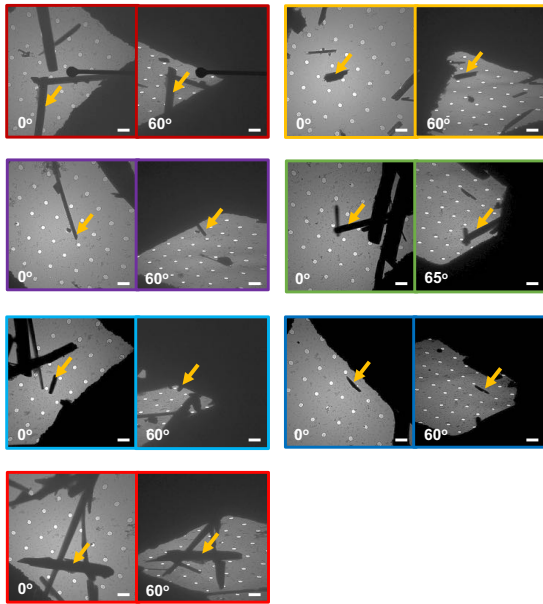
Figure-4

SI-Table-1: MicroED parameters for individual hPrP-β2α2 crystals

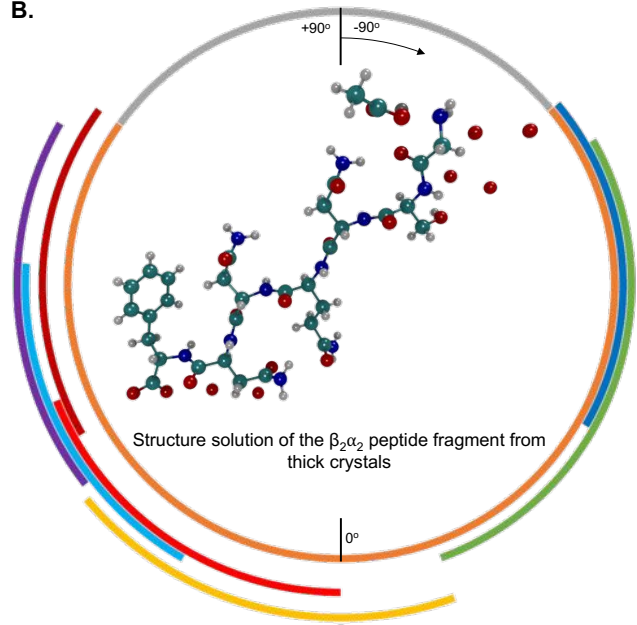
Mean Thickness (nm)	# Observed	# Unique	# Possible	Completeness	R _{observed}	R _{expected}	# Compared	I/σI	R _{measured}	CC _{1/2}	CC _{model}	R _{difference}	Crystal #
91	1420	771	2625	0.29	0.08	0.09	1298.00	4.78	0.11	98.90	95.80	0.03	18
106	921	518	2490	0.21	0.10	0.11	806.00	2.79	0.15	99.30	83.70	0.03	11
145	1311	720	2391	0.30	0.10	0.11	1182.00	3.37	0.14	98.60	90.00	0.02	7
150	1031	584	2622	0.22	0.11	0.12	894.00	3.14	0.15	98.40	88.10	0.03	20
155	1266	676	2544	0.27	0.10	0.13	1180.00	3.46	0.14	98.30	89.00	0.02	12
209	1390	747	2596	0.29	0.08	0.11	1286.00	3.82	0.11	99.20	95.20	0.01	5
217	1313	700	2582	0.27	0.11	0.12	1226.00	2.19	0.16	98.20	90.70	0.03	13
223	1396	767	2566	0.30	0.10	0.11	1258.00	3.67	0.14	97.50	93.20	0.03	8
225	1280	690	2587	0.27	0.10	0.11	1180.00	4.07	0.14	98.00	71.80	0.03	3
277	1415	775	2585	0.30	0.08	0.08	1280.00	5.74	0.11	98.30	97.10	0.03	21
291	942	572	2543	0.23	0.06	0.07	740.00	3.59	0.09	99.60	80.80	0.02	14
360	1471	790	2624	0.30	0.08	0.09	1362.00	4.97	0.11	98.10	96.00	0.03	9
372	1419	769	2526	0.30	0.10	0.12	1300.00	3.83	0.14	98.50	93.00	0.02	17
437	1351	725	2526	0.29	0.09	0.10	1252.00	4.04	0.13	98.00	92.30	0.03	10
481	1280	689	2665	0.26	0.06	0.08	1182.00	5.54	0.09	99.40	85.60	0.01	1
578	1505	816	2644	0.31	0.09	0.11	1378.00	4.05	0.13	98.30	94.60	0.02	16
664	1504	796	2554	0.31	0.07	0.08	1416.00	5.34	0.11	98.80	93.70	0.02	2
1012	1157	597	2593	0.23	0.10	0.11	1060.00	3.77	0.14	96.80	91.50	0.03	15
1107	686	378	2180	0.17	0.13	0.13	616.00	3.99	0.19	93.20	77.70	0.06	19

SI-Table-1

A.

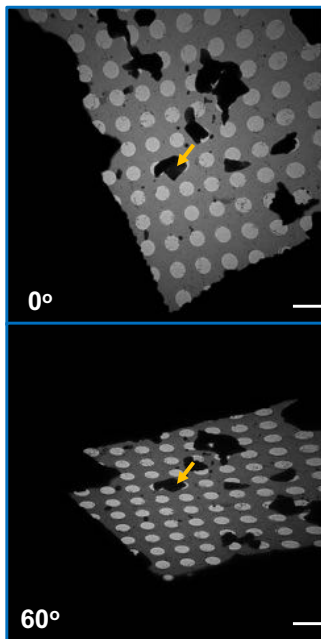


B.

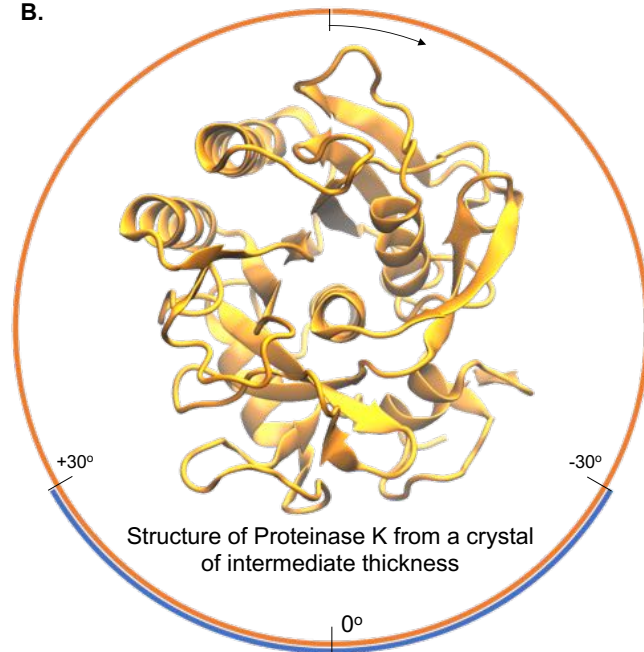


SI-Figure-1

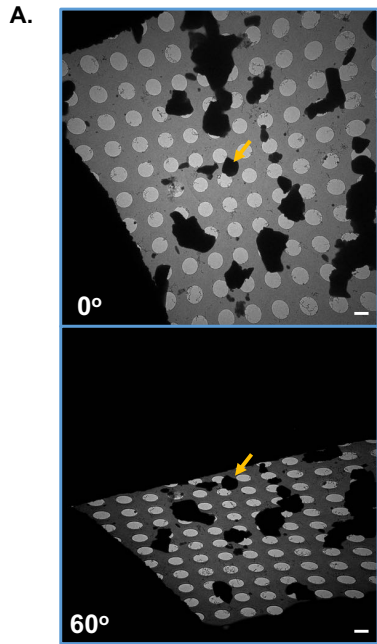
A.



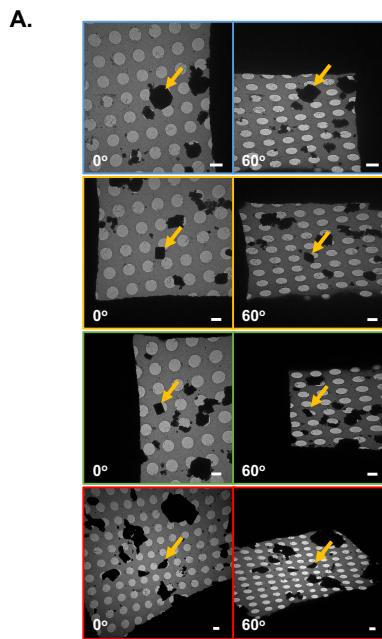
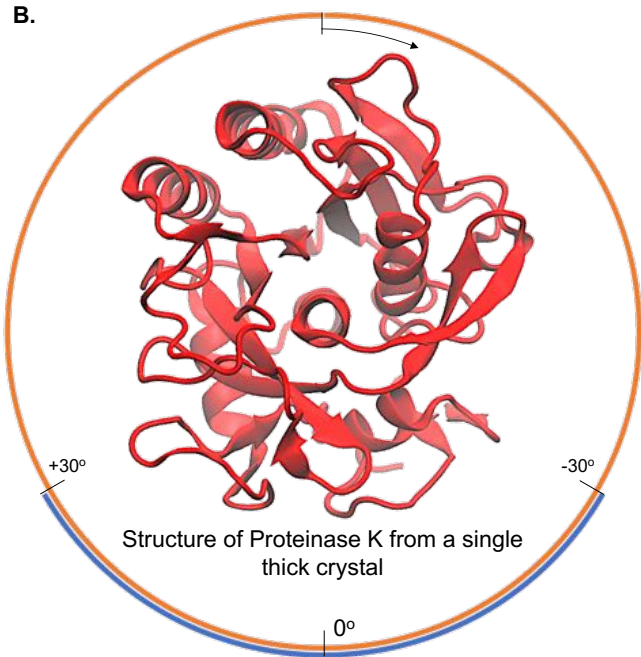
B.



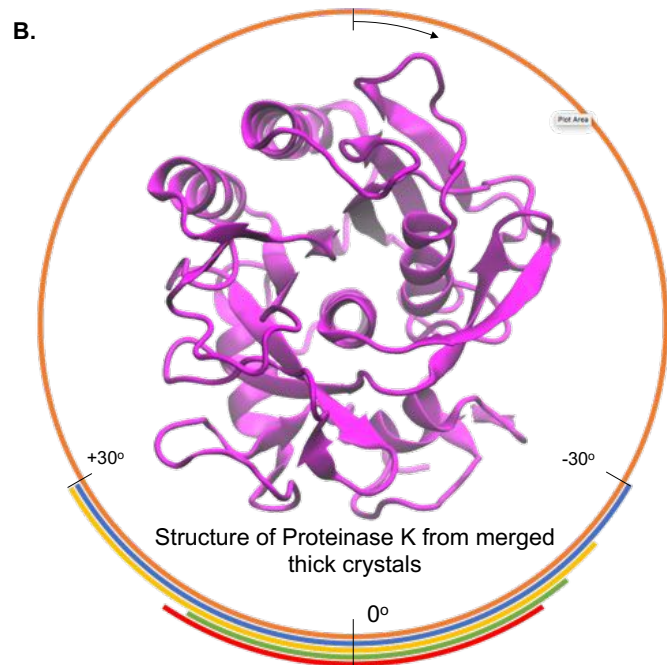
SI-Figure-2

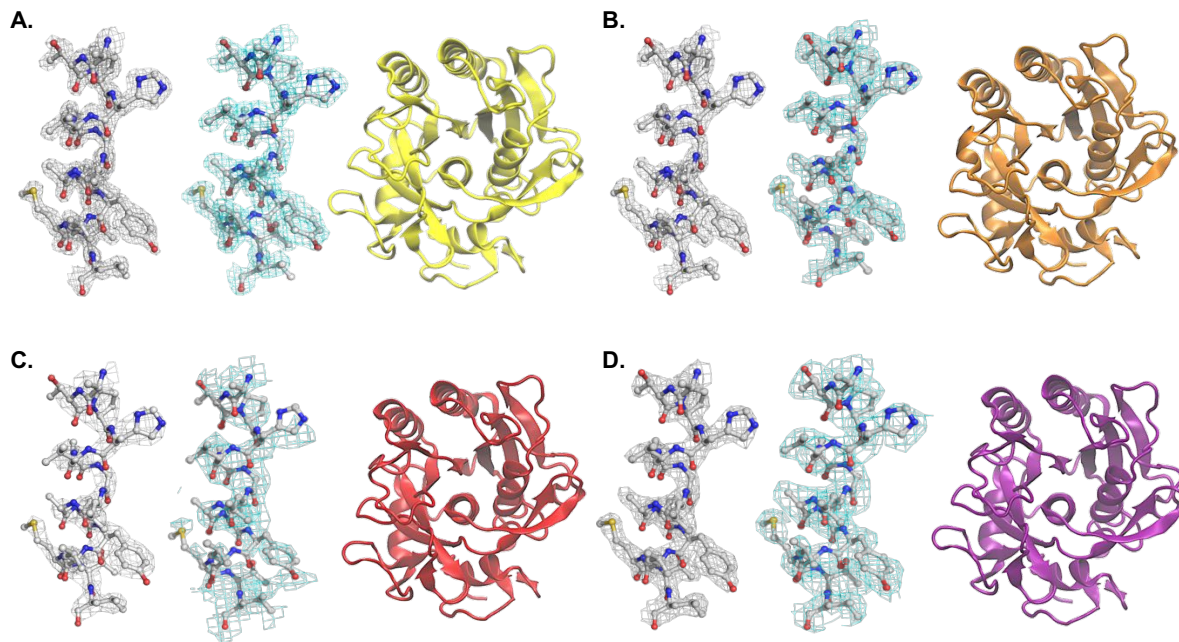


SI-Figure-3

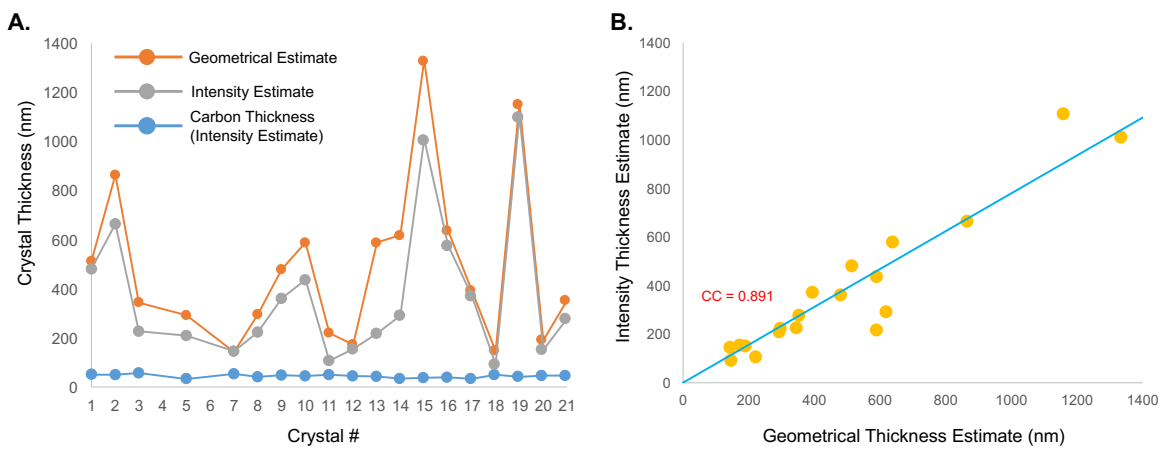


SI-Figure-4





SI-Figure-5

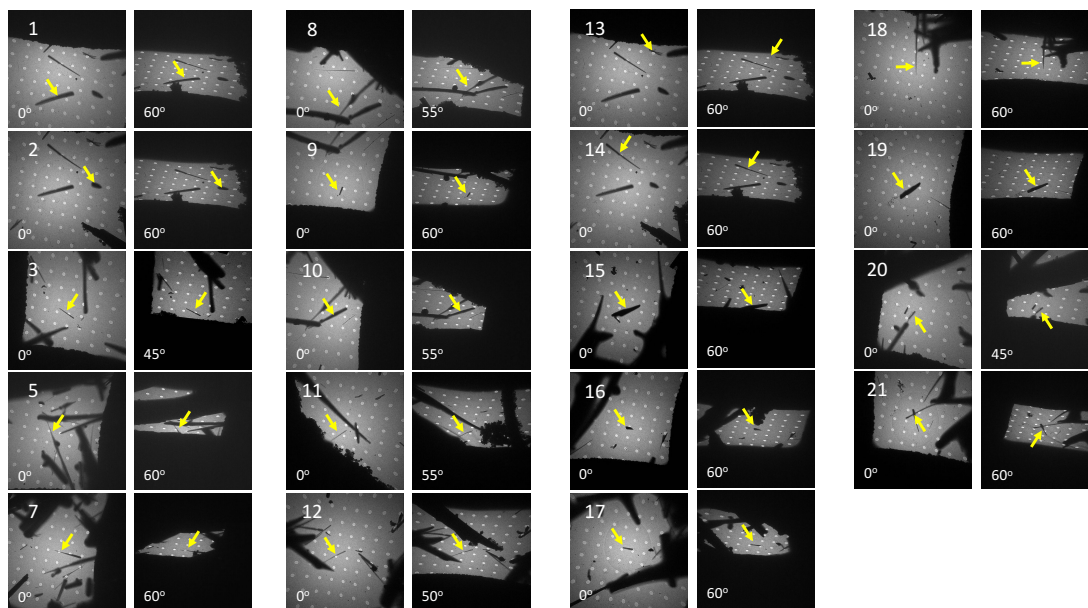


SI-Figure-6

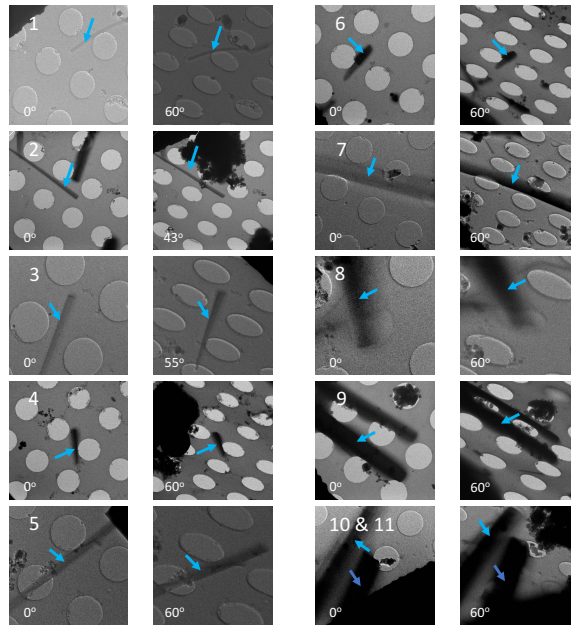
SI-Document-2: Peptide tilt angle images (diffraction data crystals) - PDF

SI-Document-3: EELS Data (images and fits) - PDF

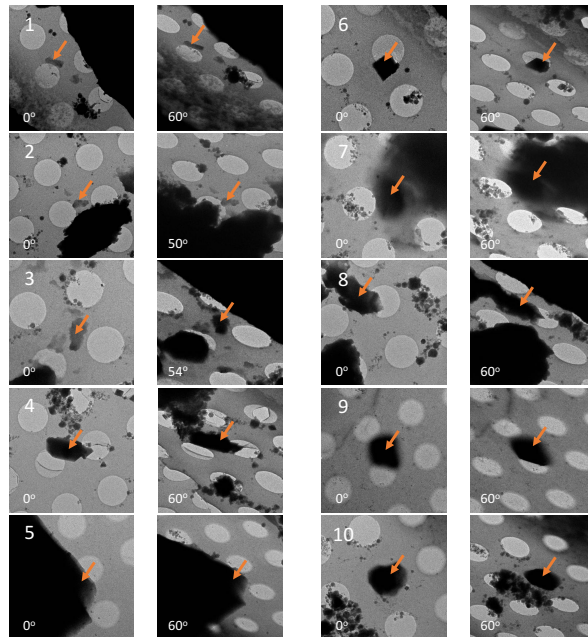
Crystals of hPrP - $\beta_2\alpha_2$ used to generate statistics in Figure-3 and SI-Table-1



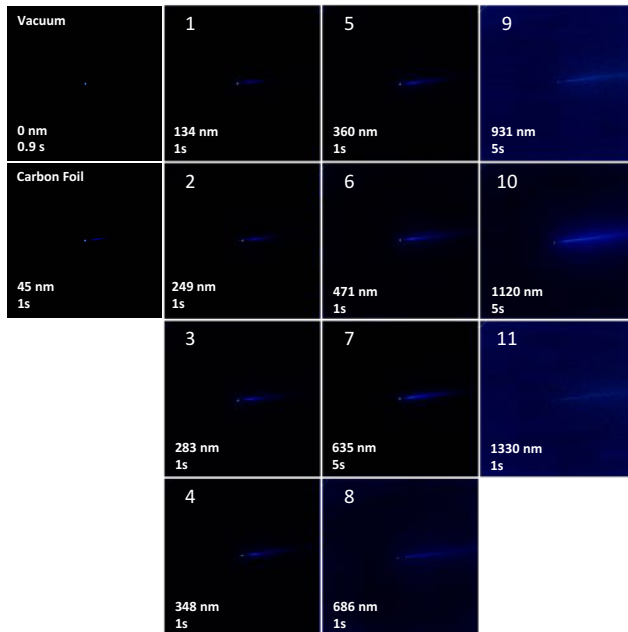
hPrP - B₂α₂ Crystals used in EELS Experiments



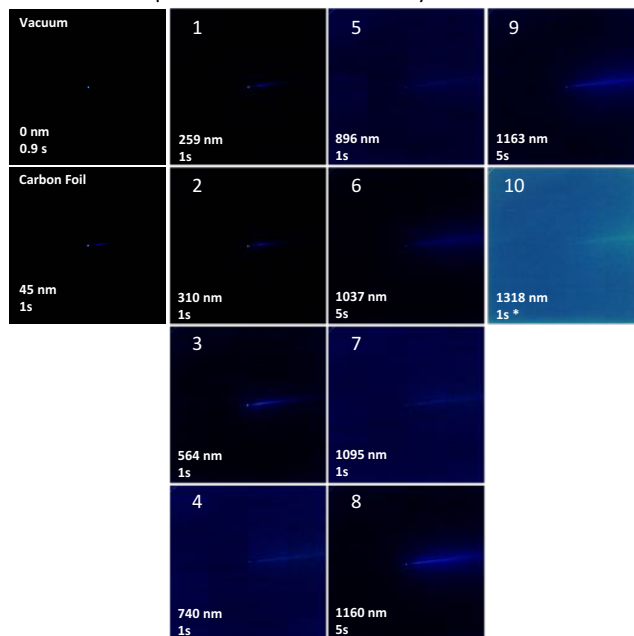
Proteinase K Crystals in EELS Experiments



EELS Spectra from $\beta_2\alpha_2$ crystals at 300 kV

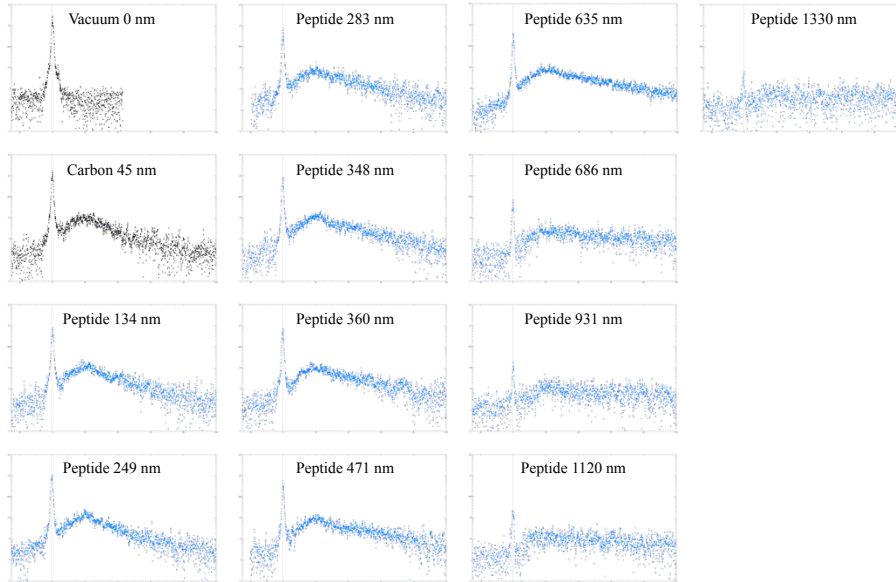


EELS Spectra from Proteinase K Crystals at 300kV

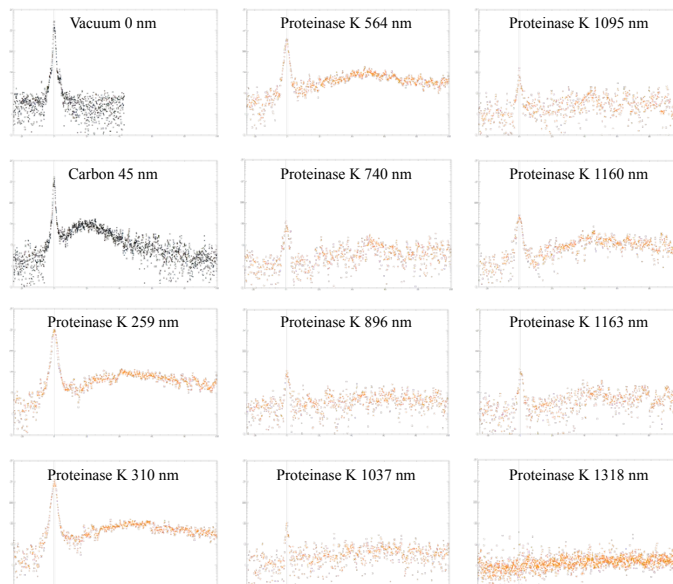


* Contrast enhanced 0.3%

Line Scans of EELS Spectra for hPrP - $\beta_2\alpha_2$ at 300 kV



Line Scans of EELS Spectra for Proteinase K Crystals



Appendix Chapter 2

Analysis of Global and Site-Specific Radiation Damage in Cryo-EM

The work described in this chapter has been reproduced from:

Johan Hattne, Dan Shi, Calina Glynn, Chih-Te Zee, Marcus Gallagher-Jones, Michael W. Martynowycz, Jose A. Rodriguez, Tamir Gonen. "Analysis of Global and Site-Specific Radiation Damage in Cryo-EM." *Structure*, 2018.

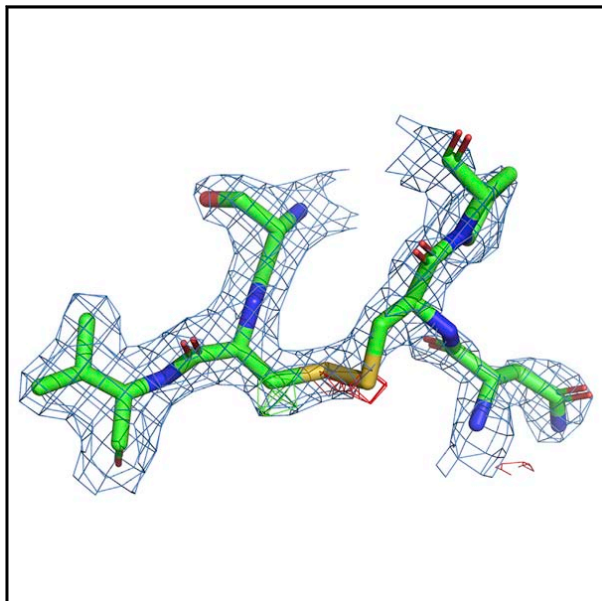
Copyright 2018

Johan Hattne, Dan Shi, Calina Glynn, Chih-Te Zee, Marcus Gallagher-Jones, Michael W. Martynowycz, Jose A. Rodriguez, Tamir Gonen.

Structure

Analysis of Global and Site-Specific Radiation Damage in Cryo-EM

Graphical Abstract



Authors

Johan Hattne, Dan Shi, Calina Glynn, ...,
Michael W. Martynowycz,
Jose A. Rodriguez, Tamir Gonen

Correspondence

tgonen@ucla.edu

In Brief

The scattered electrons in a cryo-EM measurement provide the information necessary to determine the atomic structure, but inevitably damage the sample. Radiation damage must be controlled to avoid compromising the result. Here, the relationship between exposure and damage is assessed in two different crystalline samples.

Highlights

- Damage to the crystalline order is quantified in reciprocal space
- Site-specific damage to sensitive moieties is assessed in real space
- The effects of damage are particularly pronounced at resolutions better than 3 Å
- At high resolution, damage is apparent at exposures as low as $1 \text{ e}^- \text{Å}^{-2}$



Hattne et al., 2018, Structure 26, 759–766
May 1, 2018 © 2018 Elsevier Ltd.
<https://doi.org/10.1016/j.str.2018.03.021>

CellPress

Analysis of Global and Site-Specific Radiation Damage in Cryo-EM

Johan Hattne,^{1,2} Dan Shi,² Calina Glynn,³ Chih-Te Zee,³ Marcus Gallagher-Jones,³ Michael W. Martynowycz,^{1,2} Jose A. Rodriguez,³ and Tamir Gonen^{1,2,4,5,*}

¹Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles CA 90095, USA

²Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA

³Department of Chemistry and Biochemistry, UCLA-DOE Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

⁴Departments of Physiology and Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles CA 90095, USA

⁵Lead Contact

*Correspondence: tgonen@ucla.edu

<https://doi.org/10.1016/j.str.2018.03.021>

SUMMARY

Micro-crystal electron diffraction (MicroED) combines the efficiency of electron scattering with diffraction to allow structure determination from nano-sized crystalline samples in cryoelectron microscopy (cryo-EM). It has been used to solve structures of a diverse set of biomolecules and materials, in some cases to sub-atomic resolution. However, little is known about the damaging effects of the electron beam on samples during such measurements. We assess global and site-specific damage from electron radiation on nanocrystals of proteinase K and of a prion hepta-peptide and find that the dynamics of electron-induced damage follow well-established trends observed in X-ray crystallography. Metal ions are perturbed, disulfide bonds are broken, and acidic side chains are decarboxylated while the diffracted intensities decay exponentially with increasing exposure. A better understanding of radiation damage in MicroED improves our assessment and processing of all types of cryo-EM data.

INTRODUCTION

Structure determination relies on interpreting the outcome of interactions of a beam of quanta with a sample. Most quanta pass through samples without any interaction at all (Henderson, 1995). Other quanta scatter either elastically, whereby they interact with the sample without losing any energy, or inelastically by depositing part of their energy to the sample. In a conventional diffraction measurement, the information that can be gained increases with the number of elastically scattered quanta. Inelastic scattering events manifest as damage, introduce noise, and ultimately limit the signal that can be extracted from the sample.

The success of crystallographic structure determination depends on the ratio of elastic to inelastic scattering events (Nave and Hill, 2005). While this ratio is greater in electron diffrac-

tion than in X-ray diffraction (Henderson, 1995), a single incident electron carries sufficient energy to knock out several electrons from an atom in an inelastic scattering event. These ejected, secondary electrons and their associated Auger electrons are mobile even at 77 K (Jones et al., 1987), and can, depending on the chemical composition of the sample and its surrounding mother liquor, cause additional ionization and excitation events in the crystal (Garman, 2010). Further damage from thermal diffusion of atomic and molecular radicals produced by these ionization events is curbed by keeping the sample at cryogenic temperatures, where diffusion is limited (Henderson and Unwin, 1975; Hayward and Glaeser, 1979; Uyeda et al., 1980; Jeng and Chiu, 1984). The absorbed energy, which is related to radiation damage, depends on the chemical composition of the crystal and the medium in which it is embedded, the temperature at which the measurement is performed, and the energy of the incident electrons.

As damage accumulates, its effects become apparent at specific sites as well as throughout the entire crystal. As crystalline order deteriorates, the fraction of unit cells contributing to crystalline diffraction decreases, the B factor generally increases (Kmetko et al., 2006), and the observed unit cell volume may increase as the lattice expands (Ravelli et al., 2002). Consequently, the signal from the obtained diffraction pattern, which varies with the square of the number of scattering unit cells, decreases. To some extent, damage can be compensated for by appropriate scaling procedures (Diederichs, 2006), and may be mitigated in nanocrystals as the probability that secondary electrons escape before causing further damage is higher than in large crystals (Nave and Hill, 2005; Sanishvili et al., 2011).

Site-specific damage may be observed if the impact of radiation damage on the crystal is not uniformly distributed but more selective toward certain moieties (Weik et al., 2000). Unlike global damage, site-specific damage can only be assessed once the dataset has been merged and phased, and a real-space density map is calculated. Site-specific damage becomes apparent when certain bonds are more susceptible to damage than others and may remain invisible if it only occurs in a small fraction of the unit cells or if it is masked by phases calculated from an undamaged model.

Exposure of the sample to the electron beam results in immediate damage even at cryogenic temperatures; low-dose



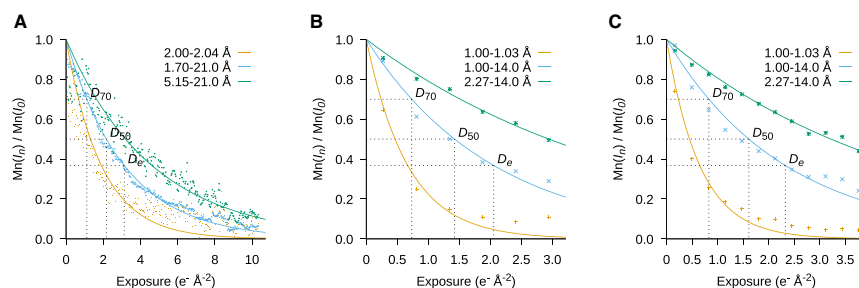


Figure 1. Exposure Dependency of the Mean Intensity of the Unmerged Integrated Reflections (A–C) Proteinase K (A), the hepta-peptide, GSNQNNF (B), recorded at an exposure rate of $0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$, and GSNQNNF (C) at $0.0017 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$. Spots at high resolution fade significantly faster than spots at low resolution. For proteinase K, the electron exposure at which the mean intensity across the entire recorded resolution range is reduced to 70%, 50%, and e^{-1} of its extrapolated value at zero dose is estimated to be $D_{70} = 1.1 \text{ e}^- \text{ \AA}^{-2}$, $D_{50} = 2.2 \text{ e}^- \text{ \AA}^{-2}$, and $D_e = 3.1 \text{ e}^- \text{ \AA}^{-2}$. The corresponding exposures for the $0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$ and $0.0017 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$ peptide datasets are $D_{70} = 0.73 \text{ e}^- \text{ \AA}^{-2}$, $D_{50} = 1.4 \text{ e}^- \text{ \AA}^{-2}$, and $D_e = 2.1 \text{ e}^- \text{ \AA}^{-2}$, and $D_{70} = 0.83 \text{ e}^- \text{ \AA}^{-2}$, $D_{50} = 1.6 \text{ e}^- \text{ \AA}^{-2}$, and $D_e = 2.3 \text{ e}^- \text{ \AA}^{-2}$, respectively.

procedures help minimize the exposure to the sample prior to data recording (Uyeda et al., 1980). Early studies indicated that organic samples deteriorate 4–5 \times faster at room temperature compared with cryogenic temperatures (Unwin and Henderson, 1975; Hayward and Glaeser, 1979; Jeng and Chiu, 1984). In these studies, two- and/or three-dimensional crystals were used in electron diffraction to look at the overall decay of reflections following exposure to the electron beam. These studies indicated that atomic resolution information (defined as better than 3 Å) was lost after exposure of the sample to only $3 \text{ e}^- \text{ \AA}^{-2}$. In 2013, a new method for cryoelectron microscopy (cryo-EM) was unveiled and termed micro-crystal electron diffraction (MicroED), or three-dimensional electron crystallography of microscopic crystals (Shi et al., 2013; Nannenga et al., 2014b). With continuous rotation MicroED, which is the preferred method of data collection in MicroED, significant loss of diffraction intensity was observed at resolutions better than $\sim 3 \text{ \AA}$ when only $\sim 3 \text{ e}^- \text{ \AA}^{-2}$ were used, consistent with past studies (Hayward and Glaeser, 1979; Jeng and Chiu, 1984; Baker et al., 2010).

Owing to the strong interaction of electrons with matter, and the fact that only diffraction data are collected (no imaging), high-resolution structures can be determined by MicroED from three-dimensional nanocrystals with significantly less total exposure than what is normally used for other cryo-EM modalities. Recent MicroED experiments demonstrated that complete datasets could be collected from a single nanocrystal using a total exposure of less than $\sim 1\text{--}2 \text{ e}^- \text{ \AA}^{-2}$ (Nannenga et al., 2014a; de la Cruz et al., 2017), making it possible to design new and improved experiments to test for beam-induced damage to the specimen with increasing exposure.

In this study, we set out to determine the damaging effects of electron radiation using MicroED and nanocrystals of a well-characterized sample, proteinase K, and a short hepta-peptide with a bound metal. MicroED data were collected using exposures of $0.0017\text{--}0.007 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$. Such low exposures allowed us to repeatedly measure the same wedge of reciprocal space from the same crystal and to compare the data obtained over

increasing exposure as the experiment progressed. The data were also sufficiently complete to allow us to investigate the effects not only in reciprocal space but also in real space, such that both global and site-specific damage can be observed. The data indicate that beam damage is a limiting factor in high-resolution cryo-EM methods, and the results have implications for all EM methods that use considerably higher exposures for imaging.

RESULTS

Indicators of Global Damage

For both the globular proteinase K and hepta-peptide samples, the overall weakening of the diffraction spots resulting from the loss of crystalline order can be modeled by exponential decay as a function of absorbed dose (Blake and Phillips, 1962; Liebschner et al., 2015) (Figure 1). We note that this model appears to systematically underestimate the intensities of the weakest reflections at high exposures for all samples. This observation does not necessarily invalidate the model: weak reflections are difficult to measure accurately due to noise; to the extent these reflections can be measured at all, they are not discernible by eye. Profiles derived from stronger reflections are likely to overestimate their intensities, and outlier rejection due to, for example, ill-fitting background models, introduces further bias toward more intense reflections.

After an exposure of $1 \text{ e}^- \text{ \AA}^{-2}$ the average intensity of all observed reflections in proteinase K decreased to 73% of its extrapolated value at zero dose (Figure 1). After an additional $1.6 \text{ e}^- \text{ \AA}^{-2}$, the high-resolution limit of the data dropped from 1.7 to 1.9 Å (Table 1). Similar trends can be seen in the peptide images at high and low exposure rates where the intensities have dropped to 61% and 65%, respectively, by $1 \text{ e}^- \text{ \AA}^{-2}$. However, the highest-resolution reflections in the peptide data are much stronger than those in proteinase K, and the effect of exposure on optical resolution is consequently not as pronounced. For both peptide datasets, the optical resolution decreases to $\sim 1.1 \text{ \AA}$ by $2 \text{ e}^- \text{ \AA}^{-2}$ (Tables 2 and Table 3).

Table 1. Processing and Refinement Statistics for Proteinase K

Proteinase K	Set 1 (6cl7, EMD-7490)	Set 2 (6cl8, EMD-7491)	Set 3 (6cl9, EMD-7492)	Set 4 (6cla, EMD-7493)	Set 5 (6clb, EMD-7494)
No. of crystals	6	6	6	5	4
$\langle t_{\text{exp}} \rangle$ (s)	123.1	370.0	617.3	862.6	1,107.9
Resolution (Å)	20.74–1.71 (1.74–1.71)	20.78–2.00 (2.05–2.00)	20.81–2.20 (2.27–2.20)	20.87–2.80 (2.95–2.80)	20.92–3.20 (3.46–3.20)
Completeness (%)	93.4 (71.9)	96.9 (97.0)	94.3 (94.6)	88.4 (87.7)	76.8 (77.7)
Multiplicity	6.1 (4.9)	6.2 (6.3)	5.7 (5.8)	5.6 (5.8)	3.3 (3.4)
CC _{1/2}	0.950 (0.164)	0.948 (0.138)	0.925 (0.097)	0.908 (0.142)	0.828 (0.215)
$R_{\text{work}}/R_{\text{free}}$	22.13/25.34	21.86/25.88	23.85/30.18	26.30/32.56	21.40/32.62

The average exposure time of a multi-crystal dataset is denoted $\langle t_{\text{exp}} \rangle$; it is defined as the mean cumulative irradiated time of all the frames in the dataset. Numbers in parentheses refer to the highest-resolution shell for refinement. Owing to varying response to radiation, damage-induced non-isomorphism causes the number of datasets that can be merged to decrease at the higher exposures.

The highest-resolution reflections, those providing atomic resolution, are most sensitive to disruption of crystalline order. Consequently, as a result of subtle long-range changes in the lattice, weak, high-resolution reflections fade into the background faster than strong, low-resolution reflections (Blake and Phillips, 1962; Howells et al., 2009), and the highest observable resolution decreases with dose (Tables 1, 2, and 3). In conjunction, fine features in real space density maps disappear at a higher rate than the overall molecular envelope. Generally, high-resolution information, defined as better than 2 Å, was significantly decayed when exposures greater than $\sim 3 \text{ e}^- \text{ \AA}^{-2}$ were used. Beyond $\sim 4 \text{ e}^- \text{ \AA}^{-2}$, the reflections at a resolution finer than 2 Å have dropped to 10% of their extrapolated value at zero dose in diffraction patterns of proteinase K. The corresponding exposure for the peptide at high and low dose rates are similar at 2.7 and $3.2 \text{ e}^- \text{ \AA}^{-2}$, respectively.

As an additional proxy for global damage by electrons we use changes in the relative B factor, B_{rel} (Kmetko et al., 2006), calculated over the reflections in the resolution range common to all datasets of a given sample and exposure rate. For all but the poorest diffracting crystals, B_{rel} increases monotonically with absorbed dose (Figure 2). We also find that the unit cell volume, although a much less reliable indicator of global radiation damage, generally increases with exposure (Murray and Garman, 2002). In our measurements, the unit cell volume of proteinase K increased by a modest 1.8%

from 452 to 460 nm³; the trends for the peptide are much less clear.

Site-Specific Damage

Localized chemical changes within the macromolecule can be analyzed by observing changes in the density attributed to specific atoms in real space. We observe site-specific damage even at exposures as small as $0.1 \text{ e}^- \text{ \AA}^{-2}$. For example, the presence of positive $mF_o - DF_c$ difference density is detected around the sulfur atoms of the disulfide bonds in proteinase K, indicating that the disulfide bridge was breaking in a significant fraction of the unit cells (Helliwell, 1988), even at total exposure $< 0.9 \text{ e}^- \text{ \AA}^{-2}$ (Figure 3). At this exposure, the overall diffraction intensity was reduced to 75% of its extrapolated value at zero dose, indicating that 86% of the unit cells are still diffracting to high resolution (Blake and Phillips, 1962). As the exposure increases, the positive difference density is replaced by negative difference density in the location of the bond and the $2mF_o - DF_c$ density progressively weakens.

We also observe site-specific damage in the decarboxylation of the acidic side chains (Figure 3). Like the disulfide bonds, these moieties have been observed to be particularly sensitive to radiation in both X-ray crystallography (Weik et al., 2000) and single-particle cryo-EM (Bartesaghi et al., 2014; Barad et al., 2015). In proteinase K, the density around the side chains of glutamate and aspartate begins to deteriorate, starting at a

Table 2. Processing and Refinement Statistics for the Hepta-Peptide GSNQNNF at $0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$

GSNQNNF ($0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$)	Set 1 (6clc, EMD-7495)	Set 2 (6clcd, EMD-7496)	Set 3 (6clce, EMD-7497)	Set 4 (6clcf, EMD-7498)	Set 5 (6clcg, EMD-7499)	Set 6 (6clch, EMD-7500)
No. of crystals	8	10	9	10	8	6
$\langle t_{\text{exp}} \rangle$ (s)	96.7	289.1	480.1	670.1	860.5	1,050.5
Resolution (Å)	13.96–1.01 (1.03–1.01)	13.95–1.01 (1.03–1.01)	14.04–1.01 (1.03–1.01)	13.98–1.15 (1.20–1.15)	13.75–1.35 (1.44–1.35)	13.96–1.37 (1.46–1.37)
Completeness (%)	76.6 (67.0)	82.4 (57.2)	77.2 (37.6)	81.2 (82.9)	77.2 (78.1)	73.5 (74.1)
Multiplicity	5.9 (5.4)	6.4 (4.5)	5.7 (3.1)	6.5 (6.0)	5.8 (5.8)	4.7 (4.5)
CC _{1/2}	0.980 (0.892)	0.983 (0.842)	0.988 (0.758)	0.986 (0.850)	0.986 (0.590)	0.968 (0.385)
$R_{\text{work}}/R_{\text{free}}$	17.77/15.86	22.03/21.31	24.52/24.38	26.59/29.28	21.95/26.86	22.50/29.75

Rows as per Table 1.

Table 3. Processing and Refinement Statistics for the Hepta-Peptide GSNQNNF at 0.0017 e⁻ Å⁻² s⁻¹

	Set 1 (6crl, EMD-7501)		Set 2 (6crl, EMD-7502)		Set 3 (6crl, EMD-7503)		Set 4 (6crl, EMD-7504)		Set 5 (6crl, EMD-7505)		Set 6 (6crl, EMD-7506)		Set 7 (6crl, EMD-7507)		Set 8 (6crl, EMD-7508)		Set 9 (6crl, EMD-7509)		Set 10 (6crl, EMD-7510)		Set 11 (6crl, EMD-7511)		Set 12 (6crl, EMD-7512)			
	11	11	9	10	10	11	11	10	10	11	9	11	11	11	10	10	12	10	10	10	10	9	9	9		
No. of crystals	11	11	9	10	10	11	11	10	10	11	9	11	11	11	10	10	12	10	10	10	10	9	9	9		
$\langle t_{exp} \rangle$ (s)	97.2	291.3	484.6	677.1	677.1	869.0	1,060.6	1,251.0	1,442.9	1,442.9	1,636.2	1,832.2	2,024.3	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	2,214.0	
Resolution (Å)	14.00–1.01	13.96–1.01	13.92–1.01	11.39–1.02	13.93–1.01	13.87–1.15	13.76–1.15	13.79–1.16	13.79–1.21	13.77–1.31	13.79–1.21	13.77–1.31	13.79–1.46	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45	13.73–1.45
Completeness (%)	78.0 (61.6)	79.6 (74.0)	80.0 (74.9)	72.5 (64.8)	78.1 (66.3)	79.7 (62.6)	78.8 (79.6)	76.0 (78.9)	79.3 (77.9)	79.4 (80.3)	78.8 (80.3)	79.3 (77.9)	79.4 (80.3)	78.7 (80.0)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)	77.6 (77.7)
Multiplicity	7.6 (6.1)	7.8 (6.0)	6.4 (4.3)	7.4 (5.2)	7.4 (4.5)	5.8 (6.2)	7.1 (6.9)	6.9 (6.7)	7.9 (7.2)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)	6.7 (5.8)
CC _{1/2}	0.972	0.973	0.948	0.895	0.978	0.940	0.917	0.988	0.974	0.980	0.974	0.980	0.980	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984	0.984
R _{work} /R _{free}	19.12/19.62	20.51/19.60	26.05/26.74	24.64/27.98	22.90/25.54	21.66/22.12	26.55/26.79	23.92/27.72	23.60/26.17	24.07/26.17	23.92/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17	24.07/26.17

Rows as per Table 1.

total exposure of $\sim 2 \text{ e}^- \text{ \AA}^{-2}$, and completely disappears after approximately $5 \text{ e}^- \text{ \AA}^{-2}$ (Figure 3).

The peptide unit cell contains a zinc atom, which displays significant signs of site-specific radiation damage at exposures $> 0.8 \text{ e}^- \text{ \AA}^{-2}$ (Figure 4). For the bound zinc, the radiation damage is primarily modeled using the atomic displacement parameter (ADP). Unlike occupancies, which model large-scale discrete disorder and were fixed at unity in all models, the ADPs describe harmonic vibrations around the mean position of the atoms. While the density around the zinc atom in the model remains positive even at the highest exposure, its ADP generally increases with exposure (Figure 4). The displacement begins with as little total exposure as $0.2 \text{ e}^- \text{ \AA}^{-2}$, when the ADP of the zinc is $1.3\times$ higher than the average ADP in the peptide model.

Site specific damage was further assessed in real space using RIDL (Bury et al., 2015), which calculates the maximum density loss D_{loss} for each atom in the model. This provides a means to objectively establish the sensitivity of different amino acids to electron radiation. Generally, the results from MicroED are consistent with the order and appearance of site-specific damage observed in X-ray crystallography: metals are significantly more prone to damage, and glutamate, aspartate, and cysteine residues accumulate damage even at very low electron beam exposures (Figure 5).

DISCUSSION

Past studies of radiation damage in cryo-EM described the exposure-dependent decay of diffraction intensities up to and including 3 \AA resolution (Henderson and Unwin, 1975; Hayward and Glaeser, 1979; Jeng and Chiu, 1984; Baker et al., 2010). No data were analyzed at resolutions better than 3 \AA , likely because such data were not recorded at the time. Unfortunately, even a recent study that reported a 2.6 \AA resolution single-particle reconstruction included an analysis of beam-induced damage only to lower resolution (Grant and Grigorieff, 2015). Moreover, while the early experiments analyzed the decay of diffraction intensities, little or no analysis of the effects of exposure on real space was described. This is likely because the effects of specific damage in real space are difficult to characterize to any degree of accuracy at low resolution.

We performed a deep analysis of the effects of electron radiation damage on biological samples at resolutions better than 1 \AA using MicroED (Shi et al., 2013; Nannenga et al., 2014b). Using ultra-low exposures allowed us to collect sufficient data from several crystals for structure determination. Each crystal was then sequentially exposed to the electron beam and additional structures were determined, from the very same crystals, at increasing levels of total exposure. In this way, we could investigate the effects of exposure on both reciprocal and real space informing us on both global and site-specific damage to the sample. With this approach, we could follow the trends of beam-induced damage in biological matter at very high resolutions (better than 1 \AA) in a way not previously possible in cryo-EM.

Our real-space analysis shows that site-specific damage is apparent at high resolution, even with exposures less than $1 \text{ e}^- \text{ \AA}^{-2}$; these exposures are well below those currently used in other cryo-EM modalities, for example imaging in single-particle EM. This analysis therefore holds important

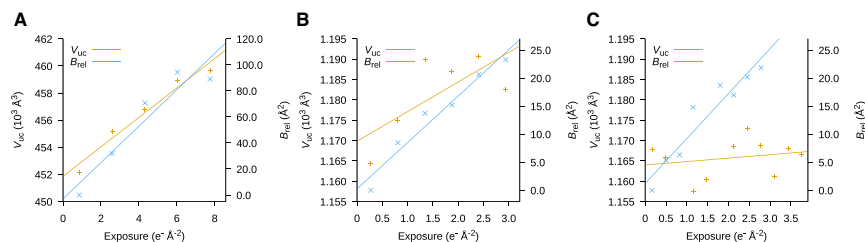


Figure 2. Exposure Dependency of the Unit Cell Volume, V_{uc} , and Relative B Factor, B_{rel}
 V_{uc} and B_{rel} were averaged across all the crystals at each exposure. For B_{rel} only reflections in a sufficiently large resolution range common to all datasets were considered (20.8–3.20 Å for proteinase K; 14.0–1.2 Å for GSNQNNF). (A) proteinase K, (B) the hepta-peptide, GSNQNNF, recorded at an exposure rate of $0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$, and (C) GSNQNNF at $0.0017 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$.

implications for all cryo-EM methods, particularly single-particle EM, as resolutions that approach those commonly observed in crystallography have recently been reported (Merk et al., 2016).

Imaging in single-particle cryo-EM allows the determination of protein structures from a collection of thousands of projection images of individual particles oriented randomly in vitrified ice. When a sufficiently fast camera is available, single-particle cryo-EM data are often collected as movie. The exposure delivered to the sample reflects a trade-off between contrast and loss of high-resolution information to radiation damage. Using catalase crystals, it was previously suggested that the optimal trade-off between signal and damage was $\sim 20 \text{ e}^- \text{ \AA}^{-2}$ for the target resolution of 20 Å, while for 3 Å resolution it was recommended that $10 \text{ e}^- \text{ \AA}^{-2}$ be used (Baker et al., 2010). In single-par-

ticle EM, the total exposure that is typically more than $20 \text{ e}^- \text{ \AA}^{-2}$, is fractionated over a sequence of short exposures. This allows individual frames to be corrected for specimen drift and beam-induced movement prior to averaging, while the first frame is usually discarded during processing. The last frames aid alignment but are excluded from the final average because they contribute little high-resolution information. The average of the first ~ 2 –5 frames from the movie therefore reflect a superposition of the same particles exposed to electron doses in the range of 2–10 $\text{e}^- \text{ \AA}^{-2}$, and these are used for the final reconstruction when combined with data from thousands of other particles.

Since the damage mechanisms in all these EM methods originate from the same phenomena, it is likely that the effects of

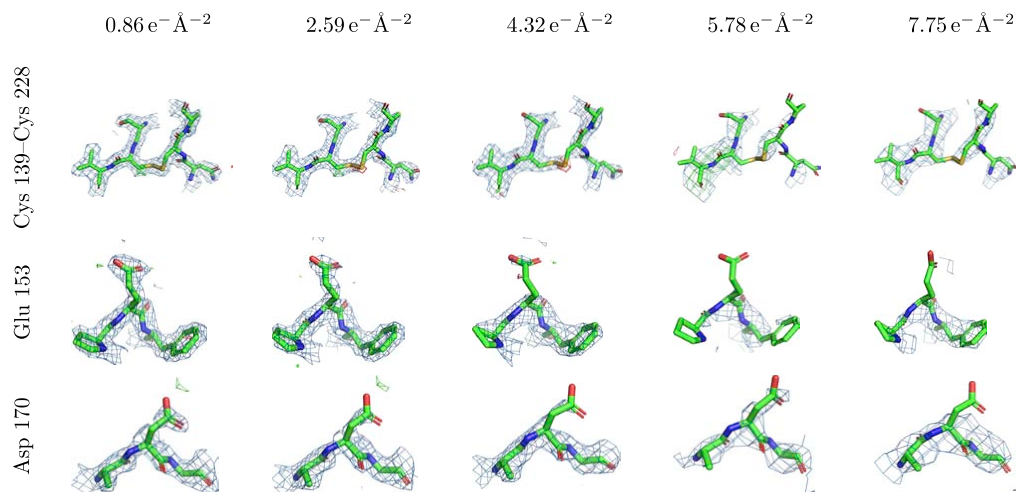


Figure 3. Disulfide Bond Breakage and Decarboxylation of Acidic Side Chains Indicate Site-Specific Radiation Damage in Proteinase K
 $2mF_o - DF_c$ maps (blue meshes) are contoured at 1.5σ above the mean, $mF_o - DF_c$ difference densities (green/red meshes) are contoured at $\pm 3\sigma$ above/below the mean. Maps up to and including those calculated at $4.3 \text{ e}^- \text{ \AA}^{-2}$ use data extending to 2.2 Å; the two maps at the highest exposure only use reflections up to 3.2 Å. Densities are carved to 2 Å around the selected atoms. All figures were generated using PyMol (Schrödinger, 2014).

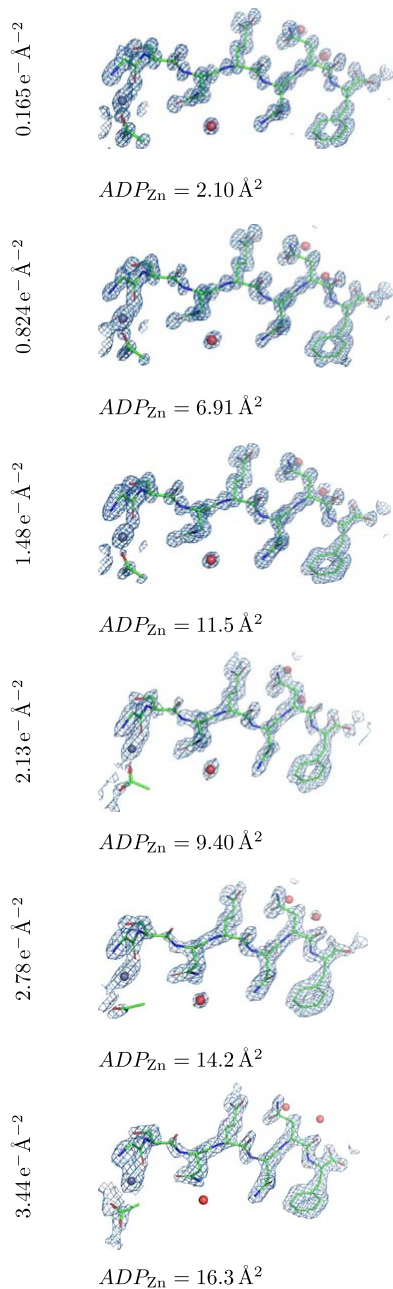


Figure 4. Exposure Dependency on the Hepta-Peptide Density

$2mF_o - DF_c$ (blue meshes, contoured at 1.5σ above the mean) and $mF_o - DF_c$ density (red/green meshes, contoured at $\pm 3\sigma$ above/below the mean) from the GSNQNNF hepta-peptide at $0.0017 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$. All densities are carved to 2 Å around the model and the maps at electron exposures up to and including $1.5 \text{ e}^- \text{ \AA}^{-2}$ are calculated using all observed reflections to 1.01 Å; remaining maps use reflections up to 1.45 Å. The atomic displacement parameter of the Zn atom, ADP_{Zn} (purple sphere to the left), generally increases over the course of exposure.

randomly distributed damage events are washed out during the immense averaging and the use of methods to exclude certain particles from the final reconstruction in single-particle EM. However, site-specific damage to acidic side chains has already been observed at 3.2 Å in $10 \text{ e}^- \text{ \AA}^{-2}$ exposures (Bartesaghi et al., 2014). We surmise that the underlying damage to the sample may limit the attainable resolution in single-particle cryo-EM at the doses currently used (Grant and Grigorieff, 2015; Merk et al., 2016), and that with such exposure levels the collected data are of damaged particles. When known structures are reconstructed by single-particle EM, even if the resulting map is noisy because of damage, it can still be readily interpreted because the correct answer is available. But for novel structures, where the correct structure is unknown, building structures *de novo* with noisy maps of damaged protein is very challenging and could be prohibitive.

Quantification of radiation damage and estimates of crystal lifetime under irradiation not only depend on the sample (e.g., the number of scattering unit cells and their size, composition and thickness of the surrounding mother liquor and the embedding vitrified ice), and the measurement setup (e.g., quanta, dose, and temperature), but also on how data were processed and analyzed. Measures such as the upper resolution limit, B factors, and unit cell volume are often the result of some optimization procedure and may be affected by factors other than the actual damage to the crystal (Kmetko et al., 2006). This is reflected in the literature by the wide spread of dose limits. The D_{50} value of $2.2 \text{ e}^- \text{ \AA}^{-2}$ calculated from reflections of proteinase K in the 21.0–1.7 Å interval is consistent with past measurements using electron diffraction from two-dimensional crystals (Stark et al., 1996) as well as three-dimensional crystals (Unwin and Henderson, 1975; Jeng and Chiu, 1984; Baker et al., 2010). When D_{50} was calculated from the reflections in the 14.0–1.7 Å interval of the much smaller hepta-peptide, its value was 2.0 and $2.2 \text{ e}^- \text{ \AA}^{-2}$ for high and low exposure rates, respectively, very close to the value obtained from proteinase K. In line with previous studies in synchrotron X-ray crystallography at comparable flux densities and temperatures, we do not see any effects from the dose rate on the observed global damage (Holton, 2009; Warkentin et al., 2013).

Given that the absorbed dose will ultimately limit the amount of meaningful data that can be extracted from a sample, data collection in the face of radiation damage may be viewed as an optimization problem. The more electrons are delivered to the sample in cryo-EM the stronger the signal, but then noise and damage accumulate and the high-resolution information suffers. Where each sample need only be exposed once, e.g., single-particle cryo-EM or serial femtosecond crystallography (Schlichting, 2015), the exposure can be tuned to maximize the signal-to-noise ratio. When pictures are recorded in

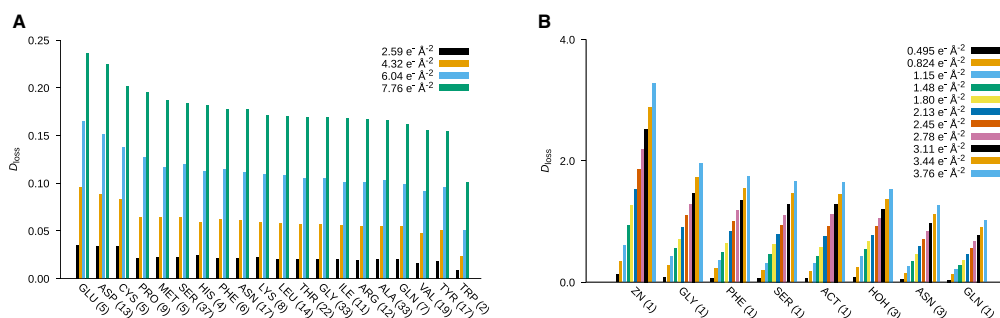


Figure 5. Accumulated Density Loss

(A and B) Density loss in arbitrary units for all the amino acids, ligands, and ions present in the refined models of (A) proteinase K and (B) the hepta-peptide. The entities are sorted in the approximate order of damage onset. The numbers in parentheses denote the occurrence of the respective amino acid in the structure. Only reflections in the resolution range common to all datasets are considered (20.8–3.20 Å for proteinase K; 14.0–1.45 Å for GSNQNNF).

cryo-EM a minimum level of exposure is required so that sufficient signal is recorded on the camera to facilitate phase contrast and faithful reconstructions. This need for phase contrast makes it hard to lower the total dose in these experiments where it was recommended to use $10 \text{ e}^- \text{ \AA}^{-2}$ for optimal trade-off between signal and noise for the target resolution of only 3 Å (Baker et al., 2010).

In MicroED, where phases are lost and only amplitudes are recorded, the minimum exposure necessary for recording the signal is significantly lower than in single-particle EM. This allows data collection at extremely low exposures $<0.01 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$ and entire datasets to be collected from a total exposure less than a single electron per \AA^2 at which point atomic resolution information can be preserved. Our estimated D_{50} rate of approximately $2 \text{ e}^- \text{ \AA}^{-2}$ sets an upper target for single-particle cryo-EM experiments seeking to maximize resolution. This is particularly challenging, but with increasingly sensitive cameras and better algorithms to allow the use of the first few frames of the recorded movies one might be able to achieve such a feat.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
 - Proteinase K
 - GSNQNNF
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

ACKNOWLEDGMENTS

The Gonen laboratory is supported by funds from the Howard Hughes Medical Institute. J.A.R. is partly supported by NSF grant no. DMR 1548924, DOE grant DE-FC02-02ER63421, and as a Searle Scholar and a Beckman Young Investigator. M.G.J. is supported by a QCB Collaboratory Postdoctoral Fellowship.

AUTHOR CONTRIBUTIONS

Conceptualization, J.A.R. and T.G.; Investigation, D.S., C.G., and C.-T.Z.; Formal Analysis, J.H., C.-T.Z., M.G.-J., and M.W.M.; Writing, J.H., D.S., J.A.R., and T.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 9, 2017

Revised: February 1, 2018

Accepted: March 30, 2018

Published: April 26, 2018

REFERENCES

- Baker, L.A., Smith, E.A., Bueler, S.A., and Rubinstein, J.L. (2010). The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. *J. Struct. Biol.* *169*, 431–437.
- Barad, B.A., Echols, N., Wang, R.Y.R., Cheng, Y., Dimaio, F., Adams, P.D., and Fraser, J.S. (2015). EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* *12*, 943–946.
- Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A., and Subramaniam, S. (2014). Structure of α -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA* *111*, 11709–11714.
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* *67* (Pt 4), 271–281.
- Blake, C.C.F., and Phillips, D.C. (1962). Effects of X-irradiation on Single Crystals of Myoglobin (The Royal Institution), pp. 183–191.
- Bury, C., Garman, E.F., Ginn, H.M., Ravelli, R.B.G., Carmichael, I., Kneale, G., and McGeehan, J.E. (2015). Radiation damage to nucleoprotein complexes in macromolecular crystallography. *J. Synchrotron Radiat.* *22*, 213–224.
- 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., et al. (2017). Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* *14*, 399–402.
- Diederichs, K. (2006). Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* *62*, 96–101.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* *69*, 1204–1214.

- Garman, E.F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. D Biol. Crystallogr.* **66**, 339–351.
- Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**, <https://doi.org/10.7554/eLife.06980>.
- Hattne, J., Reyes, F.E., Nannenga, B.L., Shi, D., de la Cruz, M.J., Leslie, A.G.W., and Gonen, T. (2015). MicroED data collection and processing. *Acta Crystallogr. A Found. Adv.* **71**, 353–360.
- Hattne, J., Shi, D., de la Cruz, M.J., Reyes, F.E., and Gonen, T. (2016). Modeling truncated pixel values of faint reflections in MicroED images. *J. Appl. Crystallogr.* **49**, 1029–1034.
- Hayward, S.B., and Glaeser, R.M. (1979). Radiation damage of purple membrane at low temperature. *Ultramicroscopy* **4**, 201–210.
- Helliwell, J.R. (1988). Protein crystal perfection and the nature of radiation damage. *J. Cryst. Growth* **90**, 259–272.
- Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193.
- Henderson, R., and Unwin, P.N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32.
- Holton, J.M. (2009). A beginner's guide to radiation damage. *J. Synchrotron Radiat.* **16** (Pt 2), 133–142.
- Howell, P.L., and Smith, G.D. (1992). Identification of heavy-atom derivatives by normal probability methods. *J. Appl. Crystallogr.* **25**, 81–86.
- Howells, M.R., Beetz, T., Chapman, H.N., Cui, C., Holton, J.M., Jacobsen, C.J., Kirz, J., Lima, E., Marchesini, S., Miao, H., et al. (2009). An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy. *J. Electron Spectrosc. Relat. Phenom.* **170**, 4–12.
- Jeng, T.-W., and Chiu, W. (1984). Quantitative assessment of radiation damage in a thin protein crystal. *J. Microsc.* **136**, 35–44.
- Jones, G.D.D., Lea, J.S., Symons, M.C.R., and Taiwo, F.A. (1987). Structure and mobility of electron gain and loss centres in proteins. *Nature* **330**, 772–773.
- Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66** (Pt 2), 133–144.
- Kabsch, W. (2010b). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66** (Pt 2), 125–132.
- Kmetko, J., Husseini, N.S., Naides, M., Kalinin, Y., and Thorne, R.E. (2006). Quantifying X-ray radiation damage in protein crystals at cryogenic temperatures. *Acta Crystallogr. D Biol. Crystallogr.* **62** (Pt 9), 1030–1038.
- Leslie, A.G.W., and Powell, H.R. (2007). Processing diffraction data with mosflm. In *Evolving Methods for Macromolecular Crystallography* (NATO Science Series II: Mathematics, Physics and Chemistry), R.J. Read and J.L. Sussman, eds. (Springer), pp. 41–51.
- Liebschner, D., Rosenbaum, G., Dauter, M., and Dauter, Z. (2015). Radiation decay of thaumatin crystals at three X-ray energies. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 772–778.
- Martynowicz, M., Glynn, C., Miao, J., de la Cruz, M.J., Hattne, J., Shi, D., Cascio, D., Rodriguez, J., and Gonen, T. (2017). MicroED structures from micrometer thick protein crystals. *bioRxiv*. <https://doi.org/10.1101/152504>.
- Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Prangani, R., Boxer, M.B., Earl, L.A., Milne, J.L.S., and Subramaniam, S. (2016). Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* **165**, 1698–1707.
- Murray, J., and Garman, E. (2002). Investigation of possible free-radical scavengers and metrics for radiation damage in protein cryocrystallography. *J. Synchrotron Radiat.* **9**, 347–354.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67** (Pt 4), 355–367.
- Nannenga, B.L., Shi, D., Hattne, J., Reyes, F.E., and Gonen, T. (2014a). Structure of catalase determined by MicroED. *Elife* **3**, e03600.
- Nannenga, B.L., Shi, D., Leslie, A.G.W., and Gonen, T. (2014b). High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930.
- Nave, C., and Hill, M.A. (2005). Will reduced radiation damage occur with very small crystals? *J. Synchrotron Radiat.* **12** (Pt 3), 299–303.
- Oliphant, T.E. (2007). Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20.
- Ravelli, R.B.G., Theveneau, P., McSweeney, S., and Caffrey, M. (2002). Unit-cell volume change as a metric of radiation damage in crystals of macromolecules. *J. Synchrotron Radiat.* **9**, 355–360.
- Sanishvili, R., Yoder, D.W., Pothineni, S.B., Rosenbaum, G., Xu, S., Vogt, S., Stepanov, S., Makarov, O.A., Corcoran, S., Benn, R., et al. (2011). Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *Proc. Natl. Acad. Sci. USA* **108**, 6127–6132.
- Schlichting, I. (2015). Serial femtosecond crystallography: the first five years. *IUCrJ* **2**, 246–255.
- Schrödinger, L.L.C. (2014). The PyMOL Molecular Graphics System (Schrödinger), Available at: <https://www.pymol.org>
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. A Found. Crystallogr.* **64** (Pt 1), 112–122.
- Shi, D., Nannenga, B.L., Iadanza, M.G., and Gonen, T. (2013). Three-dimensional electron crystallography of protein microcrystals. *Elife* **2**, e01345.
- Shi, D., Nannenga, B.L., de la Cruz, M.J., Liu, J., Sawtelle, S., Calero, G., Reyes, F.E., Hattne, J., and Gonen, T. (2016). The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904.
- Stark, H., Zemlin, F., and Boettcher, C. (1996). Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. *Ultramicroscopy* **63**, 75–79.
- Unwin, P.N.T., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* **94**, 425–440.
- Urzhumtseva, L.M., Klaholz, B., and Urzhumtsev, A.G. (2013). On effective and optical resolutions of diffraction data sets. *Acta Crystallogr. D Biol. Crystallogr.* **69** (Pt 10), 1921–1934.
- Uyeda, N., Kobayashi, T., Ishizuka, K., and Fujiyoshi, Y. (1980). Crystal structure of Ag-TCNQ. *Nature* **285**, 95–97.
- Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
- Warkentin, M., Hopkins, J.B., Badeau, R., Mulichak, A.M., Keefe, L.J., and Thorne, R.E. (2013). Global radiation damage: temperature dependence, time dependence and how to outrun it. *J. Synchrotron Radiat.* **20** (Pt 1), 7–13.
- Weik, M., Ravelli, R.B.G., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci. USA* **97**, 623–628.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B.E.B., McCoy, A.J., McNicholas, S.J., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67** (Pt 4), 235–242.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Proteinase K	Sigma-Aldrich	Cat#P2308; CAS: 39450-01-6
Ammonium sulfate	Sigma-Aldrich	Cat#A4418; CAS: 7783-20-2
BIS-TRIS	Sigma-Aldrich	Cat#B9754; CAS: 6976-37-0
GSQNNF	GenScript	N/A
Polyethylene glycol 8,000	Hampton Research	Cat#HR2-535; CAS: 25322-68-3
MES sodium salt	Sigma-Aldrich	Cat#M3885; CAS: 71119-23-8
Zinc acetate dihydrate	Sigma-Aldrich	Cat#96459; CAS: 5970-45-6
Deposited Data		
Atomic coordinates, proteinase K crystal structure	(Hattne et al., 2016)	PDB: 5i9s
Atomic coordinates and density map of proteinase K at 0.86 e ⁻² Å ⁻²	this paper	PDB: 6cl7; EMDB: EMD-7490
Atomic coordinates and density map of proteinase K at 2.6 e ⁻² Å ⁻²	this paper	PDB: 6cl8; EMDB: EMD-7491
Atomic coordinates and density map of proteinase K at 4.3 e ⁻² Å ⁻²	this paper	PDB: 6cl9; EMDB: EMD-7492
Atomic coordinates and density map of proteinase K at 6.0 e ⁻² Å ⁻²	this paper	PDB: 6cla; EMDB: EMD-7493
Atomic coordinates and density map of proteinase K at 7.8 e ⁻² Å ⁻²	this paper	PDB: 6clb; EMDB: EMD-7494
Atomic coordinates and density map of GSQNNF at 0.27 e ⁻² Å ⁻²	this paper	PDB: 6clc; EMDB: EMD-7495
Atomic coordinates and density map of GSQNNF at 0.81 e ⁻² Å ⁻²	this paper	PDB: 6cld; EMDB: EMD-7496
Atomic coordinates and density map of GSQNNF at 1.3 e ⁻² Å ⁻²	this paper	PDB: 6cle; EMDB: EMD-7497
Atomic coordinates and density map of GSQNNF at 1.9 e ⁻² Å ⁻²	this paper	PDB: 6clf; EMDB: EMD-7498
Atomic coordinates and density map of GSQNNF at 2.4 e ⁻² Å ⁻²	this paper	PDB: 6clg; EMDB: EMD-7499
Atomic coordinates and density map of GSQNNF at 2.9 e ⁻² Å ⁻²	this paper	PDB: 6clh; EMDB: EMD-7500
Atomic coordinates and density map of GSQNNF at 0.17 e ⁻² Å ⁻²	this paper	PDB: 6cli; EMDB: EMD-7501
Atomic coordinates and density map of GSQNNF at 0.50 e ⁻² Å ⁻²	this paper	PDB: 6clj; EMDB: EMD-7502
Atomic coordinates and density map of GSQNNF at 0.82 e ⁻² Å ⁻²	this paper	PDB: 6clk; EMDB: EMD-7503
Atomic coordinates and density map of GSQNNF at 1.2 e ⁻² Å ⁻²	this paper	PDB: 6cll; EMDB: EMD-7504
Atomic coordinates and density map of GSQNNF at 1.5 e ⁻² Å ⁻²	this paper	PDB: 6clm; EMDB: EMD-7505
Atomic coordinates and density map of GSQNNF at 1.8 e ⁻² Å ⁻²	this paper	PDB: 6cln; EMDB: EMD-7506
Atomic coordinates and density map of GSQNNF at 2.1 e ⁻² Å ⁻²	this paper	PDB: 6clo; EMDB: EMD-7507
Atomic coordinates and density map of GSQNNF at 2.5 e ⁻² Å ⁻²	this paper	PDB: 6clp; EMDB: EMD-7508

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Atomic coordinates and density map of GSQNNF at 2.8 e ⁻ Å ⁻²	this paper	PDB: 6clq; EMDB: EMD-7509
Atomic coordinates and density map of GSQNNF at 3.1 e ⁻ Å ⁻²	this paper	PDB: 6clr; EMDB: EMD-7510
Atomic coordinates and density map of GSQNNF at 3.4 e ⁻ Å ⁻²	this paper	PDB: 6cls; EMDB: EMD-7511
Atomic coordinates and density map of GSQNNF at 3.8 e ⁻ Å ⁻²	this paper	PDB: 6clt; EMDB: EMD-7512
Software and Algorithms		
TVIPS tools	(Hattne et al., 2015)	N/A
iMosflm	(Leslie and Powell, 2007; Battye et al., 2011)	RRID:SCR_014217
AIMLESS	(Evans and Murshudov, 2013)	RRID:SCR_015747
MOLREP	(Vagin and Teplyakov, 1997)	RRID:SCR_007255
XDS	(Kabsch, 2010b)	RRID:SCR_015652
XSCALE	(Kabsch, 2010a)	RRID:SCR_015652
XDSCONV	(Kabsch, 2010b)	RRID:SCR_015652
SHELXD	(Sheldrick, 2008)	RRID:SCR_014220
SCALEIT	(Howell and Smith, 1992)	RRID:SCR_007255
EFRESOL	(Urzhumtseva et al., 2013)	N/A
REFMAC	(Murshudov et al., 2011)	RRID:SCR_014225
AREAIMOL	(Winn et al., 2011)	RRID:SCR_007255
RIDL	(Bury et al., 2015)	N/A
PyMol	(Schrödinger, 2014)	RRID:SCR_000305
Other		
TEM grids	Quantifoil	N/A
easiGlow glow discharge cleaning system	PELCO	N/A
Vitrobot Mark IV plunge-freezer	Thermo Fisher	N/A
Gatan 626 cryo-transfer holder	Gatan	N/A
FEI Tecnai F20	Thermo Fischer	N/A
TemCam-F416	TVIPS	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tamir Gonen (tgonen@ucla.edu).

METHOD DETAILS

Proteinase K

Crystal Growth

Proteinase K from *Engyodontium album* (Sigma-Aldrich, St Louis, MO, USA) was prepared by combining 2 ml of protein solution (50 mg ml⁻¹) with 2 ml of precipitant solution (1.0–1.3 M ammonium sulfate, 0.1 M Tris pH 8.0) (Hattne et al., 2016).

Sample Preparation

The protein solution was dispensed on a glow-discharged grid (easiGlow; Pelco) and vitrified with force position 24 in an FEI (now Thermo Fisher) Vitrobot Mark IV after blotting for 12 s at an environment humidity of 30% (Shi et al., 2016). Frozen-hydrated grids were loaded onto a Gatan 626 cryo-holder and transferred to the microscope, where the specimen temperature was maintained at ~100 K.

Data Collection

Electron diffraction datasets from separate crystals were collected using an FEI Tecnai F20 transmission electron microscope operated at 200 kV, with the objective aperture fully open to evenly illuminate an area extending beyond the sample and setting the

selected area aperture to closely match the size of the crystal. For each crystal of thickness 200–400 nm, the same 23° wedge was repeatedly collected up to five times by continuously rotating the stage from -12° to $+11^\circ$ (-38° to -15° for crystal 3) off its untilted orientation at a constant rate of $0.089^\circ \text{ s}^{-1}$ (Nannenga et al., 2014b). The rate of electron exposure was adjusted to $0.007 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$, calibrated using a Faraday cage. The individual datasets, each consisting of 49–50 frames with exposure time 5.1 s were recorded at a camera length setting of 1.2 m, corresponding to an effective detector distance of 2.2 m. All diffraction images were acquired using a TVIPS TemCam-F416 CMOS camera and corrected to account for negative pixel values (Hattne et al., 2015, 2016) prior to further processing.

Data Reduction

Proteinase K data were indexed and integrated in $P4_32_12$ using MOSFLM (Leslie and Powell, 2007) through its graphical interface iMosflm (Battye et al., 2011). Wedges from six different crystals were merged by the order in which they were collected using AIMLESS (Evans and Murshudov, 2013), and the set of free reflections was copied from the molecular replacement search model, PDB entry 5i9s. Neither of these crystals yield a complete dataset on their own, but since proteinase K does not exhibit a pronounced preferred orientation, this produced five reasonably complete datasets, each comprised of frames with a similar degree of exposure. The choice between intensities derived from summation integration and profile fitting was left up to the optimization algorithm implemented in AIMLESS; in all cases this resulted in profile-fitted intensities being used. Relative B -factors were calculated between merged single-crystal datasets with SCALEIT from the CCP4 suite (Howell and Smith, 1992; Winn et al., 2011).

To quantify the global effects of radiation damage, integrated intensities were averaged for each diffraction image. Averages from different crystals were scaled by a single factor in the range [0.10, 1.0] and simultaneously fit to a common function on the form $A \times \exp(-B \times x)$ using the BFGS minimizer implemented in scipy (Oliphant, 2007). The mean effective resolution was calculated by EFRESOL (Urzhumtseva et al., 2013) and used as an objective high-resolution cutoff for all datasets.

Phasing and Model Refinement

The first dataset was phased by molecular replacement using MOLREP (Vagin and Teplyakov, 1997) from PDB entry 5i9s, and the solution was reused for all subsequent datasets. Water molecules and ions were excluded from the refined structure: while these improve the quality of the model at high resolution, they are difficult to reliably model once damage degrades the quality of the data. This model was also used to calculate solvent-accessible areas with AREAIMOL (Winn et al., 2011).

All models were refined with REFMAC (Murshudov et al., 2011), with electron scattering factors calculated using the Mott–Bethe formula. The occupancies were set to unity for all atoms and no alternate confirmations were used to model partial damage to specific sites of the molecule. Further processing and refinement statistics are given in Table 1.

GSNQNNF

Crystal Growth

The 7-residue peptide GSNQNNF (>98% purity) was purchased from GenScript, dissolved in water at 10 mg ml^{-1} , and crystallized by the hanging-drop method in a high-throughput screen. Crystals grew as needle clusters at a 1:1 ratio of peptide solution to mother liquor in a condition containing 10% (w/v) PEG-8000, 0.1 M MES pH 6.0, and $\text{Zn}(\text{OAc})_2$ (Martynowycz et al., 2017).

Sample Preparation

Clusters were broken by pipetting and dispensed onto glow-discharged grids, which were then blotted for 20 s and vitrified with force position 24. Otherwise, GSNQNNF samples were prepared identically to those of proteinase K.

Data Collection

Crystals of the hepta-peptide that were 100–500 nm thick, were tilted over $\sim 60^\circ$ at a three-fold higher rotation rate (0.3° s^{-1}) than was used for proteinase K and up to 12 sweeps were collected from each crystal. To probe the effect of dose rate on radiation damage, peptide data were collected at both $0.0028 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$ and $0.0017 \text{ e}^- \text{ \AA}^{-2} \text{ s}^{-1}$. These rates were tuned to maximize the number of sweeps collected from an individual crystal. Single crystal datasets comprised of approximately 100 images were collected with an exposure time of 2.1 s and camera length 0.73 m which corresponds to an effective sample to detector distance of 1.2 m. Because two orders of magnitude fewer reflections are typically observed on a diffraction pattern from short segments like GSNQNNF than from proteinase K, intensities were integrated with a higher gain value and averaged for each dataset instead of for each frame when estimating the effects of global damage on the hepta-peptide. Otherwise data collection was performed as detailed for proteinase K.

Data Reduction, Phasing, and Model Refinement

The datasets were indexed and integrated in $P1$ with XDS (Kabsch, 2010b) and an isomorphous subset was scaled and merged with XSCALE (Kabsch, 2010a). Phases for the GSNQNNF data were determined *ab initio* by direct methods from the first collected data set using SHELXD (Sheldrick, 2008). XDSCONV (Kabsch, 2010b) was used on this dataset to assign a free set of reflections, which was subsequently reused for all later peptide datasets. A ligated acetate, three water molecules, and a single zinc atom were included with the GSNQNNF model, because they constitute a significant fraction of the unit cell contents, and all atoms were fixed at full occupancy. Otherwise processing was performed as detailed for proteinase K; statistics for the datasets at the high and low dose rates are given in Tables 2 and 3, respectively.



QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification and statistical analyses are given in [Tables 1, 2, and 3](#). These values were extracted from the programs used to merge and refine the respective multi-crystal datasets.

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates have been deposited in the Protein Data Bank under ID codes 6cl7 to 6clt. Density maps have been deposited in the Electron Microscopy Data Bank under ID codes EMD-7490 to EMD-7512.

Appendix Chapter 3

Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein
InaZ

The work described in this chapter has been reproduced from:

Chih-Te Zee, Calina Glynn, Marcus Gallagher-Jones, Jennifer Miao, Carlos G. Santiago, Duilio Cascio, Tamir Gonen, Michael R. Sawaya and Jose A. Rodriguez. "Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ." *IUCrJ*, 2019.

Copyright 2019

Chih-Te Zee, Calina Glynn, Marcus Gallagher-Jones, Jennifer Miao, Carlos G. Santiago, Duilio Cascio, Tamir Gonen, Michael R. Sawaya and Jose A. Rodriguez.



Homochiral and racemic MicroED structures of a peptide repeat from the ice-nucleation protein InaZ

Chih-Te Zee,^a Calina Glynn,^a Marcus Gallagher-Jones,^a Jennifer Miao,^a Carlos G. Santiago,^a Duilio Cascio,^b Tamir Gonen,^c Michael R. Sawaya^b and Jose A. Rodriguez^{a*}Received 28 August 2018
Accepted 12 December 2018

Edited by F. Sun, Chinese Academy of Sciences, China

Keywords: amyloid; racemic; electron diffraction; ice nucleation; intermolecular interactions; co-crystals; electron crystallography; structural biology.**PDB references:** racemic GSTSTA, X-ray structure, 6m7m; MicroED structure, 6m9j; L-GSTSTA, X-ray structure, 6eex; MicroED structure, 6m9i**Supporting information:** this article has supporting information at www.iucrj.org

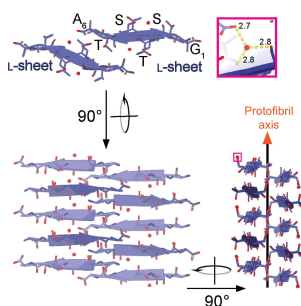
^aDepartment of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095, USA, ^bDepartment of Biological Chemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California Los Angeles, Los Angeles, CA 90095, USA, and ^cHoward Hughes Medical Institute, Departments of Physiology and Biological Chemistry, University of California Los Angeles, Los Angeles, CA 90095, USA. *Correspondence e-mail: jrodriguez@mbi.ucla.edu

The ice-nucleation protein InaZ from *Pseudomonas syringae* contains a large number of degenerate repeats that span more than a quarter of its sequence and include the segment GSTSTA. *Ab initio* structures of this repeat segment, resolved to 1.1 Å by microfocus X-ray crystallography and to 0.9 Å by the cryo-EM method MicroED, were determined from both racemic and homochiral crystals. The benefits of racemic protein crystals for structure determination by MicroED were evaluated and it was confirmed that the phase restriction introduced by crystal centrosymmetry increases the number of successful trials during the *ab initio* phasing of the electron diffraction data. Both homochiral and racemic GSTSTA form amyloid-like protofibrils with labile, corrugated antiparallel β -sheets that mate face to back. The racemic GSTSTA protofibril represents a new class of amyloid assembly in which all-left-handed sheets mate with their all-right-handed counterparts. This determination of racemic amyloid assemblies by MicroED reveals complex amyloid architectures and illustrates the racemic advantage in macromolecular crystallography, now with submicro-metre-sized crystals.

1. Introduction

Expressed by a subset of microorganisms, ice-nucleation proteins are capable of stimulating ice formation in super-cooled water (Green & Warren, 1985). The Gram-negative microbe *Pseudomonas syringae* is sold commercially as Snomax[®] for its ice-nucleating activity (Green & Warren, 1985; Cochet & Widehem, 2000). The ice-nucleation protein InaZ is produced by *P. syringae* and localized to its outer membrane (Green & Warren, 1985; Wolber *et al.*, 1986). The sequence of InaZ is approximately 1200 residues in length, over half of which includes degenerate octapeptide repeats. A subpopulation of degenerate repeats share the consensus motif GSTXT(A/S), where X represents an unconserved amino acid (Supplementary Fig. S1; Green & Warren, 1985; Warren *et al.*, 1986). These repeats are shared by other Ina proteins and may collectively contribute to ice nucleation (Green & Warren, 1985; Kobashigawa *et al.*, 2005; Han *et al.*, 2017).

Despite the crystallographic determination of structures of other ice-binding proteins (Davies, 2014; Garnham, Campbell & Davies, 2011), InaZ remains recalcitrant to crystallization. Models of full-length InaZ have proposed it to have a β -helical (Garnham, Campbell, Walker *et al.*, 2011; Graether & Jia, 2001) or solenoid-like fold rich in stacked β -strands (Cochet &



OPEN ACCESS

Widehem, 2000; Pandey *et al.*, 2016). These features are shared by amyloid filaments: their tightly mated β -sheets form fibrils that can cross-link, cluster and be functional (Nelson *et al.*, 2005; Sawaya *et al.*, 2007; Fitzpatrick *et al.*, 2017; Eisenberg & Jucker, 2012; Maury, 2009). Functional amyloid assemblies appear across the tree of life (Wasmer *et al.*, 2008; Hughes *et al.*, 2018; Maury, 2009; Tayeb-Fligelman *et al.*, 2017) and can contain low-complexity regions with degenerate repeats (Hughes *et al.*, 2018).

Success in determining amyloid structures was first achieved by crystallizing short segments that stabilize the cores of fibrils through a motif known as the steric zipper (Nelson *et al.*, 2005; Sawaya *et al.*, 2007). However, the propensity of elongated β -strands to twist or kink can limit crystal growth, sometimes yielding nanocrystals that pose a challenge for structure determination (Rodríguez *et al.*, 2015). These limits have recently been overcome in part by the development of the cryo-electron microscopy (cryo-EM) method, electron microdiffraction (MicroED; Shi *et al.*, 2013). MicroED yields high-resolution structures from protein crystals no thicker than a few hundred nanometres (Shi *et al.*, 2016; Rodríguez *et al.*, 2017). Because of this, MicroED has helped in determining the structures of a number of amyloid protofibrils (Rodríguez *et al.*, 2015; Krotee *et al.*, 2017) at atomic resolution, some *ab initio* (Sawaya *et al.*, 2016; Gallagher-Jones *et al.*, 2018).

Racemic crystallography further facilitates the crystallization of proteins and peptides (Matthews, 2009; Yeates & Kent, 2012; Patterson *et al.*, 1999), including ice-binding proteins (Pentelute *et al.*, 2008). Mixing left-handed (L) and right-handed (D) enantiomers of a macromolecule improves its likelihood of crystallization and facilitates structural analysis (Yeates & Kent, 2012; Wukovitz & Yeates, 1995). Crystallographic phases are restricted for data from centrosymmetric crystals, making the phase problem associated with the determination of their structure more tractable (Yeates & Kent, 2012). This is advantageous for structure determination by direct methods (Hauptman, 1986), where phases must be computed from measured intensities alone (Hauptman, 1986, 2001; Sheldrick, 2008). Accordingly, various polypeptide structures have been determined by racemic X-ray crystallography, including those of ester insulin, plectasin and an antifreeze protein (Pentelute *et al.*, 2008; Avital-Shmilovici *et al.*, 2013; Mandal *et al.*, 2009, 2012). While the benefits of racemic crystallography are evident for X-ray diffraction (Matthews, 2009), questions remain about the potential for exploiting these benefits in MicroED.

Hypothesizing that the repeat segments of the ice-nucleation protein InaZ may form amyloid-like assemblies, we set out to interrogate the structure of GSTSTA from both homochiral and racemic crystals by MicroED. In doing so, we also assessed the fidelity of MicroED data in racemic structure determination. By comparing the structures of homochiral and racemic GSTSTA, we gauge the effect of racemic self-assembly on protofibril architecture. With these structures of a core repeat in the InaZ protein, we begin an atomic-level investigation of peptide segments derived from ice-nucleation proteins (Pandey *et al.*, 2016).

2. Methods

2.1. Sequence analysis of ice-nucleation proteins

The sequence of the ice-nucleation protein InaZ from *P. syringae* was screened for the existence of hexameric degenerate repeat motifs that contained one or more threonine residues (Supplementary Fig. S1). The repeats were then evaluated for their propensity for amyloid fibril formation by ZipperDB (Supplementary Fig. S1). For each, a Rosetta energy score was calculated. A single repeat, GSTSTA, was chosen from this list of hexameric sequences. This segment appears five times identically in the sequence of InaZ, first at residue 707, and is part of a group with the consensus motif GSTXT(A/S) that appears 59 times in the InaZ sequence.

2.2. Synthesis, purification, characterization and crystallization of L- and D-enantiomers of the InaZ-derived peptide GSTSTA

The L-enantiomer of GSTSTA was purchased from GenScript with 98% purity. The D-enantiomer of GSTSTA was synthesized by solid-phase peptide synthesis and was purified using a Waters Breeze 2 HPLC System in reversed phase buffered with trifluoroacetic acid (Supplementary Fig. S2). The two enantiomers were qualified by ESI-MS on a Waters LCT Premier. The spectrum of the L-enantiomer showed an $[M+H]^+$ peak of 523.30 g mol⁻¹ (expected 523.22 g mol⁻¹) and a dimer $[M+M+H]^+$ peak of 1045.6 g mol⁻¹ (expected 1045.44 g mol⁻¹). The spectrum of the D-enantiomer showed an $[M+H]^+$ peak of 523.24 g mol⁻¹ (expected 523.22 g mol⁻¹) and a dimer $[M+M+H]^+$ peak of 1045.49 g mol⁻¹ (expected 1045.44 g mol⁻¹) (Supplementary Fig. S2).

Crystals of L-GSTSTA were grown as follows. Lyophilized peptide was weighed and dissolved in ultrapure water at concentrations of between 82 and 287 mM, assuming a 1:1 ratio of peptide to trifluoroacetic acid (TFA) in the lyophilized powder. Crystals were grown at room temperature by the hanging-drop method in a high-content 96-well Wizard screen. Of the many crystallization trials that yielded crystals, those obtained from a condition consisting of 0.1 M CHES buffer pH 9.5, 10% (w/v) PEG 3000 were used for microfocus X-ray data collection. Another promising condition was optimized by the hanging-drop method in 24-well plates. This condition consisted of 0.1 M McIlvaine (citrate-phosphate) buffer pH 4.2, 12.5% (w/v) PEG 8000, 0.1 M sodium chloride and was used to grow crystals of L-GSTSTA in batch.

Crystals of racemic GSTSTA were grown as follows. Lyophilized powders of L-GSTSTA and D-GSTSTA were separately weighed and dissolved in ultrapure water so that the concentrations of the two enantiomers matched. Crystal formation was screened at concentrations ranging from 82 to 123 mM after accounting for TFA. Control trays containing only L- or D-GSTSTA were prepared simultaneously alongside racemic screens. All three trays were stored and monitored at room temperature, with crystal formation observed in various conditions. Images of every well were collected after 3 h, one day, three days, five days and seven days, and crystal formation was monitored over time. A condition consisting of 0.1 M

imidazole pH 8.0, 10% (w/v) PEG 8000 produced the best crystals.

Crystals were batch grown for data collection by MicroED. Lyophilized L-GSTSTA peptide was weighed and dissolved in 0.1 M McIlvaine (citrate–phosphate) buffer pH 4.2, 12.5% (w/v) PEG 8000, 0.1 M sodium chloride to an effective final concentration of 123 mM, mimicking the final concentration of a hanging drop in the 24-well optimization. Lastly, the solution was seeded with crystal needles extracted from crystals grown in the 24-well optimization described above. Batch crystals of racemic GSTSTA were grown from lyophilized L-GSTSTA and D-GSTSTA that had been separately weighed and dissolved in 0.1 M imidazole buffer pH 8.0 containing 10% (w/v) PEG 8000 to a final concentration of 50 mM for each enantiomer after accounting for the mass contributed by TFA.

2.3. Microfocus X-ray data collection and structure determination

Crystals of L-GSTSTA were harvested from a 96-well hanging drop using MiTeGen loops and flash-cooled in liquid nitrogen. No additional cryoprotectant was used other than the PEG 3000 that was already present in the mother liquor. 72 diffraction images were collected with an oscillation range of 5° from a single crystal; 40 of these were indexed and integrated. Crystals of racemic GSTSTA were harvested from a 96-well hanging drop using MiTeGen loops and flash-cooled in liquid nitrogen. No additional cryoprotectant was used other than the PEG 8000 that was already present in the buffer. 144 diffraction images were collected with an oscillation range of 2.5° from a single crystal; 64 of these were indexed and integrated.

Diffraction data were collected from both homochiral and racemic GSTSTA crystals 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 wavelength 0.979 Å. Signal was only limited by the edge of our detector at approximately 1.1 Å; as such, higher resolution data could perhaps be achieved by modifying the experimental geometry and/or adjusting the energy of the X-ray beam in the experiment. Data from both homochiral and racemic crystals were reduced in *XDS* (Kabsch, 2010) and yielded *ab initio* solutions using *SHELXT* and *SHELXD* (Sheldrick, 2015). The phases obtained from these coordinates produced maps of sufficient quality for subsequent model building in *Coot* (Emsley *et al.*, 2010). The resulting models were refined against the measured data using *PHENIX* (Adams *et al.*, 2010).

2.4. Electron microscopy, MicroED data collection and structure determination

Crystals were prepared for MicroED data collection following a variation of the procedures detailed in Rodriguez *et al.* (2015) as follows. Following a 1:2 dilution in ultrapure water, crystals were applied onto glow-discharged grids of type (PELCO easiGlow) 300 mesh Cu 1/4. Excess liquid was

blotted off onto filter paper wetted with 4 µl ultrapure water to avoid salt-crystal formation. Grids were plunge-frozen into liquid ethane using a Vitrobot (FEI). Grids were then initially stored in liquid nitrogen before being transferred to a liquid-nitrogen-cooled Gatan 626 cryo-holder for insertion and manipulation within the electron microscope.

MicroED data were collected from three submicrometre-thick needle crystals of L-GSTSTA and two submicrometre-thick needle crystals of racemic GSTSTA. Briefly, frozen hydrated crystals of either L-GSTSTA or racemic GSTSTA were visually inspected in overfocused diffraction mode on a cryocooled FEI Tecnai F20 microscope operated at 200 kV (Janelia Research Campus). The diffraction patterns used for structure determination were collected on a TVIPS TemCam-F416 CMOS detector in rolling-shutter mode. For L-GSTSTA, diffraction patterns were collected during unidirectional rotation with 2 s exposures. For racemic GSTSTA, diffraction patterns were collected during unidirectional rotation with 3 s exposures. A rotation rate of 0.30° s⁻¹ and rotation angles ranging from -63° to 72° were used for both. Beam intensity was held constant, with an average dose rate of 0.003–0.005 e⁻ Å⁻¹ s⁻¹ or ~0.01 e⁻ Å⁻² per image, corresponding to a total dose of ~1–3 e⁻ Å⁻² per data set. Data were recorded at an effective camera length of 730 mm, which is the equivalent of a sample-to-detector distance of 1156 mm in a corresponding lenseless system. All diffraction was performed using a circular selected area aperture of ~1 µm² in projection.

Diffraction movies were converted to the SMV file format using TVIPS tools as described previously (Hattne *et al.*, 2015). Indexing and integration were performed in *XDS*. Partial data sets from three L-GSTSTA crystals were sorted and merged in *XSCALE*. Intensities from a total of 196 diffraction images were merged. An *ab initio* solution was achieved using *SHELXD* (Sheldrick, 2015). To achieve a complete data set from racemic GSTSTA crystals, the integrated diffraction intensities from partial data sets of two different crystals were sorted and merged in *XSCALE*. Intensities from a total of 145 diffraction images were merged. An *ab initio* solution was achieved using *SHELXD* and *SHELXT* (Sheldrick, 2015). Although *XDS* accurately differentiated the Laue classification for the racemic GSTSTA data, *SHELXT*, which does not rely on user input for space-group selection, ensured a correct solution for the racemic data. *SHELXT* selected *P2₁/c* as the racemic space group, a choice corroborated by the systematic absences that were present in the data. The phases obtained from the L-GSTSTA and racemic GSTSTA coordinates produced by *SHELX* were used to generate maps of sufficient quality for subsequent model building in *Coot* (Emsley *et al.*, 2010). The resulting models were refined with *PHENIX* (Adams *et al.*, 2010), using electron scattering form factors, against the measured data.

2.5. Analysis of homochiral and racemic GSTSTA structures

In the analysis of the hydrogen-bonding and assembly interactions of each L-GSTSTA structure, an assembly of four

strands, composed of two pairs in mating sheets, was used to find all unique hydrogen bonds, while racemic GSTSTA required an assembly of 12 strands composed of three strands from a pair of mating sheets and six more strands related by translation along the protofibril axis to achieve a unique set of hydrogen bonds. Hydrogen bonds were tabulated from this structure using *HBPLUS* (McDonald & Thornton, 1994).

Distances between strands along a sheet were calculated as differences between α carbons of one strand and its neighbor along the same sheet. These distances were calculated for both GSTSTA and GNNQQNY using PDB entry 1yjp (Sawaya *et al.*, 2016). The angle between a strand and its corresponding sheet was calculated against the plane formed by α carbons along that sheet.

2.6. Analysis of phases in structures determined by MicroED and X-ray crystallography

To analyze the distribution of phases associated with reflections measured from racemic crystals by both X-ray and electron diffraction, data reduction was performed in space group 1 (*P1*) and refined in *PHENIX* against a model encompassing the entire unit cell of four strands. This model was obtained by applying all symmetry operations on the asymmetric unit of the *P2₁/c* structure. Refinement in *P1* allowed symmetry to be broken, no longer restricting phases to 0 or 180°, as the phases changed in the case where coordinates deviated from their symmetry-related positions. The resulting set of reflections and phases were analyzed in *MATLAB*. We plotted the observed and calculated magnitudes of each reflection against each other and the set fitted by linear regression. For each measured magnitude, the associated phases were plotted and showed a bimodal distribution. Histograms were drawn using these data to evaluate phase distributions; the standard deviation of these was computed by merging the distributions around 0 and 180° using a modulo operation.

2.7. Analysis of paired reflections in MicroED and X-ray crystallographic data

Merged data sets collected by either MicroED or microfocal X-ray crystallography were paired for homochiral and racemic crystals of GSTSTA. MicroED data *.mtz* files were scaled against their corresponding X-ray counterparts, where corresponding reflections were paired and missing reflections were ignored within a single *.mtz* file. This was achieved using custom scripts and the *RSTATS* program, which scaled and compared common reflections between corresponding data sets. The corresponding distributions of Fourier magnitudes were then analyzed using *MATLAB*, in which a best-fit line was determined for each of the paired data sets. Zones were visualized using the *HKLVIEW* program, in which either *h*, *k* or *l* were selectively set to zero.

3. Results

3.1. Identification, synthesis and crystallization of amyloid-forming ice-nucleation protein (INP) segments

With the goal of characterizing the structural properties of degenerate repeats in INPs, we identified a group of hexapeptides within the set of InaZ repeats and evaluated their amyloid-forming propensities (Goldschmidt *et al.*, 2010; Supplementary Fig. S1). We ranked the hexapeptides based on their predicted propensity for amyloid zipper formation, their repeated appearance in INP sequences and whether they contained polar residues, including threonine (Supplementary Fig. S1). We chose to further investigate a segment whose sequence, GSTSTA, appears identically five times within InaZ at residues 707–712, 755–760, 803–808, 851–856 and 899–904. For simplicity, we numbered the segment 707–712.

We evaluated the crystallization potentials of synthesized L- and D-enantiomers of GSTSTA (Supplementary Fig. S2) compared with that of their racemic mixture by performing high-throughput crystallization trials and monitoring crystal growth. Most crystals appeared within two weeks of the start of each trial. In some conditions crystallization was observed as early as 3 h after the start of the trial. Racemic mixtures produced a greater number of successful crystallization conditions across a broad variety of trials (Supplementary Fig. S3). The number of successful conditions that were identified for racemic mixtures outpaced those identified for each enantiomer alone (Supplementary Fig. S3), which is consistent with previous predictions (Yeates & Kent, 2012). In conditions in which both racemic and single-enantiomer crystals grew, racemic crystals appeared sooner (Supplementary Fig. S3). Minor differences in the speed of crystal appearance and the total number of conditions with identifiable crystals were also seen between L- and D-enantiomers. Fewer conditions were found to produce D-enantiomer crystals across all trials (Supplementary Fig. S3). These differences may have been a consequence of subtle inequities in the amount of residual trifluoroacetic acid (TFA) associated with each enantiomer in lyophilized powders. These effects may have been magnified by the relatively high concentrations of peptide required for crystallization of these segments (~100–150 mM).

The crystallization conditions chosen for structure determination of homochiral GSTSTA (L-GSTSTA) and racemic GSTSTA (DL-GSTSTA) yielded a high density of well ordered microcrystals, each with a unique powder diffraction pattern, indicating they had formed distinct structures (Supplementary Fig. S4). Microcrystals were optimized from these conditions for microfocal X-ray diffraction; unoptimized batch conditions yielded nanocrystal slurries that were directly suitable for MicroED. Since the powder diffraction patterns of homochiral GSTSTA crystals were identical for both enantiomers (Supplementary Fig. S4), we focused our investigation on the L-enantiomer.

3.2. *Ab initio* structure determination of L-GSTSTA

We optimized crystals of L-GSTSTA for microfocal X-ray diffraction, starting from dense needle clusters and ending

with single needles (Supplementary Fig. S5). Crystals grown in batch were monodisperse rods of 1–10 μm in length and 100–500 μm in width; these diffracted to approximately 0.9 \AA resolution by MicroED (Fig. 1). X-ray diffraction from a single crystal of L-GSTSTA yielded a 91.7% complete data set to approximately 1.1 \AA resolution (Supplementary Table S1), while data sets from three crystals of L-GSTSTA obtained by MicroED were merged to achieve a data set with an overall completeness of 86.4% at 0.9 \AA resolution. It is important to note that the X-ray data in this case were limited by the detector geometry, which could be adjusted to facilitate slightly higher resolution. Atomic structure solutions were determined for L-GSTSTA from both microfocal X-ray and MicroED data by direct methods (Sheldrick, 2008; Supplementary Fig. S6).

After 50 000 trials, *SHELXD* yielded correlation figures of merit (CFOMs) of greater than 80 for both X-ray diffraction and MicroED data (Supplementary Fig. S6; Sawaya *et al.*, 2016). The initial L-GSTSTA solution with the highest CFOM shows 33 atoms for the X-ray data set and 36 atoms for the MicroED data set (Fig. 2*a* and Supplementary Fig. S7). During refinement, the number of atoms in the X-ray structure increased to 36 peptide atoms and one bound water (Supplementary Fig. S7). The final solution achieved from the 0.9 \AA resolution MicroED data also contained 36 atoms in the peptide chain and one water molecule (Fig. 2*a*).

3.3. *Ab initio* structure determination of racemic GSTSTA from centrosymmetric crystals

Like the enantiomerically pure crystals of GSTSTA, crystals of racemic GSTSTA started as dense needle clusters and were optimized to single needles, and diffracted as single crystals on a microfocal X-ray source (Supplementary Fig. S5). Batch crystals of racemic GSTSTA were also rod-shaped and were several micrometres in length and a few hundred nanometres in thickness (Fig. 1). These were immediately suitable for MicroED and diffracted to approximately 0.9 \AA resolution (Fig. 1). Data from a single crystal obtained by X-ray diffraction produced a 93.7% complete data set at 1.1 \AA resolution, while MicroED data from two nanocrystals of racemic GSTSTA were merged to reach an overall completeness of 77.4% at 0.9 \AA resolution (Supplementary Table S1). Initial atomic structure solutions for racemic GSTSTA were obtained by direct methods (Sheldrick, 2008; Fig. 2*b* and Supplementary Fig. S7).

As with L-GSTSTA, solutions for the racemic crystals yielded correlation figures of merit (CFOMs) of greater than 80 after 50 000 trials (Supplementary Fig. S6). A comparison of the racemic GSTSTA and L-GSTSTA data sets indicated that a higher number of potentially correct solutions were found for the racemic GSTSTA data (Supplementary Fig. S6). The MicroED data show a distribution of CFOM values that is shifted towards higher values, even when truncated to 1.1 \AA resolution to match the resolution of the X-ray data sets. However, the most dramatic shift in this distribution is evident at 0.9 \AA resolution (Supplementary Fig. S6).

Initial solutions with the highest CFOM show a total of 35 peptide atoms and four waters for the structure determined from X-ray data, and a total of 36 peptide atoms and one water for that determined by MicroED (Fig. 2*b*). During refinement, the number of peptide atoms in the X-ray structure increased to 36 (Supplementary Fig. S7), while the MicroED structure gained two waters (Fig. 2*b*). Linear regression of observed to calculated structure factors for the MicroED data shows an *R* value of 0.94 and a slope of 0.97 for data reduced in space

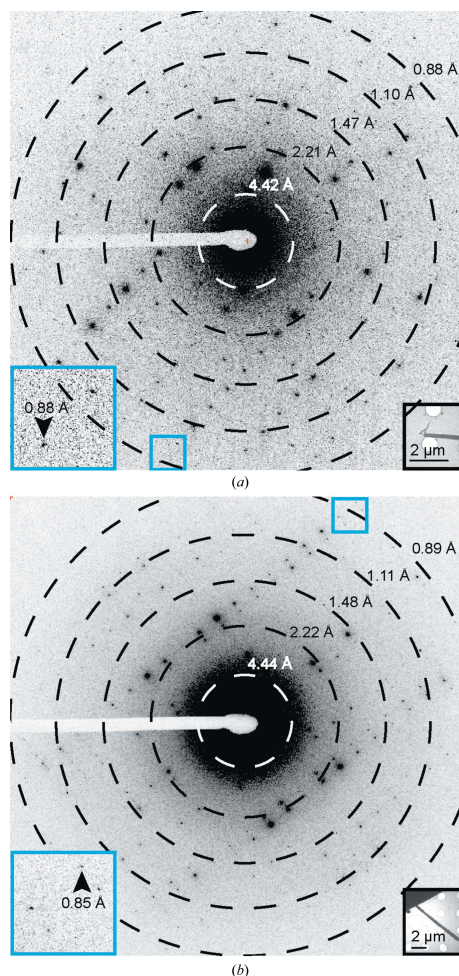


Figure 1
Single diffraction patterns of homochiral L-GSTSTA (*a*) and racemic GSTSTA (*b*) measured during continuous-rotation MicroED data collection. Each pattern corresponds to a 0.6° wedge (*a*) or a 0.9° wedge (*b*) of reciprocal space. Black insets show overfocused diffraction images of the crystals used for diffraction; blue squares correspond to magnified regions (blue insets) of the pattern that show diffraction at sub-0.9 \AA resolution (black arrows). Resolution circles are indicated by rings; scale bars are 2 μm in length.

group *P1* (Fig. 3c). These values are in good agreement with those obtained by microfocal X-ray diffraction (Supplementary Fig. S11c) and indicate a good fit between model and measurement for the racemic GSTSTA structure.

3.4. Paired comparison of Fourier magnitudes measured by X-ray crystallography or MicroED

A comparison between the X-ray and MicroED data sets for homochiral crystals of GSTSTA shows that these two types of measurement are in close agreement (Supplementary Figs. S8 and S9), although slightly higher merge errors are observed in the MicroED data across resolution bins (Supplementary Table S2). A direct comparison of Fourier magnitudes for paired reflections between these data sets is fitted by a line with a slope of 0.921 and an *R* value of 0.826 (Supplementary Fig. S8). In contrast, the comparison between X-ray and MicroED data for racemic GSTSTA shows a greater difference between the two sets and a lower *R* value for the best-fit line comparing the Fourier magnitudes of paired X-ray and MicroED reflections (Supplementary Figs. S8 and S10). This difference is likely to be owing to a lack of isomorphism between the unit cells of the racemic GSTSTA crystals used for MicroED data collection versus X-ray data collection. The unit-cell parameters for racemic GSTSTA crystals obtained by MicroED and microfocal X-ray crystallography were $a = 15.23$, $b = 9.29$, $c = 21.06$ Å, $\alpha = 90.0$, $\beta = 108.2$, $\gamma = 90.0^\circ$ and $a = 14.03$, $b = 9.22$, $c = 20.77$ Å, $\alpha = 90.0$, $\beta = 104.5$, $\gamma = 90.0^\circ$, respectively (Supplementary Table S1).

3.5. Phase restriction in centrosymmetric crystals evaluated by MicroED

Data from racemic GSTSTA crystals obtained by MicroED and reduced in the centrosymmetric space group *P21/c* satisfy refinement with imposed phases of 0 or 180° . The refinement of data from the same crystals reduced in space group *P1* results in similar residuals to those obtained for space group *P21/c* (Supplementary Table S1). The phases that result from refinement of this structure against data reduced in space group *P1* appear to be bilaterally distributed around 0 and 180° (Fig. 3a). Collapse of this bimodal phase distribution around $n\pi$ yields a standard deviation of 34.3° (Fig. 3). When the same procedure is applied to data collected from racemic GSTSTA crystals by X-ray diffraction, a similar trend appears: a normal distribution around $n\pi$ with a standard deviation of 34.4° (Supplementary Fig. S11). Bragg reflections that appear in disallowed regions of phase space (90 and 270°) for both MicroED and X-ray diffraction data are generally weakest (Fig. 3 and Supplementary Fig. S11). This suggests that the primary source of phase error

in MicroED data, as with X-ray diffraction, may come from noisy or weak reflections.

3.6. Structure of L-GSTSTA

L-GSTSTA assembles into antiparallel in-register β -sheets that mate to form a protofibril (Fig. 4a and Supplementary Figs. S12 and S13). The sheets are buckled, compressing the fibril along its length with strands spaced approximately 4.6 Å apart (Fig. 4a and Supplementary Fig. S14), closer than the typical 4.7–4.8 Å spacings seen in amyloid protofibrils (Sawaya *et al.*, 2007). This spacing equates to half of the L-GSTSTA cell edge along the *a* axis: approximately 9.2 Å (Supplementary Table S1). To accommodate this compression, the strands tilt approximately 17° with respect to the fibril axis in alternating directions along a sheet, allowing the amides to lie askew from the fibril axis (Supplementary Fig. S14) while maintaining hydrogen bonding along the protofibril axis (Supplementary Table S3). Side chains between neighboring sheets tightly interdigitate to create a close packing within the fibril (Supplementary Fig. S12); the inter-sheet distances range from 5 to 7 Å. The interface created at the fibril core is small, with 229 Å² of buried surface area, but shows a relatively high degree of shape complementarity ($S_c = 0.75$; Lawrence & Colman, 1993). The L-GSTSTA protofibril appears tightly restrained within the crystal structure, as shown by a mean *B* factor of 0.92 Å². The modeled water molecule also appears to be well ordered, particularly in the structure of L-GSTSTA determined by MicroED, where it has a *B* factor of 3.28 Å². The single coordinated water is hydrogen-bonded to Ser708, the C-terminus of a symmetry-related strand and the backbone of Thr709 in the mating sheet (Fig. 4, Supplementary Figs. S12 and S13, and Supplementary Table S4).

3.7. Structure of racemic GSTSTA

In crystals of racemic GSTSTA, homochiral strands stack to form single-enantiomer antiparallel β -sheets (Fig. 4b and

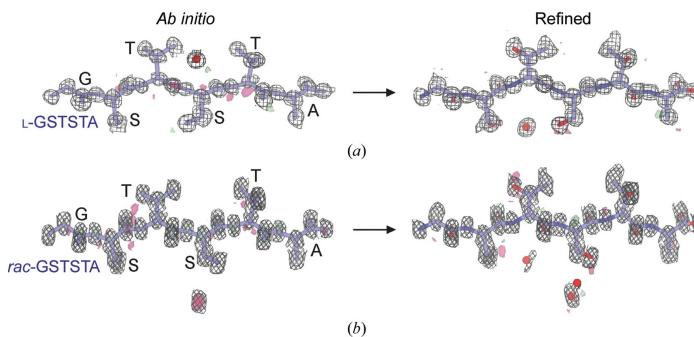


Figure 2
Ab initio structures and electrostatic potential maps of L-GSTSTA (a) and racemic GSTSTA (b). Each map in (a) is overlaid onto the initial atomic coordinates calculated by *SHELXD* from MicroED data. Each map in (b) is overlaid onto its corresponding refined model. The $2F_o - F_c$ map represented by the black mesh is contoured at 1.2σ . Green and red surfaces represent $F_o - F_c$ maps contoured at 3.0σ and -3.0 , respectively. Modeled waters are present as red spheres. The waters modeled in the *ab initio* solution in (a) and the refined structure in (b) are related by symmetry.

Supplementary Fig. S13). Like the homochiral L-GSTSTA sheets, the racemic GSTSTA sheets are buckled, with adjacent strands spaced 4.6 Å apart along each sheet (Fig. 4*b*). In the structure of racemic GSTSTA these sheets pack with alternating chirality, whereby each racemic GSTSTA protofibril is composed of one L-GSTSTA sheet and one D-GSTSTA sheet (Fig. 4*b*). The packing of D-GSTSTA sheets against their L-GSTSTA mates in the racemic fibril differs from that seen in the homochiral fibrils of L-GSTSTA. An alignment of the two protofibrils shows D-GSTSTA sheets displaced by approximately 5.3 Å compared with their corresponding L counterparts in the homochiral fibril (Fig. 4 and Supplementary Fig. S15). As a result of this displacement, the sheets are spaced farther apart (7–8 Å) in the racemic GSTSTA protofibril (Fig. 4 and Supplementary Fig. S12).

The longer spacing between sheets in the racemic GSTSTA protofibril is associated with bridging waters at its core (Supplementary Fig. S12). These waters make extensive contacts along the protofibril, with each hydrogen-bonding to at least one residue (Fig. 4, Supplementary Fig. S13 and Supplementary Table S4). Notably, the racemic GSTSTA structure shows a distinct rotamer for Ser710, which appears

bound to an ordered water, unlike its equivalent residue in the homochiral structure (Fig. 4, Supplementary S15 and Supplementary Table S4). One water (water 1; Supplementary Table S4) links Ser708 and Thr711 on the same D sheet while also coordinating Ser708 of the adjacent L sheet. This water is isolated from the other waters found within the structure. A small network of waters near the protofibril core links the carboxylate of one strand to Thr711 of a symmetry-related strand (Fig. 4, Supplementary Table S4). As in the structure of homochiral L-GSTSTA, the peptide atoms and bound waters in racemic GSTSTA show low *B* factors.

4. Discussion

Ice nucleation by *P. syringae* is linked to the expression of surface proteins, including InaZ (Wolber *et al.*, 1986). While full-length InaZ and InaZ fragments help to nucleate ice (Green & Warren, 1985; Kobashigawa *et al.*, 2005), individual InaZ repeats do not (Han *et al.*, 2017). However, at the high concentrations required for crystallization, GSTSTA repeats self-assemble into a protofibrillar structure of corrugated β -sheets (Supplementary Fig. S14). Similar structures are formed by both racemic GSTSTA and L-GSTSTA, and both contain ordered waters bridging tightly packed antiparallel β -sheets (Fig. 4, Supplementary Figs. S12 and S13). These waters may play a role in helping to stabilize the GSTSTA protofibril or could act as bridges or templates for solvent ordering at low temperatures. While we have no evidence to suggest that GSTXT(A/S) repeats facilitate the formation of amyloid-like InaZ protofibrils, our structures of GSTSTA present an opportunity to analyze the interactions between polar residues in InaZ repeats and ordered solvent molecules at atomic resolution.

The structures of enantiomerically pure and racemic GSTSTA present a platform for the comparison of homochiral and racemic amyloid protofibrils (Supplementary Fig. S16). To evaluate the packing of each GSTSTA protofibril, we look at the categorization of strand packing in amyloid fibrils through homosteric zipper classes, which were first proposed by Sawaya *et al.* (2007) and later by Stroud (2013). Many of these classes have been experimentally observed in amyloid crystals (Nelson *et al.*, 2005; Sawaya *et al.*, 2007). Homochiral GSTSTA forms a class 8 zipper in which two in-register, antiparallel β -sheets meet, related by a 180° rotation normal to the protofibril growth axis (Sawaya *et al.*, 2007; Stroud, 2013). The racemic GSTSTA structure resembles a class 8 zipper but is distinct in that two sheets of opposite handedness come together to form the protofibril (Supplementary Fig. S16). Because of this similarity to a class 8 zipper, we label this arrangement class 8 bar (Supplementary Fig. S16).

The increased propensity for crystallization by racemic mixtures could be exploited to facilitate the growth of amyloid crystals. The symmetry present in racemic amyloid crystals would have to accommodate the packing of homochiral protofibrils into the racemic structure or allow the formation of racemic protofibrils (Yeates & Kent, 2012), as is the case with GSTSTA. Our experiments in high-throughput crystallo-

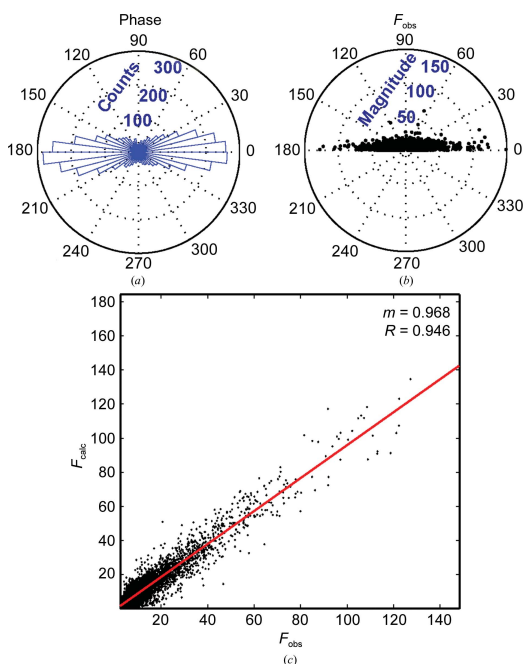


Figure 3
(a) The calculated phase associated with each reflection in the *P1* refinement of racemic GSTSTA data obtained by MicroED was analyzed and plotted as a histogram along the unit circle. (b) The magnitude of each reflection is plotted as a function of the absolute value of its associated phase. (c) A plot of F_o versus F_c values for each reflection in this data set shows a distribution that can be fitted by linear regression, shown as a red line, with slope $m = 0.97$ and *R* value 0.95.

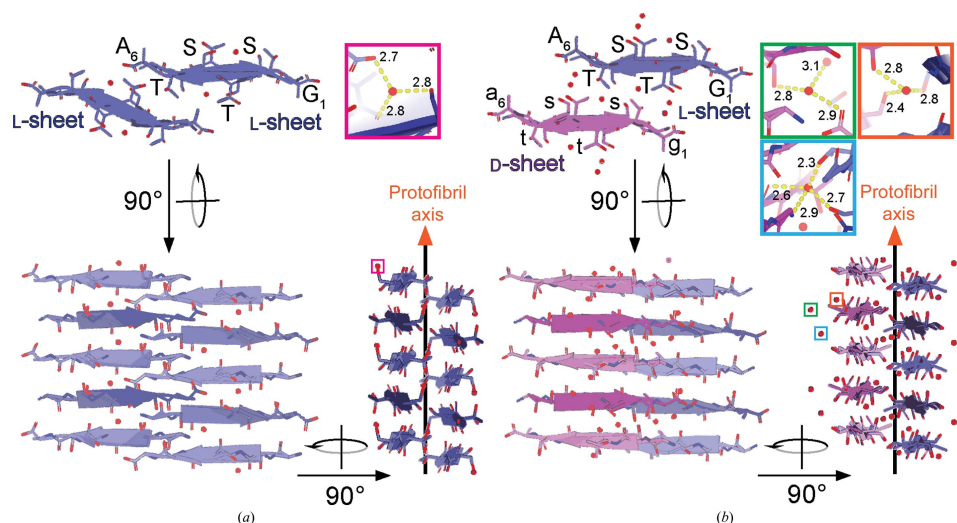


Figure 4

Views of protofibrils of L-GSTSTA (a) and racemic GSTSTA (b) represented by a pair of sheets with a view down the protofibril axis; both structures were derived by MicroED. A 90° rotation shows a side view of the protofibril with strands stacked along each sheet in an antiparallel fashion. Another 90° rotation shows a side view of the protofibril along the strand axis, showing a buckling of each sheet owing to the tilting of strands away from or towards the protofibril axis. Chains are colored such that blue represents L-peptides while magenta represents D-peptides. Lighter and darker shades of each color differentiate the orientations of strands within a sheet. Ordered waters found in each asymmetric unit are indicated by colored squares that correspond to the insets of matching colors. The insets show magnified views of each water molecule, with hydrogen bonds represented by yellow dashed lines and labeled with their corresponding distances in Å.

graphic trials of GSTSTA confirm the expected higher propensity for crystallization of racemic mixtures (Yeates & Kent, 2012; Supplementary Fig. S3), yielding a high number of conditions that contain submicrometre-sized crystals suitable for MicroED. The facile determination of *ab initio* structures from these crystals demonstrates how MicroED combined with solid-phase polypeptide synthesis (Dawson *et al.*, 1994; Merrifield, 1986) can expand the reach of racemic crystallography to submicrometre-sized crystals.

Acknowledgements

This project was inspired by conversations on racemic crystallography with Todd Yeates (UCLA). We thank Dan Shi (HHMI, Janelia Research Campus), David Boyer and Daniel Anderson (UCLA) for their assistance in data collection, Johan Hattne (UCLA) for helpful discussions and Janak Dadhaniya for assistance with figures.

Funding information

The following funding is acknowledged: US Department of Energy (grant No. DE-FC02-02ER63421); Howard Hughes Medical Institute – Janelia Research Visitor Program; National Institutes of Health, National Institute of General Medical Sciences (grant No. P41 GM103403); National Science Foundation, Office of Integrative Activities (grant No. DMR 1548924); Arnold and Mabel Beckman Foundation (award to Jose A. Rodriguez); Searle Scholars Program

(award to Jose A. Rodriguez); Pew Charitable Trusts (award to Jose A. Rodriguez); QCB Collaboratory (award to Marcus Gallagher-Jones); National Institutes of Health, Division of Biomedical Research Workforce (award No. GM007185 to Calina Glynn); Howard Hughes Medical Institute (award to Tamir Gonen).

References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213–221.
- Avital-Shmilovici, M., Mandal, K., Gates, Z. P., Phillips, N. B., Weiss, M. A. & Kent, S. B. H. (2013). *J. Am. Chem. Soc.* **135**, 3173–3185.
- Cochet, N. & Widehem, P. (2000). *Appl. Microbiol. Biotechnol.* **54**, 153–161.
- Davies, P. L. (2014). *Trends Biochem. Sci.* **39**, 548–555.
- Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. (1994). *Science*, **266**, 776–779.
- Eisenberg, D. & Jucker, M. (2012). *Cell*, **148**, 1188–1203.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., Crowther, R. A., Ghetti, B., Goedert, M. & Scheres, S. H. W. (2017). *Nature (London)*, **547**, 185–190.
- 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. (2018). *Nat. Struct. Mol. Biol.* **25**, 131–134.

- Garnham, C. P., Campbell, R. L. & Davies, P. L. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 7363–7367.
- Garnham, C. P., Campbell, R. L., Walker, V. K. & Davies, P. L. (2011). *BMC Struct. Biol.* **11**, 36.
- Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 3487–3492.
- Graether, S. P. & Jia, Z. (2001). *Biophys. J.* **80**, 1169–1173.
- Green, R. L. & Warren, G. J. (1985). *Nature (London)*, **317**, 645–648.
- Han, Y. J., Song, H., Lee, C. W., Ly, N. H., Joo, S.-W., Lee, J. H., Kim, S.-J. & Park, S. (2017). *Int. J. Biol. Macromol.* **94**, 634–641.
- Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., de la Cruz, M. J., Leslie, A. G. W. & Gonen, T. (2015). *Acta Cryst. A71*, 353–360.
- Hauptman, H. (1986). *Science*, **233**, 178–183.
- Hauptman, H. A. (2001). *Twentieth Century Harmonic Analysis – A Celebration*, edited by J. S. Byrnes, pp. 163–171. Dordrecht: Springer.
- Hughes, M. P., Sawaya, M. R., Boyer, D. R., Goldschmidt, L., Rodriguez, J. A., Cascio, D., Chong, L., Gonen, T. & Eisenberg, D. S. (2018). *Science*, **359**, 698–701.
- Kabsch, W. (2010). *Acta Cryst. D66*, 125–132.
- Kobashigawa, Y., Nishimiya, Y., Miura, K., Ohgiya, S., Miura, A. & Tsuda, S. (2005). *FEBS Lett.* **579**, 1493–1497.
- 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., Westermarck, G. T., Gonen, T. & Eisenberg, D. S. (2017). *Elife*, **6**, e19273.
- Lawrence, M. C. & Colman, P. M. (1993). *J. Mol. Biol.* **234**, 946–950.
- Mandal, K., Pentelute, B. L., Tereshko, V., Thammavongsa, V., Schneewind, O., Kossiakoff, A. A. & Kent, S. B. H. (2009). *Protein Sci.* **18**, 1146–1154.
- Mandal, K., Uppalapati, M., Ault-Riché, D., Kenney, J., Lowitz, J., Sidhu, S. S. & Kent, S. B. H. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 14779–14784.
- Matthews, B. W. (2009). *Protein Sci.* **18**, 1135–1138.
- Maury, C. P. J. (2009). *J. Intern. Med.* **265**, 329–334.
- McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
- Merrifield, B. (1986). *Science*, **232**, 341–347.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riekel, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.
- Pandey, R., Usui, K., Livingstone, R. A., Fischer, S. A., Pfaendtner, J., Backus, E. H. G., Nagata, Y., Fröhlich-Nowoisky, J., Schmäser, L., Mauri, S., Scheel, J. F., Knopf, D. A., Pöschl, U., Bonn, M. & Weidner, T. (2016). *Sci. Adv.* **2**, e1501630.
- Patterson, W. R., Anderson, D. H., DeGrado, W. F., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1410–1422.
- Pentelute, B. L., Gates, Z. P., Tereshko, V., Dashnau, J. L., Vanderkooi, J. M., Kossiakoff, A. A. & Kent, S. B. H. (2008). *J. Am. Chem. Soc.* **130**, 9695–9701.
- Rodriguez, J. A., Eisenberg, D. S. & Gonen, T. (2017). *Curr. Opin. Struct. Biol.* **46**, 79–86.
- 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. (2015). *Nature (London)*, **525**, 486–490.
- Sawaya, M. R., Rodriguez, J., Cascio, D., Collazo, M. J., Shi, D., Reyes, F. E., Hattne, J., Gonen, T. & Eisenberg, D. S. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 11232–11236.
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. Ø., Riekel, C. & Eisenberg, D. (2007). *Nature (London)*, **447**, 453–457.
- Sheldrick, G. M. (2008). *Acta Cryst. A64*, 112–122.
- Sheldrick, G. M. (2015). *Acta Cryst. C71*, 3–8.
- Shi, D., Nannenga, B. L., de la Cruz, M. J., Liu, J., Sawtelle, S., Calero, G., Reyes, F. E., Hattne, J. & Gonen, T. (2016). *Nat. Protoc.* **11**, 895–904.
- Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *Elife*, **2**, e01345.
- Stroud, J. C. (2013). *Acta Cryst. D69*, 540–545.
- Tayeb-Fligelman, E., Tabachnikov, O., Moshe, A., Goldshmidt-Tran, O., Sawaya, M. R., Coquelle, N., Colletier, J.-P. & Landau, M. (2017). *Science*, **355**, 831–833.
- Warren, G., Corotto, L. & Wolber, P. (1986). *Nucleic Acids Res.* **14**, 8047–8060.
- Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R. & Meier, B. H. (2008). *Science*, **319**, 1523–1526.
- Wolber, P. K., Deininger, C. A., Southworth, M. W., Vandekerckhove, J., van Montagu, M. & Warren, G. J. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 7256–7260.
- Wukovitz, S. W. & Yeates, T. O. (1995). *Nat. Struct. Mol. Biol.* **2**, 1062–1067.
- Yeates, T. O. & Kent, S. B. H. (2012). *Annu. Rev. Biophys.* **41**, 41–61.

IUCrJ

Volume 6 (2019)

Supporting information for article:

**Homochiral and racemic MicroED structures of a peptide repeat
from the ice-nucleation protein InaZ**

**Chih-Te Zee, Calina Glynn, Marcus Gallagher-Jones, Jennifer Miao, Carlos G.
Santiago, Duilio Cascio, Tamir Gonen, Michael R. Sawaya and Jose A.
Rodriguez**

Table S1 Crystallographic data collection and refinement statistics.

Values in parentheses are for the highest-resolution shell. All modelled waters have an occupancy of 1 after refinement.

Crystal	L-GSTSTA			Racemic GSTSTA								
Data collection												
Beam Type	x-ray		electron		x-ray		electron					
Space group	P2 ₁ 2 ₁ 2 ₁		P2 ₁ 2 ₁ 2 ₁		P2 ₁ /C		P2 ₁ /C					
Cell dimensions												
a, b, c (Å)	9.21	11.98	22.80	9.19	11.89	22.43	15.23	9.29	21.06	14.03	9.22	20.77
α, β, γ (°)	90.0	90.0	90.0	90.0	90.0	90.0	90.0	108.2	90.0	90.0	104.5	90.0
Resolution limit (Å)	1.1	(1.14–1.1)		0.90	(0.93–0.90)		1.1	(1.14–1.1)		0.9	(0.93–0.90)	
Wavelength (Å)	0.979		0.0251		0.979		0.0251					
No. of crystals merged	1		3		1		2					
R _{merge}	0.125 (0.241)		0.1618 (0.3651)		0.076 (0.163)		0.128 (0.352)					
R _{meas}	0.136 (0.265)		0.185 (0.422)		0.094 (0.206)		0.151 (0.411)					
R _{pim}	0.053 (0.108)		0.084 (0.200)		0.053 (0.123)		0.078 (0.207)					
I/ σ_1	10.6 (6.7)		5.52 (2.60)		6.99 (2.98)		5.46 (2.47)					
CC _{1/2}	0.99 (0.99)		0.99 (0.74)		0.99 (0.99)		0.99 (0.96)					
Completeness (%)	91.7 (72.7)		86.4 (80.0)		93.7 (82.4)		77.4 (78.4)					
No. reflections	7054 (516)		7494 (624)		5980 (477)		9884 (963)					
No. unique reflections	1079 (80)		1750 (164)		2097 (191)		2906 (304)					
Multiplicity	6.5 (6.5)		4.3 (3.8)		2.9 (2.5)		3.4 (3.2)					
Refinement												
Resolution range (Å)	11.4–1.1	(1.14–1.1)		7.1–0.90	(0.94–0.9)		14.4–1.1	(1.14–1.1)		7.6–0.90	(0.94–0.9)	
No. of Reflections (work)	968 (72)		1573 (16)		1873 (168)		2607 (274)					
R-work	0.061 (0.078)		0.217 (0.305)		0.199 (0.342)		0.233 (0.311)					
R-free	0.069 (0.087)		0.232 (0.287)		0.237 (0.311)		0.252 (0.421)					
CC(work)	0.997 (0.995)		0.940 (0.786)		0.989 (0.897)		0.953 (0.862)					
CC(free)	0.987 (1.000)		0.950 (0.744)		0.989 (0.725)		0.953 (0.706)					
No. of hydrogen atoms	31		31		31		31					
No. of non-hydrogen atoms	37		37		40		39					
Peptide	36		36		36		36					
Water	1		1		4		3					
B-factors (Å ²)												
Peptide	1.77		0.92		9.17		2.40					
Water	4.78		3.28		15.1		20.81					

Rms deviations

RMS(bonds, Å)	0.007	0.009	0.008	0.008
RMS(angles, °)	1.55	1.06	0.76	1.12

Table S2 Data reduction statistics for homochiral and racemic GSTSTA crystals for microfocal x-ray diffraction and MicroED.

Resolution	Number of Reflections			R-Factor			I/σ	Rmeas	CC(1/2)	
	Limit	Observed	Unique Possible	Completeness	Observed	Expected				Compared
Racemic GSTSTA										
MicroED										
1.79	1149	357	471	75.8%	8.4%	9.1%	1146	9.94	10.0%	99.1*
1.43	1230	363	470	77.2%	13.4%	12.3%	1226	7.67	15.8%	98.1*
1.25	1248	364	473	77.0%	15.0%	15.6%	1242	6.07	17.6%	98.9*
1.13	1238	358	459	78.0%	14.5%	15.4%	1233	5.41	17.1%	99.2*
1.05	1252	360	462	77.9%	16.9%	18.5%	1242	5.20	19.7%	98.1*
0.99	1271	360	466	77.3%	24.0%	26.3%	1262	3.99	27.9%	98.2*
0.94	1288	366	458	79.9%	35.4%	40.1%	1283	3.12	41.5%	90.3*
0.90	1208	378	487	77.6%	36.3%	44.3%	1149	2.45	42.6%	94.3*
total	9884	2906	3746	77.6%	12.8%	13.6%	9783	5.46	15.1%	99.1*
Microfocal x-ray										
1.79	1484	506	514	98.4%	6.7%	7.3%	1442	12.25	8.2%	99.5*
1.43	1364	481	508	94.7%	9.0%	9.3%	1315	7.94	11.1%	99.3*
1.25	1355	469	502	93.4%	13.1%	13.3%	1309	5.15	15.8%	99.3*
1.13	1459	509	541	94.1%	15.2%	16.0%	1415	3.66	18.7%	99.5*
1.10	318	132	164	80.5%	19.8%	18.7%	292	2.70	25.2%	99.4*
total	5980	2097	2229	94.1%	7.6%	8.2%	5773	6.99	9.4%	99.6*
L-GSTSTA										
MicroED										
1.80	902	237	298	79.5%	9.8%	12.2%	896	9.12	11.3%	99.2*
1.43	972	229	261	87.7%	16.4%	15.8%	965	7.25	18.6%	96.7*
1.25	974	223	254	87.8%	20.2%	19.3%	965	6.32	23.1%	94.4*
1.14	952	219	251	87.3%	21.0%	21.6%	949	5.46	23.8%	95.5*
1.05	992	218	245	89.0%	22.6%	23.7%	981	4.96	25.4%	96.1*
0.99	944	211	243	86.8%	22.8%	27.0%	938	4.25	25.8%	97.2*
0.94	972	216	241	89.6%	35.4%	39.1%	962	3.42	40.3%	70.2*
0.90	800	205	250	82.0%	37.8%	45.7%	775	2.61	43.5%	71.0*
total	7508	1758	2043	86.0%	16.2%	17.9%	7431	5.50	18.5%	99.2*
Microfocal x-ray										
1.80	1768	290	304	95.4%	8.7%	9.7%	1757	14.25	9.7%	98.8*
1.43	1681	254	270	94.1%	14.5%	14.3%	1677	10.57	15.7%	98.8*
1.25	1666	246	263	93.5%	18.7%	17.7%	1662	9.15	20.3%	98.0*
1.134	1423	220	239	92.1%	21.1%	19.4%	1413	7.92	23.0%	97.2*
1.10	540	83	116	71.6%	23.7%	23.5%	539	6.86	26.1%	98.4*
total	7078	1093	1192	91.7%	12.5%	12.7%	7048	10.41	13.6%	98.8*

Table S3 Hydrogen bonding between adjacent strands in homochiral L-GSTSTA sheets.

Peptide donor and acceptors are denoted as: chain name followed by a three-digit residue number and the associated three letter code. This table includes only unique hydrogen bonds between strands in a single beta sheet along the protofibril axis. All distances are measured in Ångströms, angles in degrees.

Donor	Atom	Acceptor	Atom	DA dist.	Type	CA to	DHA	H-A Dist.	H-A-AA	D-A-AA
B707-Gly	N	A712-Ala	O	2.80	MM	5.1	165.5	1.80	139.8	144.7
A712-Ala	N	B707-Gly	O	2.95	MM	5.1	153.1	2.03	140.6	142.1
B709-Thr	N	A710-Ser	O	2.97	MM	5.2	154.6	2.04	139.6	146.2
A710-Ser	N	B709-Thr	O	2.82	MM	5.2	151.3	1.90	147.2	157.0
A709-Thr	OG1	B710-Ser	OG	3.40	SS	4.1	153.7	2.48	94.0	99.2
B711-Thr	N	A708-Ser	O	2.89	MM	5.0	144.8	2.01	128.8	139.2
A708-Ser	N	B711-Thr	O	2.92	MM	5.0	154.7	1.98	137.6	138.0
B711-Thr	OG1	A708-Ser	OG	2.68	SS	5.0	165.0	1.70	130.4	125.4

Table S4 Hydrogen bonds of modelled waters in homochiral and racemic GSTSTA structures.

Peptide donor and acceptor names are abbreviated as the chain name followed by a three-digit residue number, a dash, and the associated three letter code. All hydrogen bonds in the table include a water as a donor or acceptor and are restricted to waters found in one asymmetric unit. All distances are measured in ångströms, angles in degrees. Only hydrogen bonds with distances below 3.2 Å are listed. The list includes potential hydrogen bonding partners, though not all might be satisfied at a time for a given atom.

Donor Name	Atom Type	Acceptor Name	Atom Type	DA dist.	D-A-AA angle
L- GSTSTA, X-Ray					
Water 1	O	A709-Thr	O	2.84	113.9
Water 1	O	C712-Ala	O	2.78	109.1
B708-Ser	OG1	Water 1	O	2.72	
Racemic GSTSTA, X-Ray					
Water 1	O	A709-Thr	O	2.93	121.1
B708-Ser	OG1	Water 1	O	2.82	
B710-Ser	OG1	Water 1	O	2.85	
Water 2	O	G712-D-Ala	O	2.89	107.2
Water 2	O	Water 3	O	2.71	
E711-D-Thr	OG1	Water 2	O	2.93	
Water 3	O	Water 4	O	2.94	
Water 3	O	E712-D-Ala	O	2.86	102.6
H707-Gly	N	Water 3	O	2.80	
C709-Thr	OG1	Water 3	O	2.67	
Water 4	O	C708-Ser	O	2.78	122.3
Water 4	O	B711-Thr	O	2.89	127.8
H707-Gly	N	Water 4	O	2.90	
C709-Thr	OG1	Water 4	O	3.02	
L-GSTSTA, MicroED					
Water 1	O	D712-Ala	O	2.66	112.7
Water 1	O	B709-Thr	O	2.76	113.3
A708-Ser	OG1	Water 1	O	2.76	
Racemic GSTSTA, MicroED					
Water 1	O	E711-D-Thr	OG1	2.81	126.9
D708-D-Ser	OG1	Water 1	O	2.42	
B708-Ser	OG1	Water 1	O	2.76	
Water 2	O	Water 3	O	3.05	
Water 2	O	G712-D-Ala	O	2.87	107.4
E711-D-Thr	OG1	Water 2	O	2.76	
Water 3	O	H712-D-Ala	O	2.62	108.1
H707-Gly	N	Water 3	O	2.93	
C709-Thr	OG1	Water 3	O	2.34	
B710-Ser	OG1	Water 3	O	2.71	

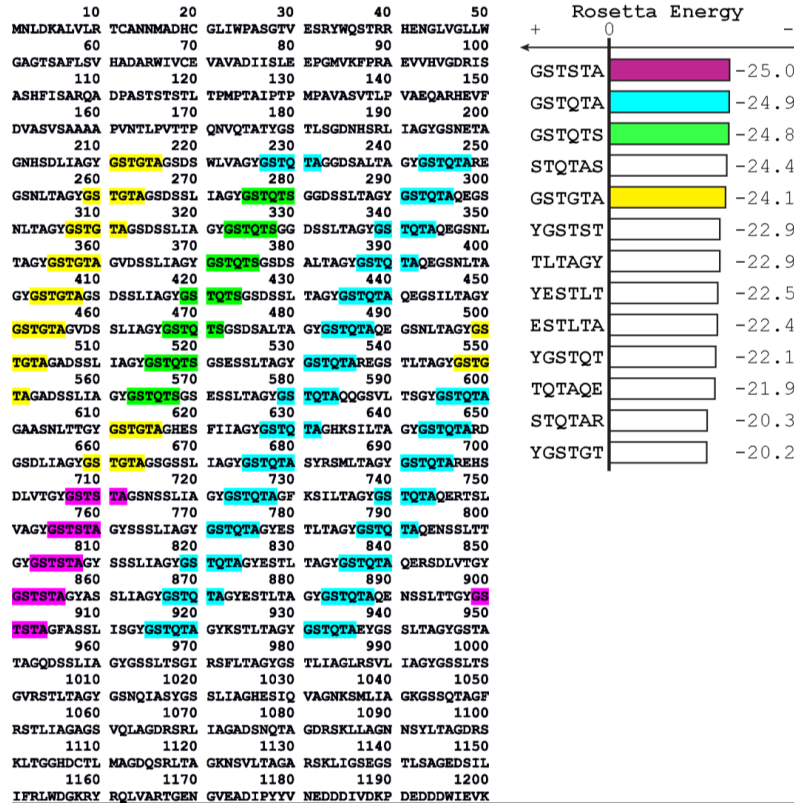


Figure S1 Sequence of the ice nucleation protein InaZ is shown with its degenerate hexameric repeats highlighted as follows: GSTGTA (yellow), GSTQTA (cyan), GSTSTA (magenta), and GSTQTS (green). The propensity for the hexamers to form steric zippers is shown on the left as Rosetta energy scores, determined by ZipperDB (Goldschmidt *et al.*, 2010). This list of repeats is limited to those with Rosetta energy lower than -20 that containing at least two threonine residues and appear with frequency greater than or equal to five across the InaZ sequence.

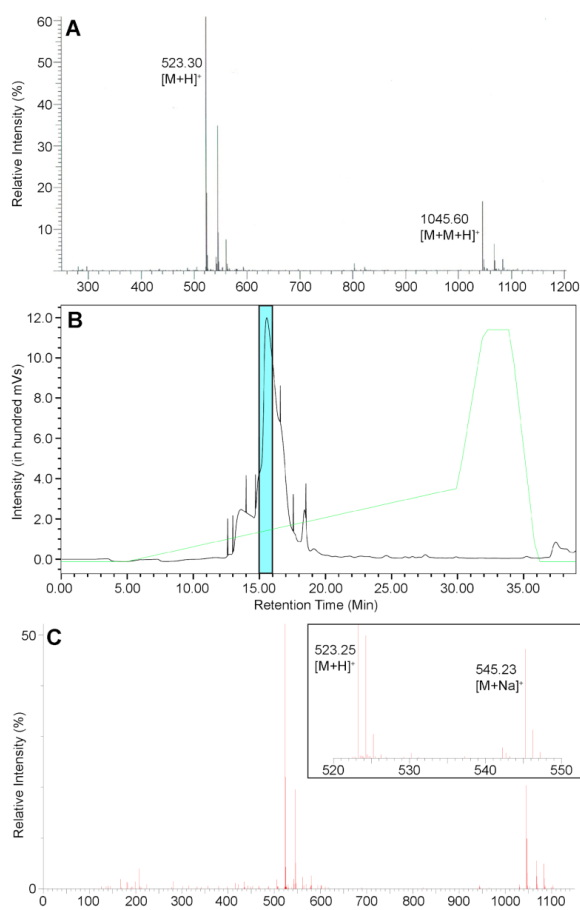


Figure S2 The mass trace of L-GSTSTA (A), purchased from Genscript, shows a [M+H]⁺ peak of 523.30 g/mol (expected 523.22) and a dimer [M+M+H]⁺ peak of 1045.6 g/mol (expected 1045.44). D-GSTSTA was synthesized and purified in-house by reverse-phase HPLC (B). The cyan shaded region highlights the HPLC fraction collected and lyophilized for crystallization experiments. The mass spectrum of D-GSTSTA (C) shows a [M+H]⁺ peak of 523.24 g/mol (expected 523.22) and a [M+Na]⁺ peak of 545.23 g/mol (expected 545.22).

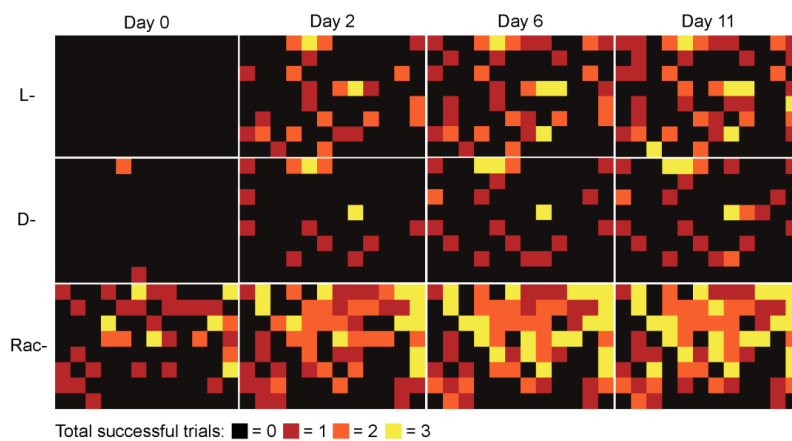


Figure S3 The number of successful crystallization trials in the 96-well crystal screens for L-GSTSTA, D-GSTSTA, and racemic GSTSTA were monitored and converted to heatmaps. Each well contained three hanging drops, each with different protein to buffer ratios. A count of crystals found in each of these three conditions was given a score from 0 to 3. No crystals in drops (black), one drop with crystals (red), two drops with crystals (orange), crystals in all three drops (yellow). The initial time point was three hours after setting up the screens (day 0), with data points collected up to 11 days post-setup.

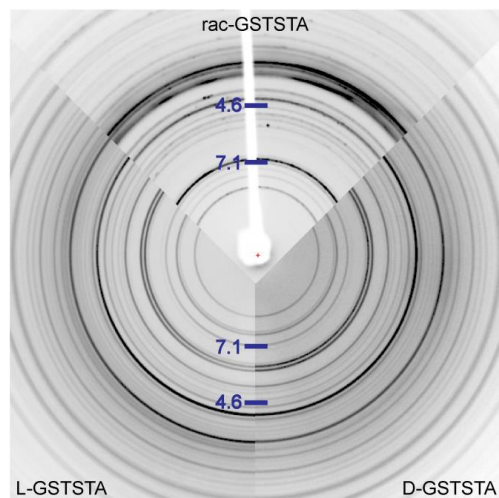


Figure S4 Comparison of powder diffraction patterns measured from L-GSTSTA (left), D-GSTSTA (right), and racemic GSTSTA (top) slurries show differences across all resolutions. Both patterns show faint rings at approximately 4.6 Å, representing the approximate distance between strands along the fibril axis. The racemic pattern contains a prominent reflection at ~7.1 Å, representative of overall sheet-to-sheet distances in its structure.

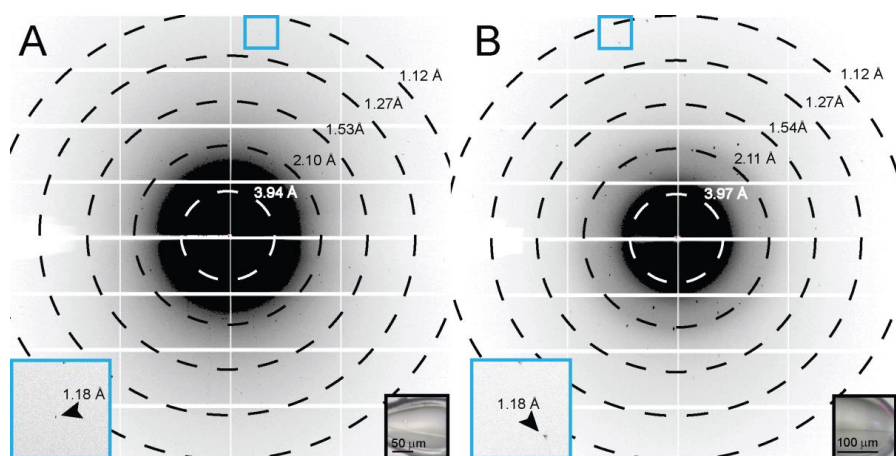


Figure S5 Single diffraction patterns of homochiral L-GSTSTA (A) or racemic GSTSTA (B) measured during continuous vector scanning, microfocus x-ray data collection. Each pattern corresponds to a 5° wedge (A) or 2.5° wedge (B) of reciprocal space. Black insets show in-line images of the crystals that were diffracted; blue squares correspond to magnified regions (blue insets) of the pattern that show diffraction near the detector edge at approximately 1.1 Å resolution (black arrows). Resolution circles are indicated by rings; scale bars are 50 to 100 μm for L-GSTSTA and racemic GSTSTA, respectively.

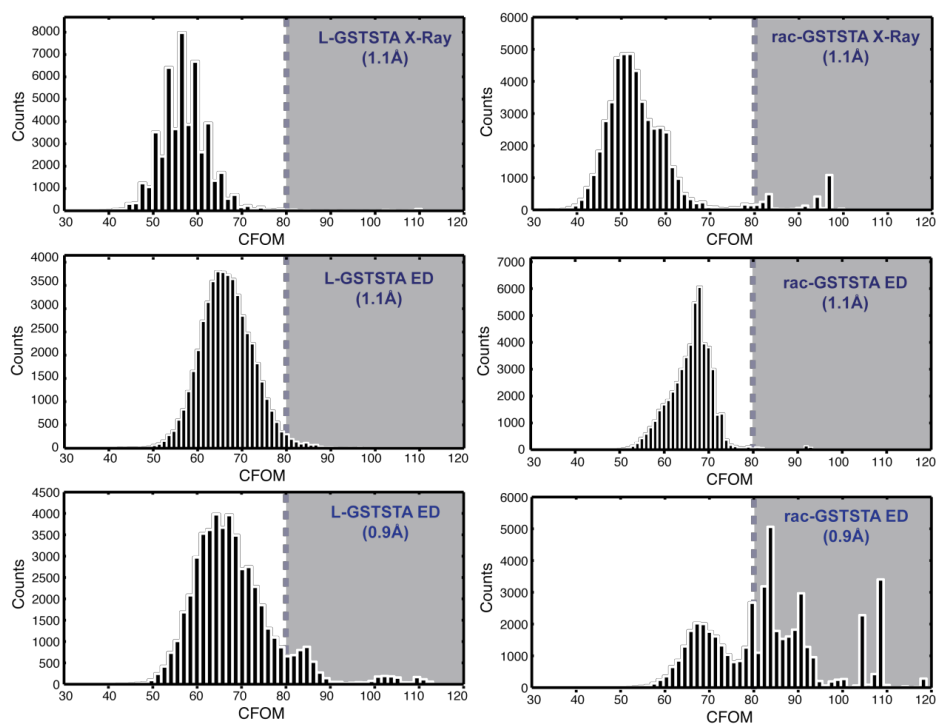


Figure S6 Histograms of the combined figure of merit (CFOM) scores for 50,000 trials by SHELXD indicate the approximate frequency of correct solutions. The shaded region in each plot, where CFOM scores are greater than 80, represents an area in which solutions have a high probability of being correct. Two sets of plots were generated from MicroED data: results of attempts using a truncated dataset that matches the resolution of the x-ray data (1.1Å) are shown in middle panels, while bottom panels show results of attempts using the measured resolution (0.9Å) for that data.

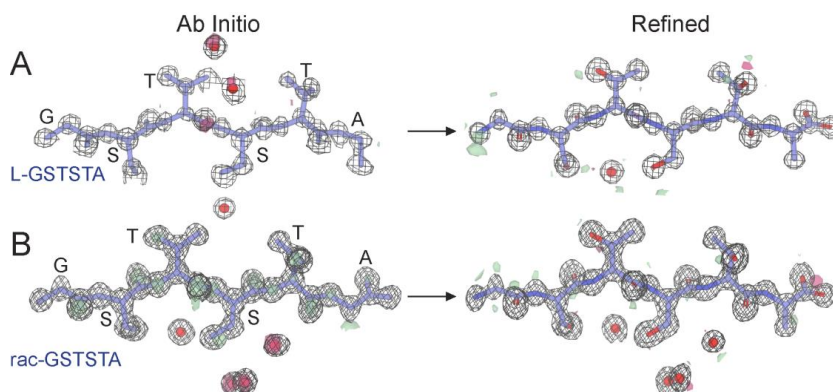


Figure S7 *Ab Initio* structures and electron density maps of L-GSTSTA (A) or racemic GSTSTA (B). Each map in A is overlaid onto the initial atomic coordinates calculated by SHELXD from x-ray diffraction data. Each map in B is overlaid onto its corresponding refined model. The 2F_o-F_c map represented by the black mesh is contoured at 1.2 σ . Green and red surfaces represent the F_o-F_c maps contoured at 3.0 and -3.0 σ . Modelled waters are present as red spheres.

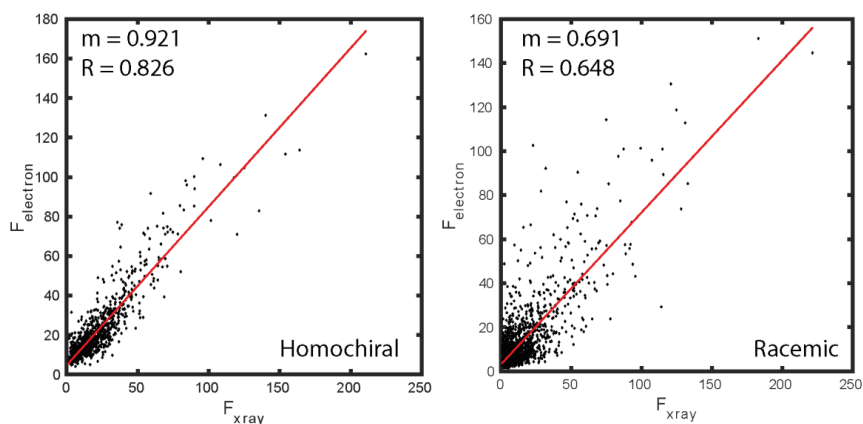


Figure S8 A plot of magnitudes (F) compared between reflections in datasets collected from homochiral crystals (left) or racemic crystals (right) by either electron or x-ray scattering shows a distribution that can be fit by linear regression, indicated by red lines with slope (m) and R-value.

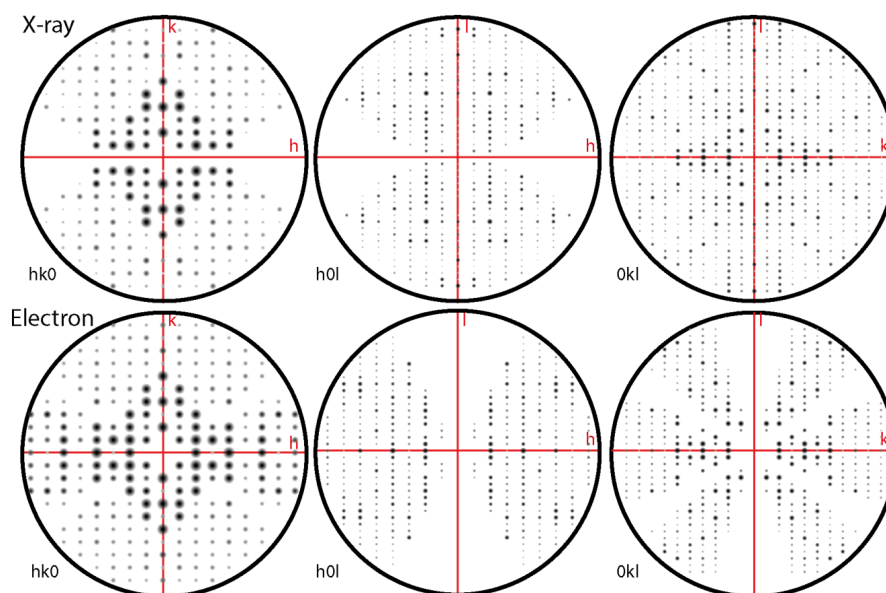


Figure S9 HKL Zone analysis for crystals of L-GSTSTA. Fourier magnitudes are shown as displayed by the HKL view software for reflections along principal zones of the reciprocal lattice. Zones where $l=0$ (left), $k=0$ (middle), $h=0$ (right) are shown for merged data collected by x-ray (top) and electron (bottom) diffraction. The black circle in each zone plot represents a resolution of 1.1Å ; zone axes are labeled in red.

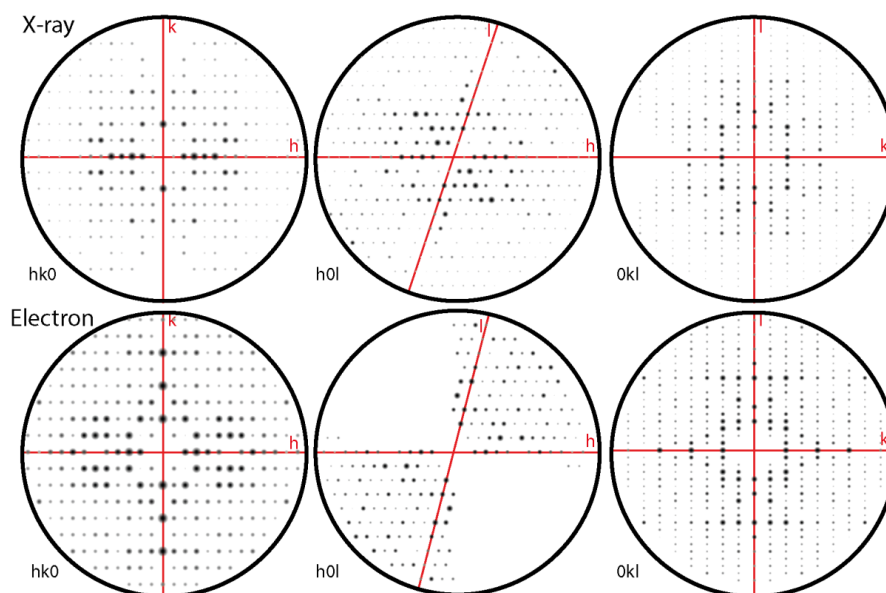


Figure S10 HKL Zone analysis for crystals of racemic GSTSTA. Fourier magnitudes are shown as displayed by the HKL view software for reflections along principal zones of the reciprocal lattice. Zones where $l=0$ (left), $k=0$ (middle), $h=0$ (right) are shown for merged data collected by x-ray (top) and electron (bottom) diffraction. The black circle in each zone plot represents a resolution of 1.1 \AA ; zone axes are labeled in red.

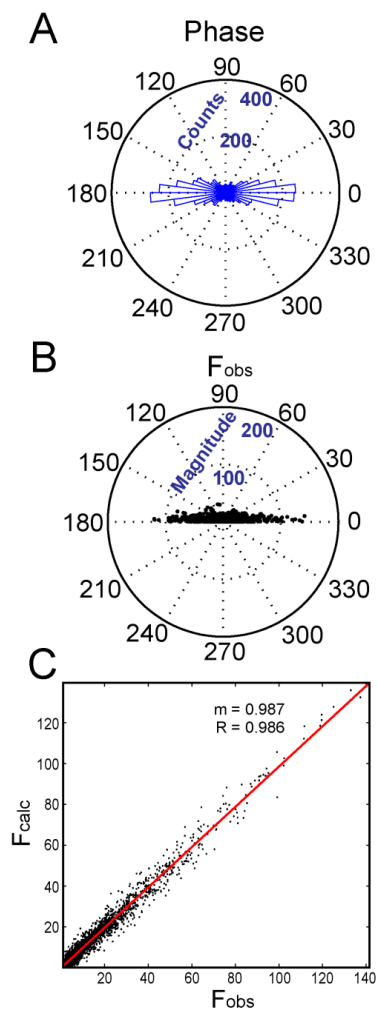


Figure S11 (A) The calculated phase associated with each reflection in the P1 refinement of racemic GSTSTA data obtained by x-ray diffraction was analyzed and plotted as a histogram along the unit circle. (B) The magnitude of each reflection is plotted as a function of the absolute value of its associated phase. (C) A plot of F_o vs. F_c values for each reflection in this data set shows a distribution that can be fit by linear regression, shown as a red line with slope $m=0.987$ and R-value 0.986.

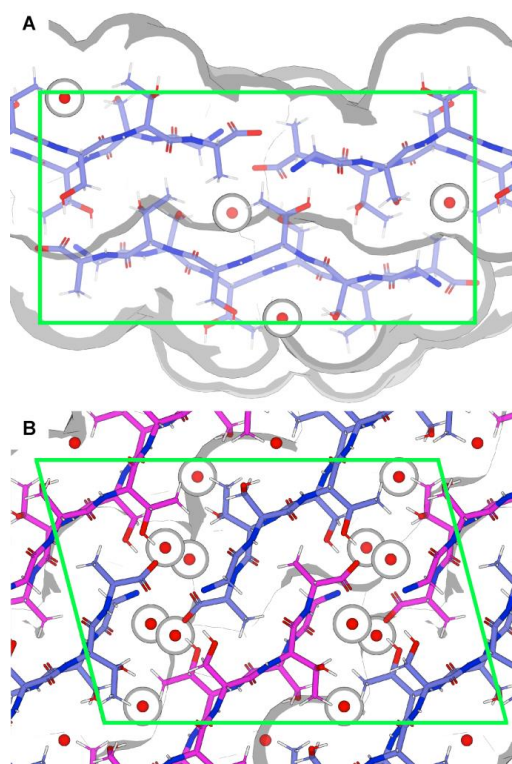


Figure S12 Views down the a-axis of L-GSTSTA (A) and racemic GSTSTA (B) structures are enclosed in green by their respective unit cells. L-GSTSTA strands are colored blue while D-GSTSTA strands are colored magenta. The volumes occupied by each structure are shown in white with edges defined by the shaded grey regions. Space-fill models represent the solvent accessible surface; ordered waters are represented by van der Waals radii of 1.4Å.

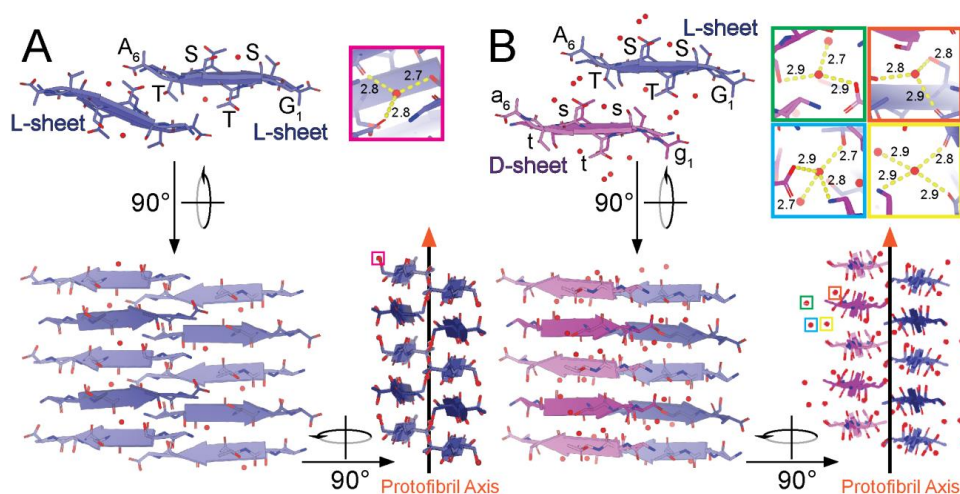


Figure S13 Views of protofibrils of L-GSTSTA (A) and racemic GSTSTA (B) represented by pair of sheets with a view down the protofibril axis; both structures derived by x-ray diffraction. A 90° rotation shows a side view of the protofibril with strands stacked along each sheet in an antiparallel fashion. Another 90° rotation shows a side view of the protofibril along the strand axis, showing a buckling of each sheet due to the tilting of strands away from or toward the protofibril axis. Chains colored such that blue represent L-peptides while magenta represents D-peptides. Lighter and darker shades of each color differentiate the orientation of strands within a sheet. Ordered waters found in each asymmetric unit are indicated by colored squares that correspond to insets of matching colors. Insets show magnified views of each water molecule with hydrogen bonds represented by the yellow dashed lines, labelled with their corresponding distances in Å

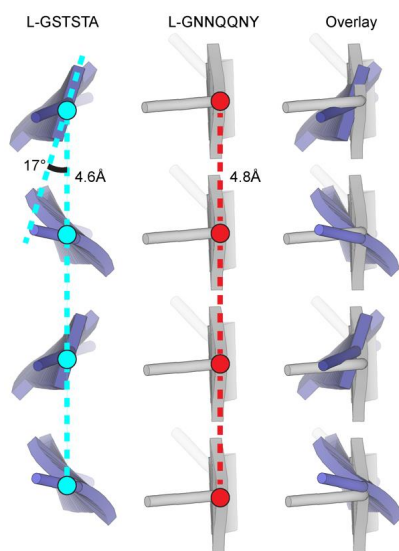


Figure S14 Beta sheets in the structure of L-GSTSTA have an inter-strand distance of 4.6 Å with amide carbonyls angled away from the protofibril axis by approximately 17° (left). Canonical beta sheets formed by the yeast prion segment L-GNNQQNY (RCSB PDB: 1YJP) show inter-strand distance of 4.8 Å and a near 0° deviation of amide hydrogen bonding down the protofibril axis (center) (Nelson *et al.*, 2005). An overlay (right) illustrates compaction of the L-GSTSTA sheet along its length compared to a sheet formed by L-GNNQQNY.

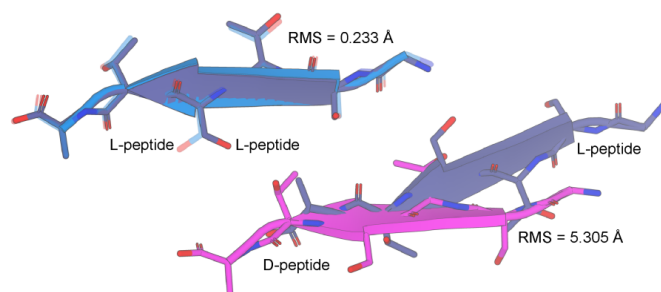


Figure S15 Pairs of mated strands representing the homochiral and racemic protofibrils of GSTSTA. Alignment of these protofibrils based on a common L-GSTSTA sheet shows a displacement of their paired sheet. The RMSD between the common L-GSTSTA sheets is 0.23 Å, while that between the pairing L and D sheets of the homochiral and racemic protofibrils is 5.3 Å. L-GSTSTA strands are colored blue and purple while D-GSTSTA is colored magenta.

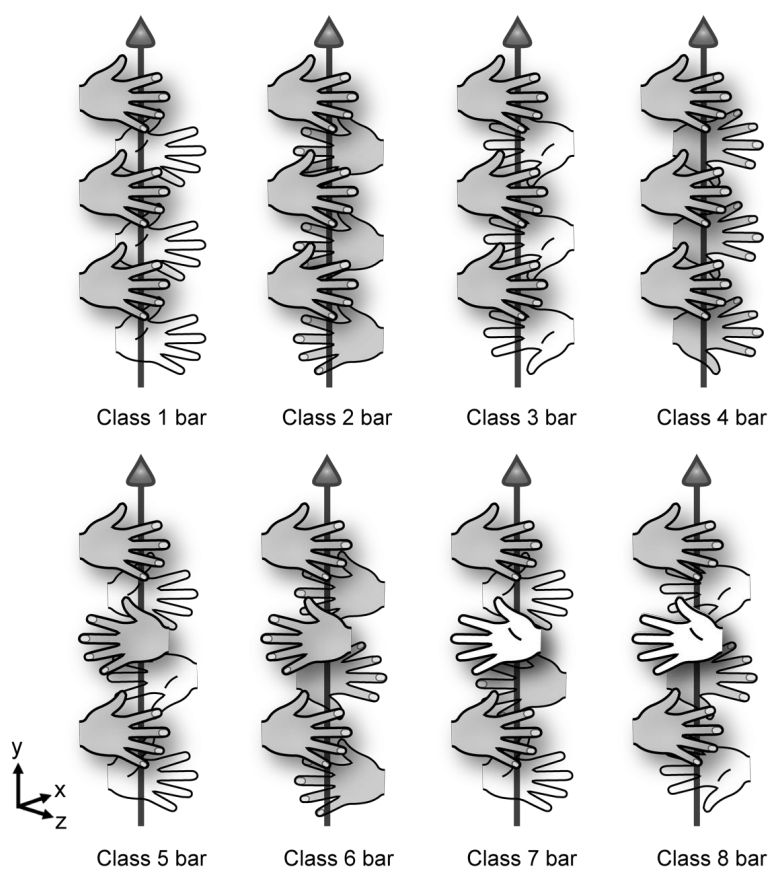


Figure S16 Eight new potential steric zipper classes are enabled by racemic assemblies. All new classes are based on those originally described in Sawaya *et al.* (Sawaya *et al.*, 2007) but now contain a mirror plane, a glide plane, or an inversion center. Strands are represented by left and right hands, each equivalent to the enantiomers present in a zipper class. The asymmetry of side chains on either side of a strand is portrayed by the palm and back of each hand. The up or down orientation of the thumbs and the direction in which the fingers point indicates the direction of each strand. An arrow indicates the axis of fibril growth, which here is coincident with the y direction. While additional symmetry classes with racemic mixtures are possible, only eight are illustrated here.

Appendix Chapter 4

Data-driven challenges and opportunities in crystallography

The work described in this chapter has been reproduced from:

Calina Glynn and Jose A. Rodriguez. "Data-driven challenges and opportunities in crystallography." *Emerging Topics in Life sciences*, 2019.

Copyright 2019

Calina Glynn and Jose A. Rodriguez

Review Article

Data-driven challenges and opportunities in crystallography

Calina Glynn and  Jose A. Rodriguez

Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, U.S.A

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



Structural biology is in the midst of a revolution fueled by faster and more powerful instruments capable of delivering orders of magnitude more data than their predecessors. This increased pace in data gathering introduces new experimental and computational challenges, frustrating real-time processing and interpretation of data and requiring long-term solutions for data archival and retrieval. This combination of challenges and opportunities is driving the exploration of new areas of structural biology, including studies of macromolecular dynamics and the investigation of molecular ensembles in search of a better understanding of conformational landscapes. The next generation of instruments promises to yield even greater data rates, requiring a concerted effort by institutions, centers and individuals to extract meaning from every bit and make data accessible to the community at large, facilitating data mining efforts by individuals or groups as analysis tools improve.

Introduction

The crystallographic analysis of macromolecules is increasingly dependent on data-driven approaches [1]. Growing data output is a central feature of the new tools that facilitate crystallographic analysis. These include ultra-bright ultra-fast X-ray sources, upgraded synchrotron facilities and accessible instruments for routine electron diffraction. Single experiments can now yield gigabytes or terabytes of data from one or more samples allowing unprecedented exploration of ultra-fast processes or from increasingly small crystallites [2–4]. Here, we discuss two areas of rapid growth in structural biology that are increasingly reliant on large-scale data gathering: serial X-ray crystallography [1,5,6] and electron crystallography of micro and nanocrystals [7–10]. While other structural biology approaches are without a doubt also reliant on large datasets [11], data demands in crystallography are rapidly growing. We focus on these two closely related crystallographic approaches and their evolution from conventional data gathering to large-scale data collection and analysis.

While the two crystallographic approaches we highlight are somewhat similar, some key differences are important to note. Most of the X-ray crystallography measurements we describe are performed in a small number of large, government-funded facilities with staff and computational infrastructure matched to the scale of the experiment. In contrast, electron diffraction efforts are generally driven by individual groups and are implemented across a large number of small-scale facilities. Nonetheless, data collected by each approach is largely sparse and must be reduced to a small set of measurements for structure determination. The focus of this review is, therefore, the challenge of data reduction and the extraction of information from large sets of individual diffraction patterns belonging to one or more crystals.

Conventional X-ray crystallography efforts set the stage for the analysis of this type of data by establishing methods for routine data collection and processing from macro-scale crystals. The success of these approaches has yielded a consistently high number of structures deposited in the protein data bank (PDB) per year, a number that remains unmatched by any other structural biology method (Figure 1). Likewise, efforts in electron cryo-microscopy (cryoEM) [11–19] have established analogous

Received: 1 May 2019
Revised: 13 June 2019
Accepted: 24 June 2019

Version of Record published:
26 July 2019

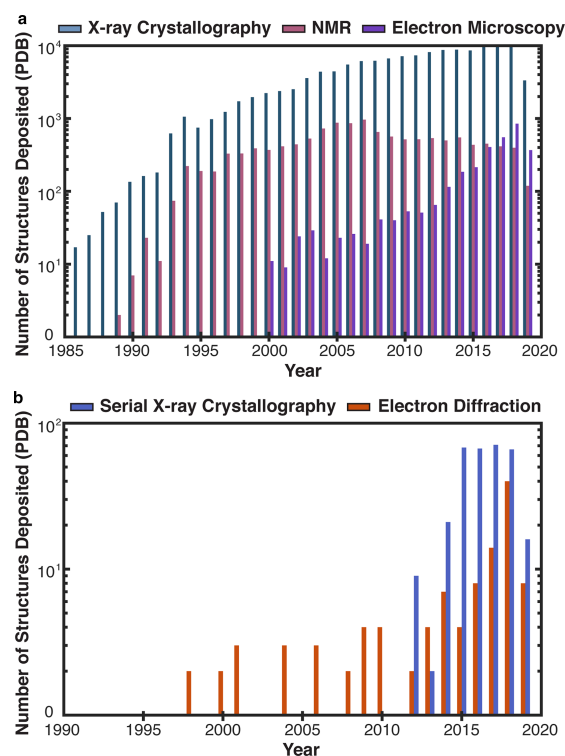


Figure 1. Number of structures deposited in the PDB per year [20].

(a) Structures deposited per year from 1985 to 2019 associated with all X-ray crystallographic methods, compared with nuclear magnetic resonance (NMR) and electron cryo-microscopy (cryoEM). Each counted structure is associated with X-ray crystallography (blue), NMR (pink) or electron microscopy (magenta). (b) A subset of structures deposited per year from 1990 to 2019 associated with serial X-ray crystallography or electron crystallography, including MicroED. Each counted structure is associated with data collection by serial X-ray crystallography methods, including those acquired at X-ray free electron laser facilities (purple), compared with those acquired by electron crystallography methods (orange).

pipelines for structure determination [11,19] that directly benefit electron crystallographic approaches [8], giving rise to a growing number of deposited structures per year (Figure 1) as catalogued by the protein data bank (PDB) [20].

The rise in structures determined by these new methods is accompanied by increasing challenges associated with storage, analysis and accessibility of data [20–23]. The Electron Microscopy Public Image Archive has been established to facilitate the public archival of electron microscopy and diffraction data [22]. Likewise, data repositories like the coherent X-ray imaging data bank are providing resources for depositing the terabytes of data collected at X-ray free electron laser facilities [23]. As the use of fast detectors grows routine in crystallography, the number of sites worldwide collecting large datasets will place additional demands on these and analogous resources.

Advances in X-ray diffraction

The advent of high brightness sources for crystallographic study of macromolecules took a quantum leap with the development and introduction of ultra-bright, pulsed X-ray lasers [6,24,25]. While initially developed with

the goal of determining structures of single molecules, acquiring a signal before inducing a coulombic explosion of each molecule [26,27], X-ray lasers have opened new doors for the crystallographic study of proteins and nucleic acids [1,6,28]. The first structure published by this approach was that of photosystem I, for which more than 3 million diffraction patterns were collected [2]. Among the first demonstrations of the feasibility of this approach was an effort to demonstrate the routine ability to determine structures of proteins from crystal slurries (Figure 2). This was demonstrated with the structure of lysozyme determined from a slurry of microcrystals [29], although many other sample delivery strategies have since been demonstrated [30] (Figure 2). This new kind of experiment, termed serial femtosecond crystallography (SFX) [2,6], allows the detection of high-resolution diffraction from individual microcrystals micrometers on edge. Lysozyme crystals were delivered to the 9.4 keV X-ray beam via a focused jet of liquid with a 10-micrometer focus, and out of 1.5 million individual diffraction frames collected, about 12 000 were indexed and integrated to yield a high-resolution structure [29]. Diffraction frames measured the interaction of crystals with individual 40 or 5 femtosecond pulses at a rate of 120 Hz, yielding a data stream of gigabytes per minute [29]. This approach has now been applied to a whole variety of samples, including entirely unknown structures, some of whose crystals are grown in cells [31–38]. A key feature of this data-driven crystallographic approach is the ability to probe dynamics, capturing the atomic scale motions within proteins [28,32,39,40].

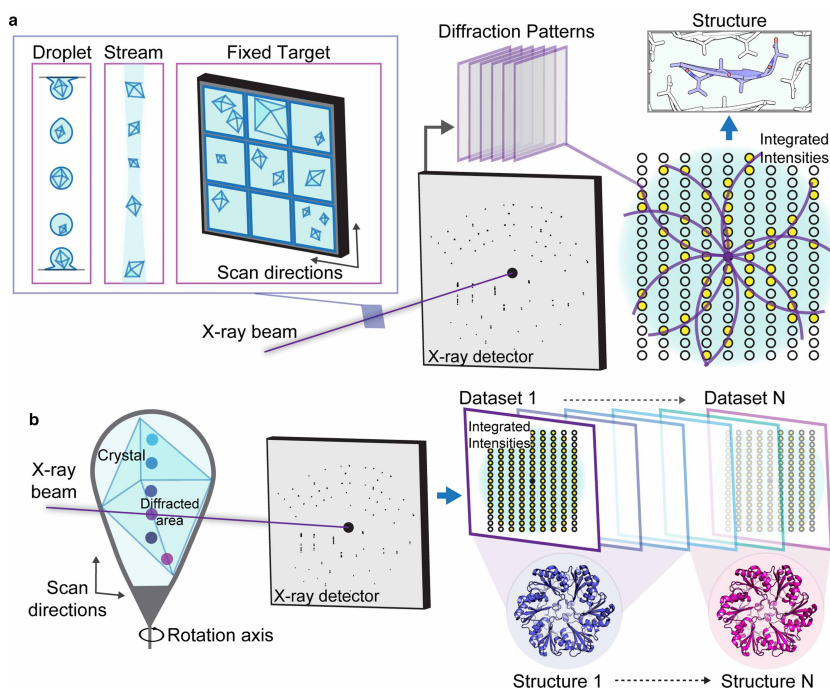


Figure 2. Examples of data gathering by serial X-ray crystallography.

(a) Ultra-fast pulses of high energy X-rays are directed at a sample interaction region to which samples can be delivered by varying means. Three methods of sample delivery are pictured: droplet-based delivery, stream or liquid-jet delivery, and fixed target delivery, typically by a windowed device. Data from a train of patterns recorded from different crystals are combined to yield one or more datasets that are used to determine a molecular structure. (b) Approaches that pursue multiple structures from a single crystal can train an X-ray beam on particular locations on a crystal, collecting multiple datasets that can then be combined or individually yield a structure. In this way, several structures can be obtained from a single crystal or a set of similar crystals. This example illustrates determination of the monoclinic structure of EutL, PDB ID 4TLH.

The latest generation of ultra-fast X-ray sources offers even greater throughput [41]. The megahertz X-ray free electron laser facility, part of the European XFEL project, delivers 10 pulse trains of megahertz X-ray sequences delivering up to 27 000 pulses per second [41]. However, efficiently utilizing a torrent of pulses presents a challenge for real-time processing and data storage. Diffraction produced by 50-pulse trains impinging on crystals in a liquid jet every 100 milliseconds was captured by a megapixel Adaptive Gain Integrating Pixel Detector [41,42]. A jet with flow rates of ~50 m/s allows the fresh sample to be probed by the series of pulses in each train and the recording of individual diffraction patterns within a train, with pulses in a train separated by ~900 ns [41].

Serial femtosecond crystallography inspired the adaptation of methods and technologies for serial data collection efforts at synchrotron beamlines [43–45]. There, millisecond timescale shutterless detection of the interaction between focused X-ray beams and a high viscosity crystal matrix or crystals held in fixed target mounts has facilitated routine room-temperature determination of structures from thousands of diffraction patterns within minutes to hours [43–46]. The large, fast detectors used in some of these experiments accelerate data collection [47]; 18-megapixel images collected at high frame rates (50 Hz) allow the collection of a single dataset in under 90 min [44]. Collection of room-temperature data at this rate in an accessible synchrotron beamline has enabled high-resolution structure determination [45,48], including determination of structures *de novo* and with bound ligands [44]. Acquisition frame rates of 50 Hz appear to improve the signal to noise, particularly at high resolution by allowing the collection of multiple diffraction patterns per crystal and the elimination of background from empty frames [44]. This was particularly advantageous for *de novo* structure determination through the improved measurement of the anomalous signal [44]. Importantly, information about ligand binding could be acquired from as few as 900 images (9 Gb) collected over a period of 5 min [44]. This opens the door to large-scale screening efforts for ligand binding at dedicated serial crystallography beamlines. This may be even further enabled by the development of compact X-ray sources that deliver high-repetition bright X-ray beams in a facility with the footprint of a conventional laboratory [49].

A different type of serial crystallography experiment uses bright X-ray sources to expand the amount of data collected from single crystals and extract more information out of each (Figure 2) [48,50,51]. These strategies train an X-ray beam on particular regions of a crystal or collections of crystals, collecting a partial or full dataset at each location (Figure 2). The data collected at each location can then be indexed and integrated, and the data between sampled locations correlated to yield multiple datasets per crystal [51]. This approach is especially advantageous when radiation damage is a challenge, and helps explore the possible conformational heterogeneity present within macro-scale crystals that may represent structural variations in their constituent proteins [52]. In the simplest case, regions within a crystal can be non-isomorphous and distinguished by varying unit cell parameters and described by a lattice distortion [51]. Combining data within isomorphous sets then allows for a single structure to be determined per set (Figure 2), a comparison of the structures reveals conformational differences that explain the lattice distortion [51]. With greater data sampling, the potential to mine differences in crystal structures in search for conformational heterogeneity becomes increasingly possible. These approaches, generally referred to as multiple structures from one crystal (MSOX), are not limited to spatial heterogeneity in crystals, but can also add a temporal component to assess changes induced by radiation [50,53] or changes due to environmental differences or molecular actions [54].

Advances in electron diffraction

Improvements to detector technology have played a major role in the interrogation of macromolecular structures from thin three-dimensional crystals by dramatically increasing the rate of accurate data gathering [9,10,55]. The field of electron crystallography of three-dimensional microcrystallites (MicroED) is a marriage of cryoEM sample manipulation and data collection with X-ray crystallography analysis for routine structure determination of macromolecular structures [8]. The MicroED technique, initially demonstrated on the well-known protein standard lysozyme [9], has now been successful in determining a variety of structures from small molecules to macromolecular assemblies (Figure 1) [56–58]. To obtain electron diffraction from 3D protein crystals, approaches like MicroED illuminate a sub-micron thick crystal with an extremely low-dose electron beam (Figure 3). The signal from the illuminated crystal can be isolated from that of its surroundings by means of a selected area aperture (Figure 3). Diffraction is then sampled from the selected area and recorded either as a movie from a continuously rotating crystal (Figure 3) [59] or as a series of precession photographs at discrete tilt angles [60]. Data is reduced by conventional crystallography approaches to yield integrated intensities that can cover part or all of the reciprocal lattice, depending on crystal symmetry and the fraction of

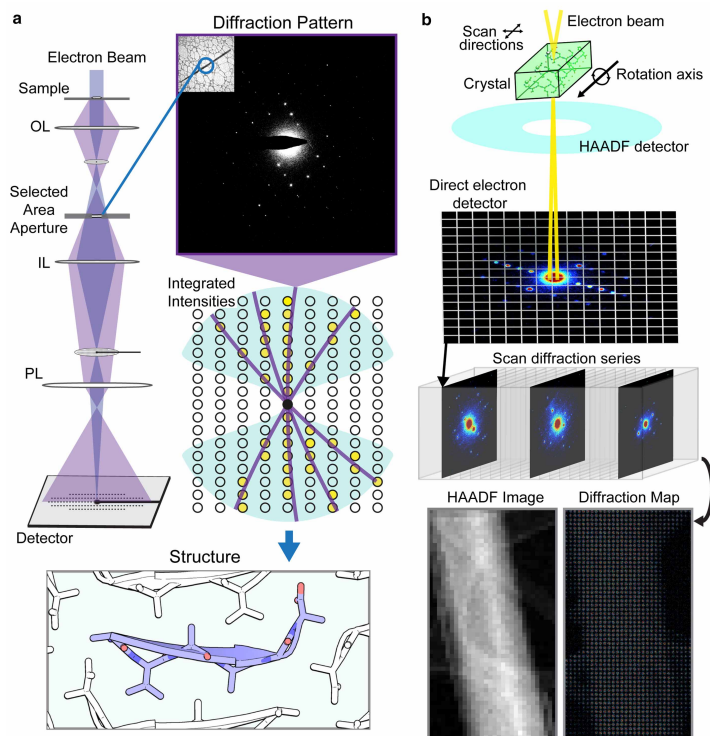


Figure 3. Examples of data gathering by selected area or scanning electron diffraction.

(a) In a conventional electron diffraction experiment, a crystal and its surroundings are illuminated with a low-dose electron beam. A selected area aperture samples diffraction from a portion or all of the illuminated crystal; diffraction patterns are sampled continuously as the crystal is unidirectionally rotated within the electron beam. Integrated intensities from these patterns populate a reciprocal lattice that is phased to determine a molecular structure. OL, objective lens; IL, intermediate lenses; PL, projection lenses. (b) In scanning nanobeam electron diffraction, a focused electron beam is scanned across a sample while diffraction patterns are recorded. Electrons scattered to high angle are integrated by a dark field detector (HAADF) to yield a visual representation of the sample, while electrons scattered to a lower angle travel through a hole in that detector and are recorded by a fast readout, pixelated electron counting detector to yield a diffraction map of the scanned region. Both panels show data collected from nanocrystals of a hexapeptide.

Bragg reflections obstructed by the supporting grid and holder (Figure 3). Structures are determined from this data either by direct methods [61–63], when the resolution of the data is atomic (~ 1.2 Å) or by molecular replacement when a suitable probe exists [8]. Over one hundred macromolecular structures have now been deposited in the PDB, determined by electron diffraction since 1990 (Figure 1). Nearly half of these are determined from 3D protein crystals, and add to a growing number of small molecule structures [64–67].

While the number of structures deposited into the PDB by MicroED is miniscule in comparison with X-ray crystallography efforts (Figure 1), growth in the number of sites collecting MicroED data and the implementation of automation will pose a challenge for efforts promoting archival and access to raw data. This challenge is most prominent for diffraction data collected on fast electron detectors [3,47], especially where multiple datasets must be combined to yield information about a single structure. The simplest of these experiments is one where data from several tilt series is merged to create a single high completeness dataset [8,68], but even this effort can be a challenge since the specific combination of datasets that yields the best merge is not known

a priori. In cases where polymorphism is high, and diffraction quality is highly variable between crystals, obtaining a quality merge may require a collection of data from dozens or more individual crystals. Widespread adoption of this approach necessitates methods for automation in data collection and processing [69,70]. Partial automation is already facilitating more complex experiments, including scanning nanobeam diffraction where thousands of patterns are collected at a given orientation from a single crystallite (Figure 3) [3,71], and serial crystallography experiments that combine diffraction from hundreds or thousands of crystallites using SerialED software [69].

Scanning diffraction experiments are particularly data rich [3]. In a scanning electron diffraction experiment, a focused electron beam is positioned at specific locations across a sample in a defined pattern (Figure 3), analogous to the experiments performed when scanning an X-ray beam across a much larger crystal (Figure 2) [51]. While the beam is positioned at each location for only a small fraction of a second, a diffraction pattern is collected on a fast detector [72]. Data volume in scanning diffraction scales varies with the number of individual scan steps. Depending on the size of a scan area and the density of locations sampled, this procedure can generate tens of thousands of diffraction patterns per scan, multiplied by the number of crystallite tilt angles sampled (Figure 3). Data in each scan is plentiful, but information is sparse; each diffraction pattern in a scan has few Bragg reflections, and each reflection may only be partially sampled [3]. This makes averaging over similar patterns advantageous. Ultimately, all patterns containing a usable signal could be combined into a single set of integrated intensities representing a subset or the whole of reciprocal space. By spatially linking reciprocal space information to specific locations on a crystal, scanning diffraction offers a rich source of information unavailable in a conventional crystallography experiment [3,51]. While a simple analysis of total diffracted intensity can reveal regions of strong or weak diffraction within a crystal, more subtle changes in the diffracted signal can be extracted by algorithms that compare deviations across patterns. Comparison of patterns across a scan can ultimately reveal changes in lattice orientation [3], and potentially changes in unit cell parameters or perhaps changes in the structure of molecules in a crystal.

The role of detectors in big data gathering

The use of advanced detectors for photon and electron counting is expediting data collection in both X-ray and electron crystallography, facilitating the interrogation of crystals with short exposure times using high-flux sources. An outstanding challenge is posed by the rapid readout Adaptive Gain Integrating Pixel Detector [42], a system capable of delivering bursts of 350 images with a 6.5-mHz frame rate [41,73]. This and other fast detectors are now key to enabling new efforts in serial crystallography [43], where their high dynamic range and sensitivity allow rapid readout from single X-ray pulses or short electron beam exposures [43,47,74].

Electron microscopy as a whole has largely benefitted from the development of advanced detectors, first by the introduction of slow readout charged couple device (CCD) sensors [75], then by faster active pixel sensors including the complementary metal-oxide semiconductors (CMOS) including direct electron detectors [76,77]. Each improvement in detector technology has grown data output, facilitating not only improved speed and accuracy of data collection but also more efficient data archival and processing [77]. While the readout rates of several electron detectors are now in the kilohertz range [65,74,76], most experiments are not performed in this regime. Fast readout is instead used to preserve resolution in electron microscopy images by limiting the influence of motion induced artifacts [78]. In continuous rotation electron diffraction, readout rates are adjusted to limit overexposure due to brightly diffracted beams, limit readout noise and ensure adequate sampling of the reciprocal lattice (Figure 3) [8,68,74].

Performing electron diffraction of single crystals at kilohertz frame rates could allow the collection of a 140-degree wedge of reciprocal space while sampling a quarter of a degree per frame in approximately half a second. However, the use of direct electron detectors for measuring diffraction from macromolecular crystals not only facilitates rapid, low-noise, low-dose data collection but can also increase the data rate per experiment by over an order of magnitude. This startling pace of acquisition would have a dramatic effect on our ability to rapidly screen samples, probe heterogeneity in crystallites and ultimately sample more experimental parameters. However, the collection of 14-megapixel, 8-bit images at 400 frames per second yields a torrential 5.6 gigabytes of raw data per second. In line conversion of this data from its archival format to one suitable for crystallographic analysis is the first of many bottlenecks. Since crystallographic data is highly sparse, the greatest improvement to processing of kilohertz diffraction will come from algorithms for rapid and automated data reduction [68,69].

Toward real-time processing

Algorithms for accurate data reduction from sparse diffraction data are key enabling technologies for emerging crystallographic experiments. Rapid hit-finding and indexing algorithms are routinely used in X-ray free electron laser facilities during serial crystallography experiments [79–85], where the rate of image collection far outpaces the amount of data that can be parsed manually in real time. Algorithms used to parse 120 Hz data generated by the Cornell-SLAC Pixel Array Detector can operate concurrently with data collection, giving the crystallographer a notion of how many frames have usable diffraction and how many reflections are present in a frame [81,83–86]. Many of these same packages can create and refine experimental parameter models, index reflections and integrate reflection intensities. Programs such as KAMO [87] and Blend [88] facilitate the identification of consistent unit cells and evaluate the capacity of data subsets to be merged. Their algorithms rely on the relationships between cell parameters or intensity correlations to select suitable merge sets. Whilst a similar feat of automated data processing has only recently been demonstrated in electron diffraction [69]. Although no fundamental hurdles limit the full applicability of these methods to continuous rotation, serial or nanobeam electron diffraction, sparse single electron diffraction patterns may be especially difficult to index without sufficient prior constraints on cell dimension and experimental geometry. The need for a centralized access to a comprehensive set of tools and protocols for analysis of large crystallographic datasets will continue to be underscored by the continued and rapid development of sources, instruments, methods, programs and algorithms [89].

Summary

- New crystallographic approaches are facing new challenges and opportunities delivered by data-rich experiments facilitated by the development of new ultra-fast detectors.
- Electron crystallography experiments are following a similar trend, where faster detectors are allowing the rapid collection of information-rich data from single crystals or collections of crystals.
- New algorithms for automated data collection and analysis are leveraging the rise in data production to yield faster structure solutions, and new insights into the macromolecular structure at the time and storage space limits.

Abbreviations

CCD, charged couple device; CMOS, complementary metal-oxide semiconductors; MSOX, multiple structures from one crystal; MicroED, microcrystallites; SFX, serial femtosecond crystallography.

Funding

This work is supported by DOE Grant DE-FC02-02ER63421, NSF Grant DMR-1548924 and the NIH-NIGMS Grant R35 GM128867. J.A.R. is supported as a Searle Scholar a Pew Scholar and a Beckman Young Investigator.

Acknowledgements

We thank Marcus Gallagher-Jones (UCLA) for his help in preparing figures and contributions to the writing of this manuscript, and Johan Hattne (UCLA) for critical discussion.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

- 1 Martin-Garcia, J.M., Conrad, C.E., Coe, J., Roy-Chowdhury, S. and Fromme, P. (2016) Serial femtosecond crystallography: a revolution in structural biology. *Arch. Biochem. Biophys.* **602**, 32–47 <https://doi.org/10.1016/j.abb.2016.03.036>

- 2 Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A. et al. (2011) Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 <https://doi.org/10.1038/nature09750>
- 3 Gallagher-Jones, M., Ophus, C., Bustillo, K.C., Boyer, D.R., Panova, O., Glynn, C. et al. (2019) Nanoscale mosaicism revealed in peptide microcrystals by scanning electron nanodiffraction. *Commun. Biol.* **2**, 26 <https://doi.org/10.1038/s42003-018-0263-8>
- 4 Gallagher-Jones, M., Rodriguez, J.A. and Miao, J. (2016) Frontier methods in coherent X-ray diffraction for high-resolution structure determination. *Q. Rev. Biophys.* **49**, e20 <https://doi.org/10.1017/S0033583516000147>
- 5 Johansson, L.C., Stauch, B., Ishchenko, A. and Cherezov, V. (2017) A bright future for serial femtosecond crystallography with XFELs. *Trends Biochem. Sci.* **42**, 749–762 <https://doi.org/10.1016/j.tibs.2017.06.007>
- 6 Schlichting, I. (2015) Serial femtosecond crystallography: the first five years. *IUCrJ* **2**, 246–255 <https://doi.org/10.1107/S205225251402702X>
- 7 Clabbers, M.T.B., van Genderen, E., Wan, W., Wieggers, E.L., Gruene, T. and Abrahams, J.P. (2017) Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Crystallogr. D Struct. Biol.* **73**, 738–748 <https://doi.org/10.1107/S2059798317010348>
- 8 Shi, D., Nannenga, B.L., de la Cruz, M.J., Liu, J., Sawtelle, S., Calero, G. et al. (2016) The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 <https://doi.org/10.1038/nprot.2016.046>
- 9 Shi, D., Nannenga, B.L., Iadanza, M.G. and Gonen, T. (2013) Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 <https://doi.org/10.7554/eLife.01345>
- 10 Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. and Toyoshima, C. (2015) Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. *Proc. Natl Acad. Sci. U.S.A.* **112**, 3368–3373 <https://doi.org/10.1073/pnas.1500724112>
- 11 Baldwin, P.R., Tan, Y.Z., Eng, E.T., Rice, W.J., Noble, A.J., Negro, C.J. et al. (2018) Big data in cryoEM: automated collection, processing and accessibility of EM data. *Curr. Opin. Microbiol.* **43**, 1–8 <https://doi.org/10.1016/j.mib.2017.10.005>
- 12 Glaeser, R.M. (2008) Cryo-electron microscopy of biological nanostructures. *Phys. Today* **61**, 48 <https://doi.org/10.1063/1.2835153>
- 13 Taylor, K.A. and Glaeser, R.M. (1974) Electron diffraction of frozen, hydrated protein crystals. *Science* **186**, 1036–1037 <https://doi.org/10.1126/science.186.4168.1036>
- 14 Taylor, K.A. and Glaeser, R.M. (1976) Electron microscopy of frozen hydrated biological specimens. *J. Ultrastruct. Res.* **55**, 448–456 [https://doi.org/10.1016/S0022-5320\(76\)80099-8](https://doi.org/10.1016/S0022-5320(76)80099-8)
- 15 Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. and Downing, K.H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 [https://doi.org/10.1016/S0022-2836\(05\)80271-2](https://doi.org/10.1016/S0022-2836(05)80271-2)
- 16 Adrian, M., Dubochet, J., Lepault, J. and McDowell, A.W. (1984) Cryo-electron microscopy of viruses. *Nature* **308**, 32–36 <https://doi.org/10.1038/308032a0>
- 17 Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W. et al. (1988) Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 <https://doi.org/10.1017/S0033583500004297>
- 18 Cheng, Y., Grigorieff, N., Penczek, P.A. and Walz, T. (2015) A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 <https://doi.org/10.1016/j.cell.2015.03.050>
- 19 Baker, M.L., Zhang, J., Ludtke, S.J. and Chiu, W. (2010) Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat. Protoc.* **5**, 1697–1708 <https://doi.org/10.1038/nprot.2010.126>
- 20 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242 <https://doi.org/10.1093/nar/28.1.235>
- 21 Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P. et al. (2011) EMDatabank.org: unified data resource for CryoEM. *Nucleic Acids Res.* **39**, D456–D464 <https://doi.org/10.1093/nar/gkq880>
- 22 Iudin, A., Korir, P.K., Salavert-Torres, J., Kleywegt, G.J. and Patwardhan, A. (2016) EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 <https://doi.org/10.1038/nmeth.3806>
- 23 Maia, F.R.N.C. (2012) The coherent X-ray imaging data bank. *Nat. Methods* **9**, 854–855 <https://doi.org/10.1038/nmeth.2110>
- 24 Pellegrini, C. (2016) X-ray free-electron lasers: from dreams to reality. *Phys. Scr.* **T 169**, 014004 <https://doi.org/10.1088/1402-4896/aa5281>
- 25 Schlichting, I. and Miao, J. (2012) Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* **22**, 613–626 <https://doi.org/10.1016/j.sbi.2012.07.015>
- 26 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J. (2000) Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 <https://doi.org/10.1038/35021099>
- 27 Barty, A., Caleman, C., Aquila, A., Timneanu, N., Lomb, L., White, T.A. et al. (2012) Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photon.* **6**, 35–40 <https://doi.org/10.1038/nphoton.2011.297>
- 28 Spence, J.C.H., Weierstall, U. and Chapman, H.N. (2012) X-ray lasers for structural and dynamic biology. *Rep. Prog. Phys.* **75**, 102601 <https://doi.org/10.1088/0034-4885/75/10/102601>
- 29 Boutet, S., Lomb, L., Williams, G.J., Barends, T.R., Aquila, A., Doak, R.B. et al. (2012) High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 <https://doi.org/10.1126/science.1217737>
- 30 Martiel, I., Müller-Werkmeister, H.M. and Cohen, A.E. (2019) Strategies for sample delivery for femtosecond crystallography. *Acta Crystallogr. D Struct. Biol.* **75**, 160–177 <https://doi.org/10.1107/S2059798318017953>
- 31 Johansson, L.C., Arnlund, D., Katona, G., White, T.A., Barty, A., DePonte, D.P. et al. (2013) Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography. *Nat. Commun.* **4**, 2911 <https://doi.org/10.1038/ncomms3911>
- 32 Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N.A., Rendek, K.N. et al. (2014) Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 <https://doi.org/10.1038/nature13453>
- 33 Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D. et al. (2013) Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521–1524 <https://doi.org/10.1126/science.1244142>
- 34 Zhang, H., Unal, H., Gati, C., Han, G.W., Liu, W., Zatsepin, N.A. et al. (2015) Structure of the angiotensin receptor revealed by serial femtosecond crystallography. *Cell* **161**, 833–844 <https://doi.org/10.1016/j.cell.2015.04.011>
- 35 Sawaya, M.R., Cascio, D., Gingery, M., Rodriguez, J., Goldschmidt, L., Colletier, J.-P. et al. (2014) Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl Acad. Sci. U.S.A.* **111**, 12769–12774 <https://doi.org/10.1073/pnas.1413456111>

- 36 Duszko, M., Redecke, L., Mudogo, C.N., Sommer, B.P., Mogk, S., Oberthuer, D. et al. (2015) *In vivo* protein crystallization in combination with highly brilliant radiation sources offers novel opportunities for the structural analysis of post-translationally modified eukaryotic proteins. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 929–937 <https://doi.org/10.1107/S2053230X15011450>
- 37 Redecke, L., Nass, K., DePonte, D.P., White, T.A., Rehders, D., Barty, A. et al. (2013) Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227–230 <https://doi.org/10.1126/science.1229663>
- 38 Colletier, J.-P., Sawaya, M.R., Gingery, M., Rodriguez, J.A., Cascio, D., Brewster, A.S. et al. (2016) *De novo* phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, 43–47 <https://doi.org/10.1038/nature19825>
- 39 Tenboer, J., Basu, S., Zatselin, N., Pande, K., Milathianaki, D., Frank, M. et al. (2014) Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 <https://doi.org/10.1126/science.1259357>
- 40 Coquelle, N., Silva, M., Woodhouse, J., Schirò, G., Adam, V., Aquila, A. et al. (2018) Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 <https://doi.org/10.1038/nchem.2853>
- 41 Grünbein, M.L., Bielecki, J., Gorel, A., Stricker, M., Bean, R., Cammarata, M. et al. (2018) Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9**, 3487 <https://doi.org/10.1038/s41467-018-05953-4>
- 42 Henrich, B., Becker, J., Dinapoli, R., Goettlicher, P., Graafsma, H., Hirsemann, H. et al. (2011) The adaptive gain integrating pixel detector AGIPD a detector for the European XFEL. *Nucl. Instrum. Methods Phys. Res. A* **633**, S11–S14 <https://doi.org/10.1016/j.nima.2010.06.107>
- 43 Martín-García, J.M., Conrad, C.E., Nelson, G., Stander, N., Zatselin, N.A., Zook, J. et al. (2017) Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. *IUCrJ* **4**, 439–454 <https://doi.org/10.1107/S205225251700570X>
- 44 Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D. et al. (2017) Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* **8**, 542 <https://doi.org/10.1038/s41467-017-00630-4>
- 45 Botha, S., Nass, K., Barends, T.R.M., Kabsch, W., Latz, B., Dworowski, F. et al. (2015) Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 387–397 <https://doi.org/10.1107/S1399004714026327>
- 46 Ebrahim, A., Appleby, M.V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D.A. et al. (2019) Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography. *Acta Crystallogr. D Struct. Biol.* **75**, 151–159 <https://doi.org/10.1107/S2059798318010240>
- 47 Casanas, A., Warshamanage, R., Finke, A.D., Panepucci, E., Olieric, V., Nöll, A. et al. (2016) EIGER detector: application in macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **72**, 1036–1048 <https://doi.org/10.1107/S2059798316012304>
- 48 Coquelle, N., Brewster, A.S., Kapp, U., Shilova, A., Weinhausen, B., Burghammer, M. et al. (2015) Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1184–1196 <https://doi.org/10.1107/S1399004715004514>
- 49 Graves, W.S., Bessuille, J., Brown, P., Carbajo, S., Dolgashev, V., Hong, K.-H. et al. (2014) Compact x-ray source based on burst-mode inverse Compton scattering at 100 kHz. *Phys. Rev. ST Accel. Beams* **17**, 120701 <https://doi.org/10.1103/PhysRevSTAB.17.120701>
- 50 Horrell, S., Antonyuk, S.V., Eady, R.R., Hasnain, S.S., Hough, M.A. and Strange, R.W. (2016) Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal. *IUCrJ* **3**, 271–281 <https://doi.org/10.1107/S205225251600823X>
- 51 Thompson, M.C., Cascio, D. and Yeates, T.O. (2018) Microfocus diffraction from different regions of a protein crystal: structural variations and unit-cell polymorphism. *Acta Crystallogr. D Struct. Biol.* **74**, 411–421 <https://doi.org/10.1107/S2059798318003479>
- 52 van den Bedem, H. and Fraser, J.S. (2015) Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat. Methods* **12**, 307–318 <https://doi.org/10.1038/nmeth.3324>
- 53 Horrell, S., Kekilli, D., Sen, K., Owen, R.L., Dworowski, F.S.N., Antonyuk, S.V. et al. (2018) Enzyme catalysis captured using multiple structures from one crystal at varying temperatures. *IUCrJ* **5**, 283–292 <https://doi.org/10.1107/S205225251800386X>
- 54 Ren, Z., Chan, P.W.Y., Moffat, K., Pai, E.F., Royer, W.E., Šrajer, V. et al. (2013) Resolution of structural heterogeneity in dynamic crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 946–959 <https://doi.org/10.1107/S0907444913003454>
- 55 Nederlof, I., van Genderen, E., Li, Y.-W. and Abrahams, J.P. (2013) A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1223–1230 <https://doi.org/10.1107/S0907444913009700>
- 56 de la Cruz, M.J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F.E. et al. (2017) Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 <https://doi.org/10.1038/nmeth.4178>
- 57 Rodriguez, J.A., Eisenberg, D.S. and Gonen, T. (2017) Taking the measure of MicroED. *Curr. Opin. Struct. Biol.* **46**, 79–86 <https://doi.org/10.1016/j.sbi.2017.06.004>
- 58 Clabbers, M.T.B. and Abrahams, J.P. (2018) Electron diffraction and three-dimensional crystallography for structural biology. *Crystallogr. Rev.* **24**, 176–204 <https://doi.org/10.1080/0889311X.2018.1446427>
- 59 Nannenga, B.L., Shi, D., Leslie, A.G.W. and Gonen, T. (2014) High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930 <https://doi.org/10.1038/nmeth.3043>
- 60 Midgley, P.A. and Eggeman, A.S. (2015) Precession electron diffraction—a topical review. *IUCrJ* **2**, 126–136 <https://doi.org/10.1107/S2052252514022283>
- 61 Hauptman, H. (1986) The direct methods of X-ray crystallography. *Science* **233**, 178–183 <https://doi.org/10.1126/science.233.4760.178>
- 62 Usón, I. and Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 [https://doi.org/10.1016/S0959-440X\(99\)00020-2](https://doi.org/10.1016/S0959-440X(99)00020-2)
- 63 Sawaya, M.R., Rodriguez, J., Cascio, D., Collazo, M.J., Shi, D., Reyes, F.E. et al. (2016) *Ab initio* structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl Acad. Sci. U.S.A.* **113**, 11232–11236 <https://doi.org/10.1073/pnas.1606287113>
- 64 Jones, C.G., Martynowycz, M.W., Hattne, J., Fulton, T.J., Stoltz, B.M., Rodriguez, J.A. et al. (2018) The CryoEM method MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **4**, 1587–1592 <https://doi.org/10.1021/acscentsci.8b00760>
- 65 Gruene, T., Wennmacher, J.T.C., Zaubitzer, C., Holstein, J.J., Heidler, J., Fecteau-Lefebvre, A. et al. (2018) Rapid structure determination of microcrystalline molecular compounds using electron diffraction. *Angew. Chem. Int. Ed.* **57**, 16313–16317 <https://doi.org/10.1002/anie.201811318>
- 66 van Genderen, E., Clabbers, M.T.B., Das, P.P., Stewart, A., Nederlof, I., Barendsen, K.C. et al. (2016) *Ab initio* structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector. *Acta Crystallogr. A Found. Adv.* **72**, 236–242 <https://doi.org/10.1107/S2053273315022500>

- 67 Palatinus, L., Brázda, P., Boullay, P., Perez, O., Klementová, M., Petit, S. et al. (2017) Hydrogen positions in single nanocrystals revealed by electron diffraction. *Science* **355**, 166–169 <https://doi.org/10.1126/science.aak9652>
- 68 Hattne, J., Reyes, F.E., Nannenga, B.L., Shi, D., de la Cruz, M.J., Leslie, A.G.W. et al. (2015) MicroED data collection and processing. *Acta Crystallogr. A Found. Adv.* **71**, 353–360 <https://doi.org/10.1107/S2053273315010669>
- 69 Smeets, S., Zou, X. and Wan, W. (2018) Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* **51**, 1262–1273 <https://doi.org/10.1107/S1600576718009500>
- 70 de la Cruz, M.J., Martynowycz, M.W., Hattne, J. and Gonen, T. (2019) MicroED data collection with SerialEM. *Ultramicroscopy* **201**, 77–80 <https://doi.org/10.1016/j.ultramic.2019.03.009>
- 71 Lanza, A., Margheritis, E., Mugnaioli, E., Cappello, V., Garau, G. and Gemmi, M. (2019) Nanobeam precession-assisted 3D electron diffraction reveals a new polymorph of hen egg-white lysozyme. *IUCr* **6**, 178–188 <https://doi.org/10.1107/S2052252518017657>
- 72 Bustillo, K.C., Panova, O., Chen, X.C., Takacs, C.J., Ciston, J., Ophus, C. et al. (2017) Nanobeam scanning diffraction for orientation mapping of polymers. *Microsc. Microanal.* **23**, 1782–1783 <https://doi.org/10.1017/S1431927617009576>
- 73 Allangholi, A., Becker, J., Delfs, A., Dinapoli, R., Goettlicher, P., Greiffenberg, D. et al. (2019) The adaptive gain integrating pixel detector at the European XFEL. *J. Synchrotron Radiat.* **26**, 74–82 <https://doi.org/10.1107/S1600577518016077>
- 74 Tinti, G., Fröjd, E., van Genderen, E., Gruene, T., Schmitt, B., de Winter, D.A.M. et al. (2018) Electron crystallography with the EIGER detector. *IUCr* **5**, 190–199 <https://doi.org/10.1107/S2052252518000945>
- 75 Fan, G.Y. and Ellisman, M.H. (2000) Digital imaging in transmission electron microscopy. *J. Microsc.* **200**, 1–13 <https://doi.org/10.1046/j.1365-2818.2000.00737.x>
- 76 Contarato, D., Denes, P., Doering, D., Joseph, J. and Krieger, B. (2012) High speed, radiation hard CMOS pixel sensors for transmission electron microscopy. *Phys. Procedia* **37**, 1504–1510 <https://doi.org/10.1016/j.phpro.2012.04.103>
- 77 Ruskin, R.S., Yu, Z. and Grigorieff, N. (2013) Quantitative characterization of electron detectors for transmission electron microscopy. *J. Struct. Biol.* **184**, 385–393 <https://doi.org/10.1016/j.jsb.2013.10.016>
- 78 Li, X., Mooney, P., Zheng, S., Booth, C.R., Braumfeld, M.B., Gubbens, S. et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 <https://doi.org/10.1038/nmeth.2472>
- 79 Brewster, A.S., Sawaya, M.R., Rodriguez, J., Hattne, J., Echols, N., McFarlane, H.T. et al. (2015) Indexing amyloid peptide diffraction from serial femtosecond crystallography: new algorithms for sparse patterns. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 357–366 <https://doi.org/10.1107/S1399004714026145>
- 80 Sauter, N.K., Hattne, J., Brewster, A.S., Echols, N., Zwart, P.H. and Adams, P.D. (2014) Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 3299–3309 <https://doi.org/10.1107/S1399004714024134>
- 81 White, T.A. (2019) Processing serial crystallography data with *CrystFEL*: a step-by-step guide. *Acta Crystallogr. D Struct. Biol.* **75**, 219–233 <https://doi.org/10.1107/S205979831801238X>
- 82 White, T.A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K.R. et al. (2016) Recent developments in *CrystFEL*. *J. Appl. Crystallogr.* **49**, 680–689 <https://doi.org/10.1107/S1600576716004751>
- 83 Sauter, N.K., Hattne, J., Grosse-Kunstleve, R.W. and Echols, N. (2013) New Python-based methods for data processing. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1274–1282 <https://doi.org/10.1107/S0907444913000863>
- 84 Nakane, T., Joti, Y., Tono, K., Yabashi, M., Nango, E., Iwata, S. et al. (2016) Data processing pipeline for serial femtosecond crystallography at SACLA. *J. Appl. Crystallogr.* **49**, 1035–1041 <https://doi.org/10.1107/S1600576716005720>
- 85 Ginn, H.M., Evans, G., Sauter, N.K. and Stuart, D.I. (2016) On the release of *fcpxfel* for processing X-ray free-electron laser images. *J. Appl. Crystallogr.* **49**, 1065–1072 <https://doi.org/10.1107/S1600576716006981>
- 86 Lyubimov, A.Y., Uevirojnangkorn, M., Zeldin, O.B., Brewster, A.S., Murray, T.D., Sauter, N.K. et al. (2016) *IOTA*: integration optimization, triage and analysis tool for the processing of XFEL diffraction images. *J. Appl. Crystallogr.* **49**, 1057–1064 <https://doi.org/10.1107/S1600576716006683>
- 87 Yamashita, K., Hirata, K. and Yamamoto, M. (2018) *KAMO*: towards automated data processing for microcrystals. *Acta Crystallogr. D Struct. Biol.* **74**, 441–449 <https://doi.org/10.1107/S2059798318004576>
- 88 Mylona, A., Carr, S., Aller, P., Moraes, I., Treisman, R., Evans, G. et al. (2017) A novel approach to data collection for difficult structures: data management for large numbers of crystals with the BLEND software. *Crystals* **7**, 242 <https://doi.org/10.3390/cryst7080242>
- 89 Grimes, J.M., Hall, D.R., Ashton, A.W., Evans, G., Owen, R.L., Wagner, A. et al. (2018) Where is crystallography going? *Acta Crystallogr. D Struct. Biol.* **74**, 152–166 <https://doi.org/10.1107/S2059798317016709>